A Comparative Study of Few-Shot Classification Models for Plant Disease Identification

Authors: Bardia Akbari, Navid Naserazad

Abstract

The accurate and timely identification of plant diseases is crucial for maintaining crop yield and ensuring food security. While deep learning has shown immense promise in this domain, its effectiveness is often hampered by the scarcity of labeled data for many diseases. Few-Shot Learning (FSL) offers a compelling solution by enabling models to generalize from a very limited number of examples. This report details a project aimed at applying and comparing different FSL techniques to the PlantVillage dataset. We investigate two prominent FSL approaches: the metric-based Prototypical Network and the reconstruction-based Feature Map Reconstruction Network (FRN). For the Prototypical Network, we systematically evaluated three distinct feature extraction backbones: a pre-trained Vision Transformer (ViT), a pre-trained ResNet50, and a custom four-layer convolutional network (Conv-4) trained from scratch. Our experiments show that the FRN model achieves the highest performance, with a 5-way 1-shot accuracy of 88.75% and a 5-way 5-shot accuracy of 95.58%, demonstrating the power of reconstruction-based methods for this task.

1. Introduction

Precision agriculture relies heavily on the automated detection of plant diseases to facilitate early intervention and minimize economic losses. Conventional deep learning models, particularly Convolutional Neural Networks (CNNs), have excelled at image classification tasks but typically require vast amounts of labeled data for training. In agriculture, collecting and annotating such large datasets for every possible plant disease is often impractical, as some diseases are rare or emerge unexpectedly.

This data bottleneck motivates the exploration of Few-Shot Learning (FSL), a paradigm that trains models to learn new classes from only a handful of examples. This project implements and rigorously evaluates two powerful FSL methodologies for classifying plant diseases from the PlantVillage dataset:

- **Feature Map Reconstruction Networks (FRN)**: A novel approach that reframes few-shot classification as a feature reconstruction problem in a latent space.
- **Prototypical Networks**: A well-established metric-based method that classifies based on distances to class prototypes in an embedding space.

To understand the impact of the feature extractor on the performance of Prototypical Networks, we tested three different backbone architectures: a pre-trained Vision Transformer (ViT), a pre-trained ResNet50, and a standard shallow CNN (Conv-4) trained from scratch. This report presents our methodology, data analysis, experimental setup, and a detailed discussion of the results, providing a clear comparison of these different FSL approaches.

2. Dataset and Analysis

We used the PlantVillage dataset, which contains a large number of images of healthy and diseased plant leaves.

2.1. Data Analysis Key Findings

- The dataset is organized into subdirectories, with each subdirectory representing a specific plant disease category or a healthy plant category.
- The dataset contains varying numbers of images per category, with *Tomato**Tomato_YellowLeaf**Curl_Virus* and *Tomato_Bacterial_spot* having significantly more images than categories like *Potato___healthy* and *Tomato__Tomato_mosaic_virus*.
- Sampled images across all categories were found to have uniform dimensions of 256x256 pixels.

Figure 1: Distribution of Images per Plant Disease Category

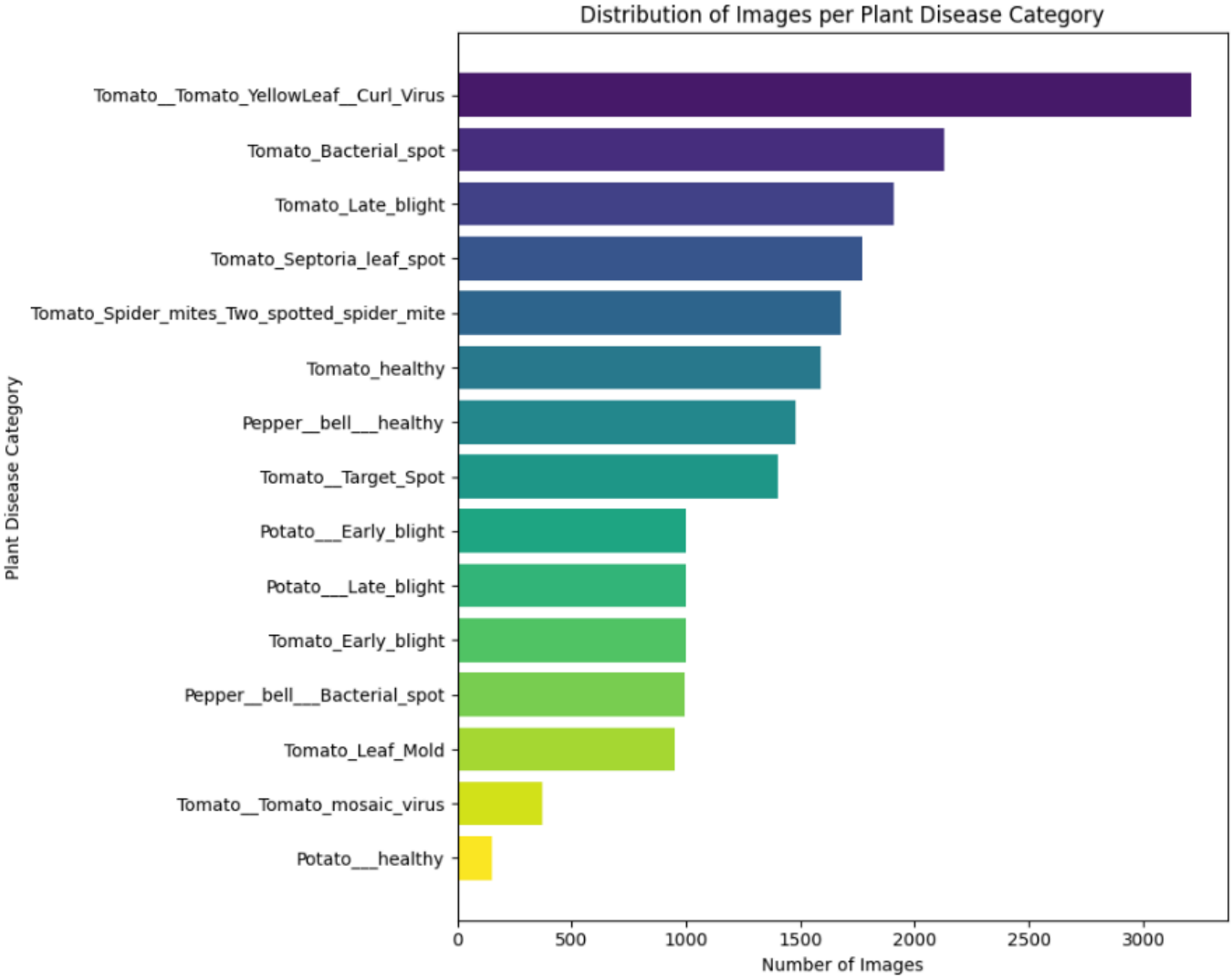
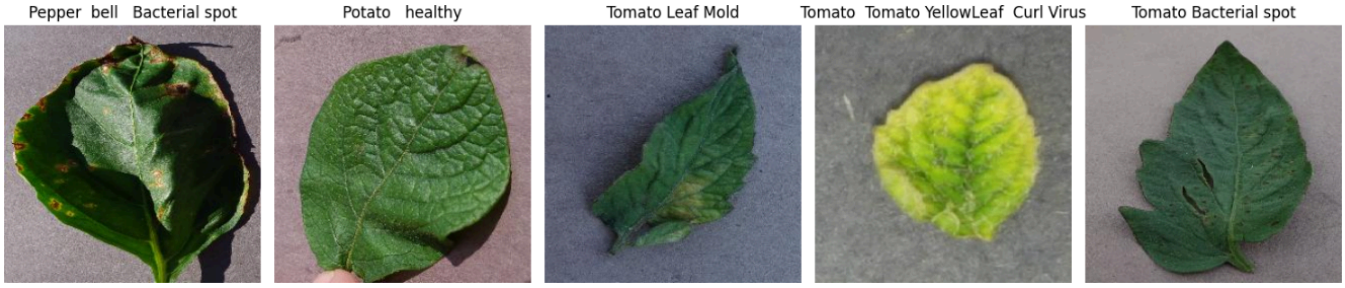


Figure 2: Sample Images from the PlantVillage Dataset



2.2. Dataset Preprocessing

To properly evaluate the FSL models, it is crucial that the classes seen during training, validation, and testing are disjoint. Our *PlantVillageDatasetPreparer* class handles this by splitting the 38 available classes into three distinct sets:

- **Meta-Train (26 classes)**: Used for training the model episodically.
- **Meta-Validation (6 classes)**: Used to monitor training progress and save the best-performing model.

- **Meta-Test (6 classes)**: A completely held-out set of classes used for final performance evaluation.

All images were resized to 224x224 pixels. For training, we applied random horizontal flipping and color jittering. All images were normalized using standard ImageNet statistics.

3. Methodology

3.1. Few-Shot Learning

Few-Shot Learning aims to solve a classification task where only a few labeled examples (the support set) are available for each new class. The model's ability to generalize is then tested on unseen examples (the query set). This is often framed as an "N-way K-shot" problem, where N is the number of classes and K is the number of support examples per class.

3.2. Feature Map Reconstruction Networks (FRN)

As proposed by Wertheimer et al., FRN reformulates few-shot classification as a reconstruction problem. Instead of comparing a query image's embedding to class prototypes, it attempts to reconstruct the query image's feature map using the feature maps from a given class's support set. The classification score is based on the reconstruction error; a lower error suggests the query image belongs to that class. This approach allows the model to leverage fine-grained spatial details in the feature maps without overfitting to a specific pose.

3.3. Prototypical Networks

This is a metric learning method that operates on a simple but powerful principle: there exists an embedding space where images from the same class cluster together. The algorithm proceeds as follows:

- **Embedding**: A neural network backbone, or encoder, maps all input images (both support and query) into a high-dimensional feature space.
- **Prototype Calculation**: For each of the N classes, a single prototype vector is computed by taking the element-wise mean of the embeddings of its K support images. This prototype represents a central point for that class in the embedding space.
- **Classification**: A query image is classified by calculating its distance (e.g., Euclidean distance) to each of the N class prototypes. The query image is assigned the label of the class corresponding to the nearest prototype.

3.4. Backbone Architectures

The effectiveness of a Prototypical Network is highly dependent on the quality of its encoder. We experimented with three diverse backbones:

- **ViT (Vision Transformer)**: We used the *vit_small_patch16_224* model, pre-trained on ImageNet.
- **ResNet50**: A 50-layer deep CNN, pre-trained on ImageNet.
- **Conv-4 (Custom CNN)**: A classic four-layer CNN trained entirely from scratch on the PlantVillage meta-training set.

4. Experimental Setup

The core of our training logic is managed by the *MetaLearner* class, which handles the episodic training loop, optimization, and evaluation.

Configurations: We evaluated all models under two standard FSL scenarios:

- **5-way 1-shot**: 5 classes, 1 support image per class.
- **5-way 5-shot**: 5 classes, 5 support images per class.

Training Details:

- **Optimizer**: AdamW with a learning rate of 1e-4.
- **Scheduler**: A StepLR scheduler that reduces the learning rate by a factor of 0.5 every 10 epochs.
- **Epochs**: The pre-trained models and FRN were trained for 10 epochs, while the from-scratch Conv-4 model was trained for 30 epochs.
- **Evaluation**: Final performance was measured on the meta-test set over 10,000 randomly generated episodes for FRN and 1,000 for the other models.

5. Results and Discussion

5.1. 5-Way 1-Shot Results

Backbone	Best Validation Accuracy	Meta-Test Accuracy (Mean ± Std Dev)	95% Confidence Interval
FRN	-	88.75% ± 0.14%	-
ViT (Pre-trained)	91.36%	87.83% ± 8.44%	[87.30%, 88.35%]
ResNet50 (Pre-trained)	81.76%	74.68% ± 10.74%	[74.01%, 75.35%]
Conv-4 (From Scratch)	65.36%	61.07% ± 11.96%	[60.33%, 61.81%]

5.2. 5-Way 5-Shot Results

Backbone	Best Validation Accuracy	Meta-Test Accuracy (Mean ± Std Dev)	95% Confidence Interval
FRN	94.945	95.58% ± 0.06%	-
ViT (Pre-trained)	97.68%	95.50% ± 4.41%	[95.22%, 95.77%]
ResNet50 (Pre-trained)	93.60%	90.94% ± 6.24%	[90.55%, 91.33%]
Conv-4 (From Scratch)	82.16%	78.96% ± 9.54%	[78.37%, 79.55%]

5.3. Discussion

The results clearly demonstrate the effectiveness of modern FSL techniques for plant disease classification. The Feature Map Reconstruction Network (FRN) emerged as the top-performing model in both the 1-shot and 5-shot scenarios, highlighting the strength of reconstruction-based methods. Its ability to leverage spatial feature details gives it a slight edge over other approaches.

Closely following is the Prototypical Network with a pre-trained ViT backbone. In the challenging 1-shot setting, ViT significantly outperforms the ResNet50 and Conv-4 backbones, underscoring the value of the powerful feature representations learned by transformers on large-scale datasets.

As expected, all models show a substantial improvement in the 5-shot setting. With more support examples, the calculated prototypes and feature reconstructions become more robust, leading to more accurate classification. The performance gap between the models persists, but the Conv-4 model benefits the most from the additional data, with its accuracy increasing by nearly 18 percentage points. This suggests that while shallow models trained from scratch are weak in extreme low-data regimes, they can become more effective as the number of shots increases.

6. Conclusion and Future Work

This project successfully implemented and compared several FSL models for plant disease classification. Our findings lead to two main conclusions:

1. FSL is a highly effective paradigm for this task, with both reconstruction-based (FRN) and metric-based (Prototypical Networks) methods achieving high accuracy. FRN showed the best performance, indicating that feature reconstruction is a very promising direction for fine-grained classification tasks.

2. The choice of backbone architecture is paramount for metric-learning methods. Pre-trained models, particularly the Vision Transformer, provide a significant advantage by leveraging knowledge transferred from large-scale datasets.

Future work could explore hybrid models that combine reconstruction and metric-learning principles or investigate the application of these models in a real-world agricultural setting with images captured directly from the field.