**B.Tech. BCSE497J - Project-I**

**Plant Disease Classification using AI**

|  |  |
| --- | --- |
| **21BCE0472** | **PRANIT PRAVEEN KAUL** |
| **21BCE2510** | **SHIVAM KOUL** |
| **21BDS0049** | **PUSHKAR KUMAR GAUTAM** |

Under the Supervision of

|  |
| --- |
| **Prof. BALAJI N** |
| Assistant Professor Sr. Grade 1 |
| School of Computer Science and Engineering (SCOPE) |

**B.Tech.**

*in*

**Computer Science and Engineering**

**School of Computer Science and Engineering**



November 2024

Plant Disease Classification using AI

# 1.Abstract

This research presents an advanced approach to automated plant disease classification using a two-stage transfer learning technique based on the ResNet-50 architecture. The primary objective is to develop a highly accurate and scalable solution for identifying plant diseases from image inputs, addressing critical challenges in precision agriculture.

Our novel methodology employs a two-stage transfer learning process:

1. In the first stage, we pre-train a ResNet-50 model (Model A) on an extensive 25 GB dataset of diverse leaf images, enabling the network to learn robust, generalizable features across a wide range of plant species and conditions.
2. The second stage involves fine-tuning Model A on the Plant Village dataset, a curated collection of plant disease images, to specialize the model for specific disease recognition tasks.

The dataset undergoes comprehensive preprocessing, including advanced data augmentation techniques, intelligent shuffling, efficient caching, and optimized batching. Our pipeline leverages TensorFlow's *tf.data* API to enhance data handling, significantly improving training efficiency and model scalability.

Experimental results demonstrate the effectiveness of our approach, with the model achieving an impressive classification accuracy exceeding 96% on the Plant Village dataset. This marks a substantial improvement over previous benchmarks and single-stage transfer learning methods.

Key contributions of this work include:

1. Development of a two-stage transfer learning methodology that effectively leverages large-scale, diverse leaf data to improve disease classification accuracy.
2. Demonstration of superior model generalization across various plant species and disease types.
3. Implementation of an efficient data pipeline capable of handling extensive datasets, enhancing training speed and scalability.
4. Achieving state-of-the-art accuracy (>96%) in plant disease classification, setting a new benchmark for automated agricultural diagnostics.

This research significantly advances the field of precision agriculture by providing a highly accurate, efficient, and scalable tool for early plant disease detection. The proposed model has the potential to revolutionize crop health management, enabling timely interventions and substantially improving agricultural productivity and sustainability.

# 2.Literature Survey and Gap Identification

## Literature Survey

Our base paper for this project is <https://ieeexplore.ieee.org/document/9726036>. Using our aim is to improve the results of the findings using 2 stage transfer learning as proposed below.   
  
We used [ResNet50](https://arxiv.org/abs/1512.03385) model as this model shows promising results when trained on big dataset and has great potential for transfer learning. ResNet models outperforms [VGG models](https://arxiv.org/abs/1512.03385) on imagenet dataset.  
  
[ResNet Strikes back](https://arxiv.org/pdf/2110.00476) paper shows how improved training and finetuning still keeps this model relevant and has great potential even to this date.

Plant disease detection and classification using deep learning techniques have become crucial in modern agriculture, offering the potential to significantly improve crop management and yield. This study aims to build upon and improve the results of Gosai et al. (2022), who proposed a comprehensive approach to plant disease detection using machine learning algorithms.

Base Study: Plant Disease Detection Using Machine Learning

Gosai et al. (2022) presented a thorough review and implementation of machine learning techniques for plant disease detection. Their study compared various classification algorithms, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN). Key aspects of their work included:

1. A systematic approach to plant disease detection, including image acquisition, preprocessing, segmentation, feature extraction, and classification.
2. Comparison of traditional machine learning algorithms with deep learning approaches.
3. Implementation of data augmentation techniques to enhance model performance.
4. Achieving an accuracy of 70.14% using a Random Forest classifier on a dataset of papaya leaf images.

While their study provided a solid foundation, there is significant room for improvement in accuracy and generalization across different plant species and diseases.

***Proposed Improvement: Two-Stage Transfer Learning with ResNet50***

To enhance the results of Gosai et al., we propose utilizing a two-stage transfer learning approach based on the ResNet50 architecture. This choice is motivated by several factors:

1. ResNet50's Strong Performance: He et al. (2016) introduced the ResNet architecture, which has become a cornerstone in deep learning for image classification tasks. ResNet models have consistently outperformed other architectures, including VGG, on large-scale datasets like ImageNet.
2. Transfer Learning Efficacy: Zhang et al. (2023), although focused on steel surface defects, demonstrated the effectiveness of transfer learning using ResNet50, achieving 99.4% accuracy. Their success in a different domain suggests the potential for similar improvements in plant disease detection.
3. Recent Advancements in ResNet Training: Wightman et al. (2021), in their "ResNet Strikes Back" paper, showed that with improved training procedures and fine-tuning techniques, ResNet models remain highly competitive and relevant. They achieved state-of-the-art performance on ImageNet classification, demonstrating the untapped potential of ResNet architectures.

## Gap Identification

### Current State of Research

The [base paper](https://ieeexplore.ieee.org/document/9726036) utilizes a deep learning approach, specifically ResNet18, to classify plant diseases. It applies a single-stage transfer learning method on a dataset of 54,306 images, achieving an impressive accuracy of 96%. This model effectively detects and classifies diseases from images of plant leaves, contributing significantly to precision agriculture by automating the disease detection process.

### Limitations and Gaps

Despite the success of the base paper, several key limitations exist:

* **Dataset Size and Specificity**: The dataset used in the base paper, while comprehensive, covers a generalized set of plant leaves and diseases. There is room for improvement by focusing on a more specialized dataset that better captures the nuances of specific plant diseases.
* **Single-Stage Transfer Learning**: The base paper relies on a single-stage transfer learning process, which may limit the model's capacity to leverage broader knowledge for enhanced fine-tuning on a smaller, more specific dataset.
* **Accuracy Limitation**: While 96% accuracy is commendable, there is a gap in further refining the model to push beyond this accuracy threshold, particularly by integrating a more nuanced, multi-stage learning process.

Given the importance of early and accurate plant disease detection in agriculture, enhancing model performance is crucial for practical implementation. The gap lies in the potential to improve accuracy beyond 96%, especially with better feature extraction from more specific datasets and optimized learning strategies. Addressing these gaps could lead to a more robust and generalizable model that performs better across various plant species and disease types.

We propose using a **2-stage transfer learning approach** to build on the work done in the base paper. Our approach first trains the ResNet50 model on a larger, general dataset of plant leaves, capturing broad feature representations. Then, we fine-tune this pre-trained model on the more specific Plant Village dataset, which focuses on distinct plant diseases. This method is expected to further refine the model's ability to generalize and improve classification accuracy, surpassing the 96% achieved in the base paper.

By addressing these identified gaps, our project aims to push the boundaries of plant disease classification accuracy and provide a more effective tool for precision agriculture.

# 3.Objectives and Project plan

# 4.Requirement Analysis

# 5.Design and Implementation