Cassava Disease Classification

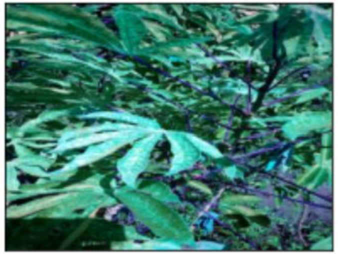
By: Matthew Ocando & Daphne Hong

Table of Contents

About the Dataset

The dataset consists of **21,397 images** belonging to the following classes:

- Cassava Bacterial Blight (CBB)
- Cassava Brown Streak Disease (CBSD)
- Cassava Green Mottle (CGM)
- Cassava Mosaic Disease (CMD)
- Healthy



Cassava Green Mottle (CGM)

Cassava Mosaic Disease (CMD)

Cassava Brown Streak Disease (CBSD)

Defining the Problem – Cassava Plants

Key to Food Security

- Second-largest source of carbohydrates in Africa
- Can withstand harsh conditions

Impacts of Viral Diseases

- Can significantly impact yield
- Plants need to be treated quickly to limit spread

Current Solution

- Manual process, labor intensive & costly
- Constraints due to limited access to resources (e.g. mobile-quality cameras, low bandwidth etc.).

Problem Statement

Can the process of Cassava disease classification be made more efficient (e.g. faster, less labor intensive, and costly)?

02.

Data Preprocessing

Preprocessing

Checking for Missing Values

Data did not have 'N/A' that would require cleaning.

Normalizing the Data

Scale pixels to [0, 1] for efficiency during the training.

Data Augmentation

Rotation, zoom, horizontal/vertical shift, shear, & flip

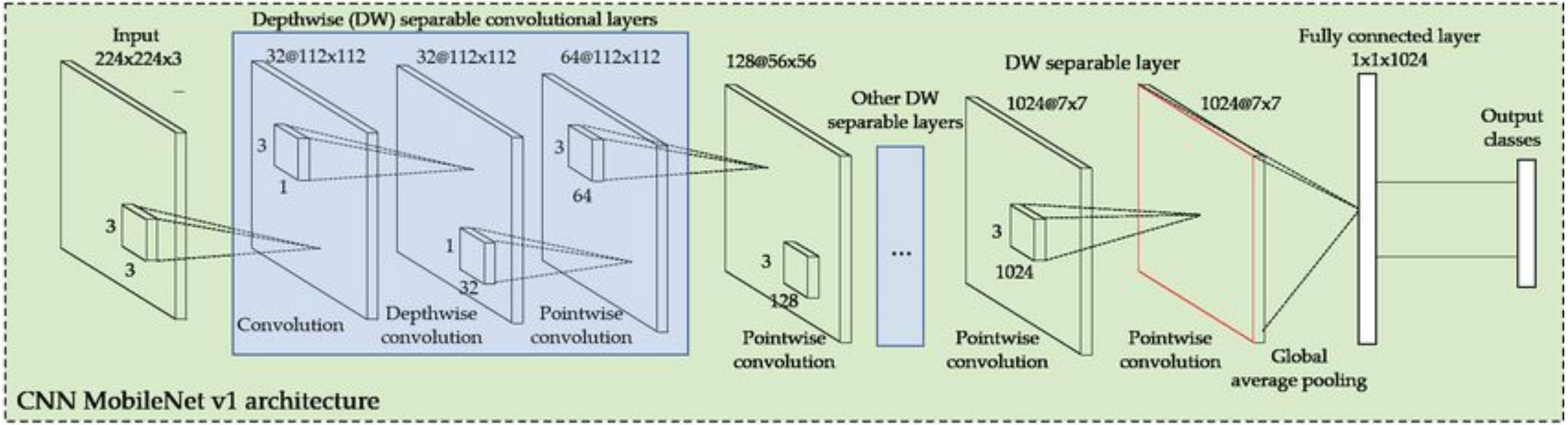
Train Test Split (80/20)

Creating training and testing set to estimate model robustness

03.

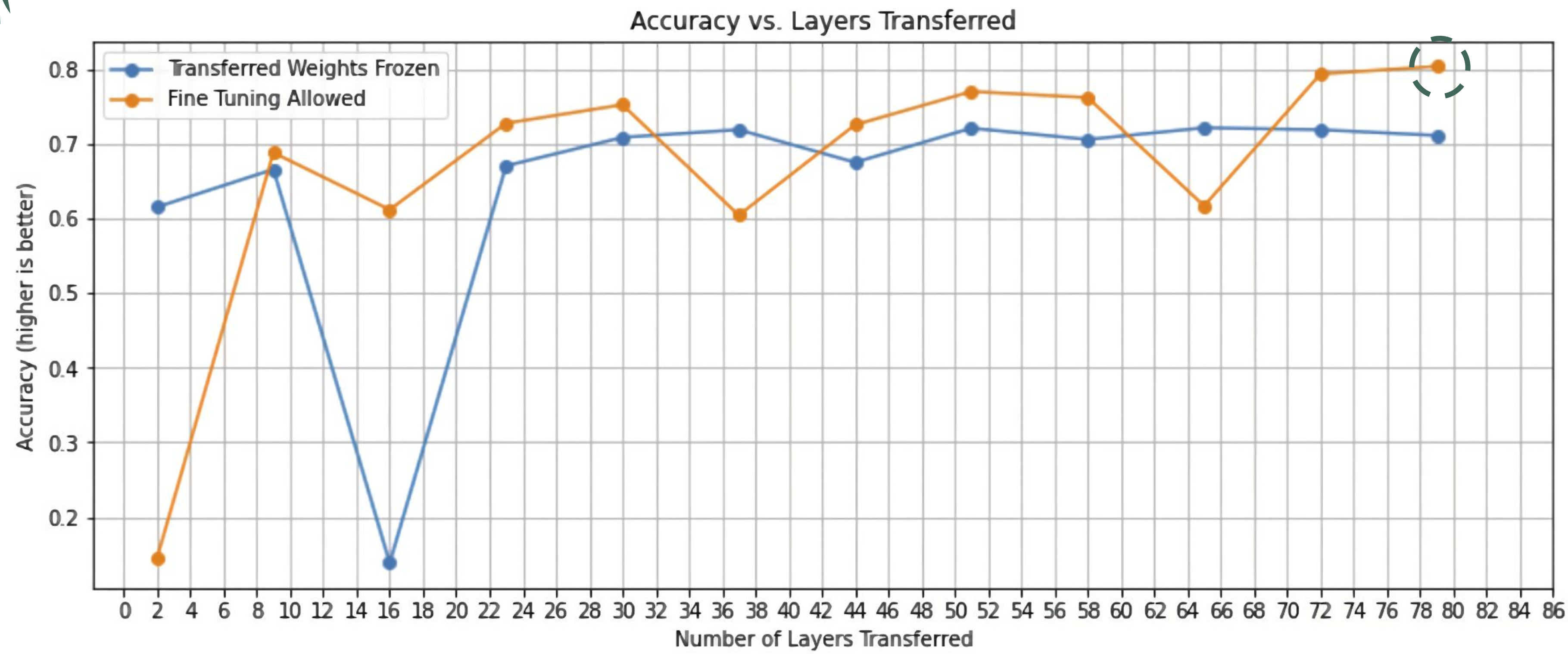
Transfer Learning

MobileNetV1 Architecture



- Pre-trained on ImageNet dataset
- Replaced output layer for 5 class classification problem
- Adjusted input layer to accommodate 128x128 dimension input images

Transfer Analysis



At the end of 10 epochs…

(base model)

Accuracy: 76.33%

Validation Accuracy: 71.68%

04.

Model Compression

Compression Techniques

Pruning (Magnitude + Iterative)

(10%, 30%, 50%, & 80%)

Quantization

(none, 8-bit, 16-bit)

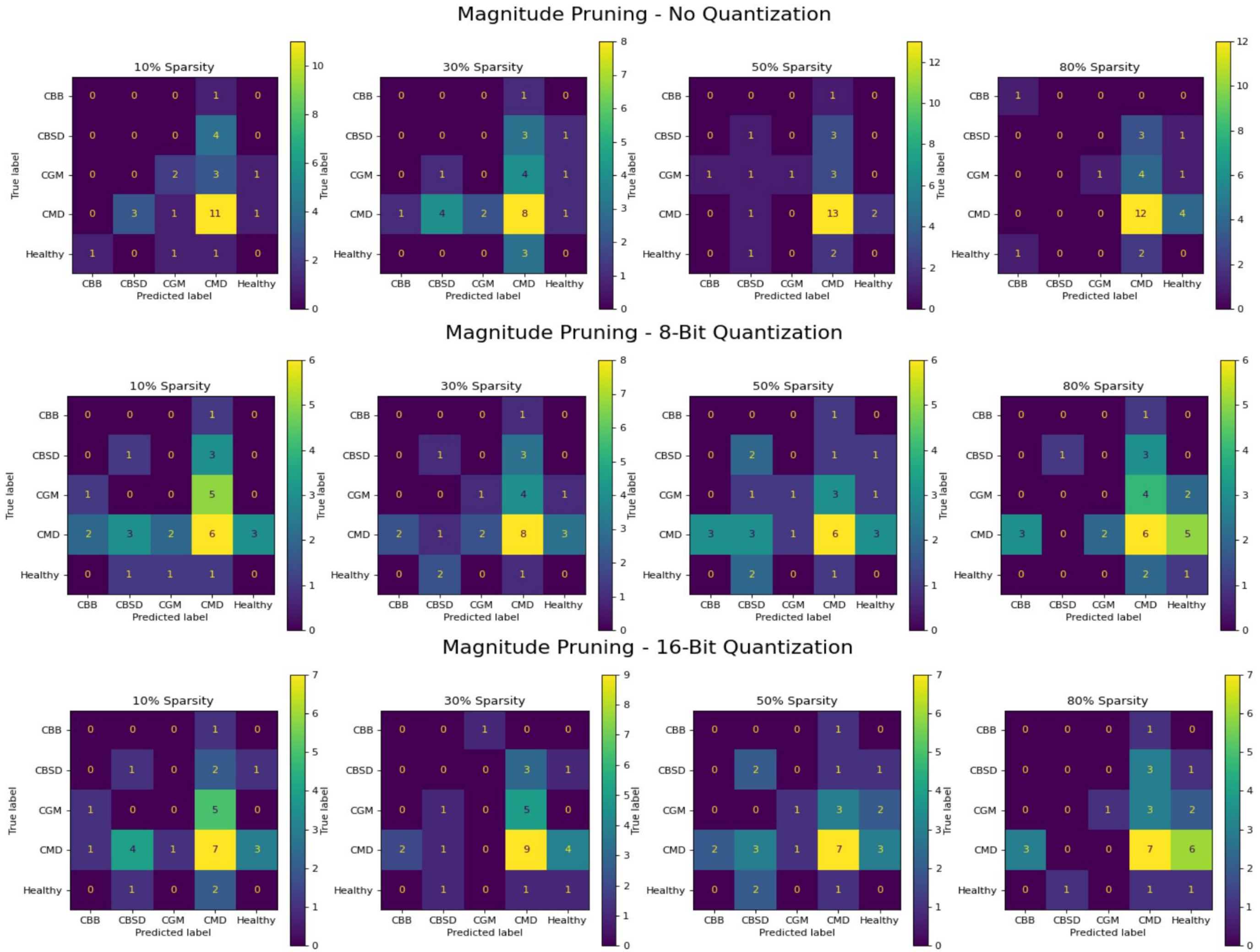
Effects of Pruning

Target Sparsity		Magnitude Pruning	Iterative Pruning
	10%	Acc: 77.66% **Val Acc: 71.45%** Avg: 74.56%	Acc: 79.01% Val Acc: 68.49% Avg: 73.78%
	30%	Acc: 79.21% Val Acc: 69.91% Avg: 74.56%	Acc: 77.73% Val Acc: 68.21% Avg: 72.97%
	50%	**Acc: 80.71%** Val Acc: 66.29% Avg: 73.50%	Acc: 79.11% **Val Acc: 71.77%** Avg: 75.44%
	80%	Acc: 77.89% Val Acc: 68.27% Avg: 73.08%	**Acc: 80.04%** Val Acc: 62.31% Avg: 71.18%

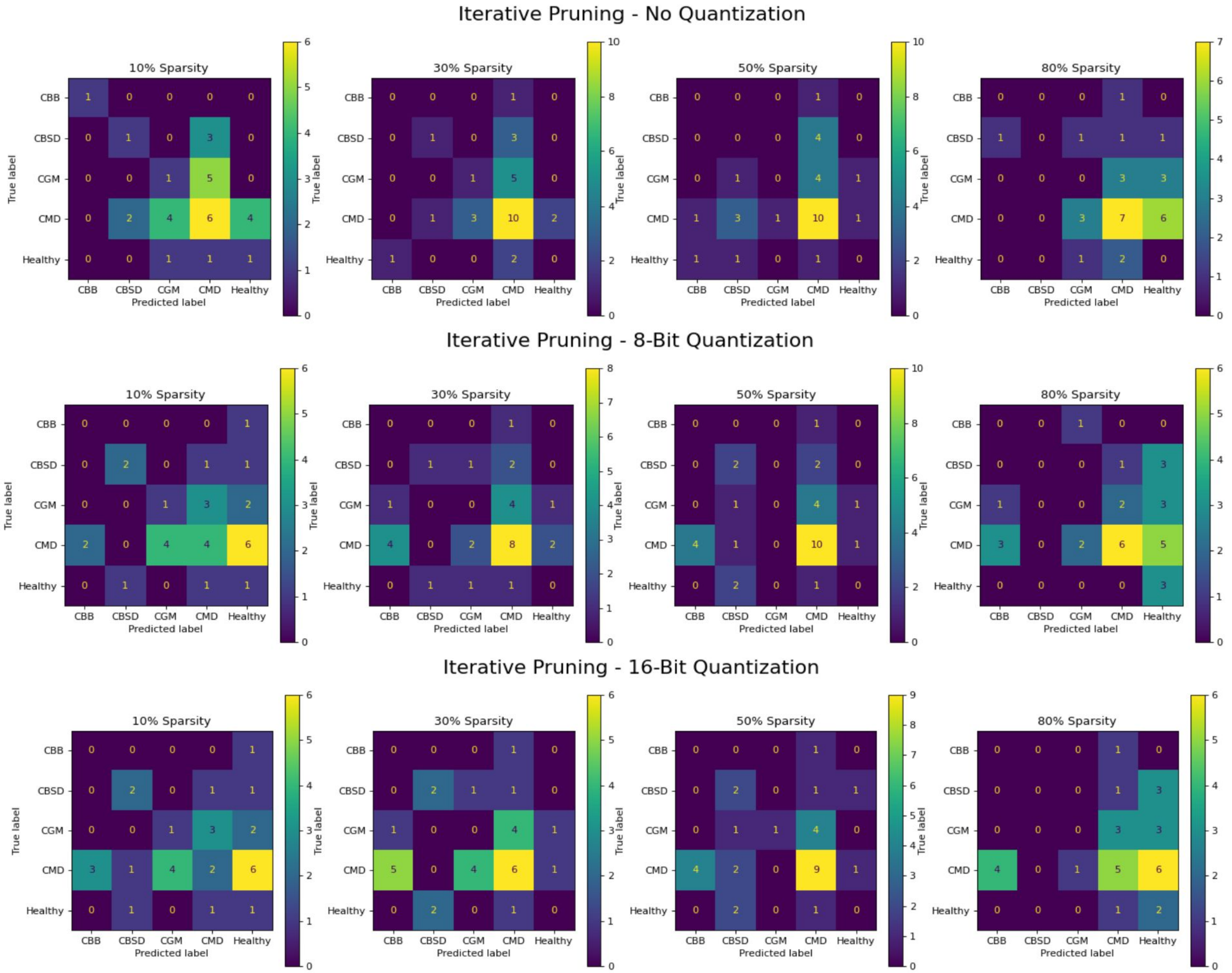
05.

Model Evaluation

Confusion Matrix (Magnitude Pruning + Quantization)

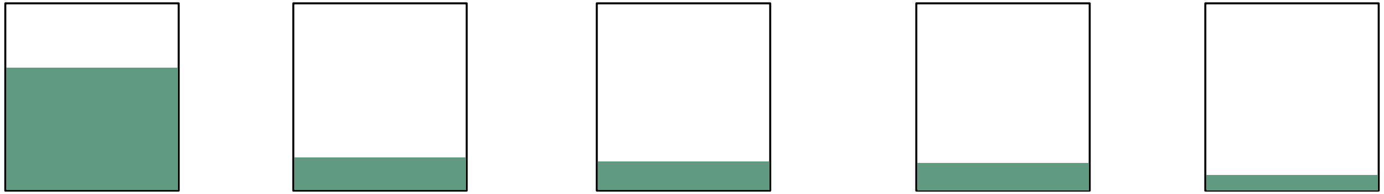

Magnitude Pruning - No Quantization
Magnitude Pruning - 8-Bit Quantization
Magnitude Pruning - 16-Bit Quantization
10% Sparsity
30% Sparsity
50% Sparsity
80% Sparsity
True label
Predicted label
CBB
CBSD
CGM
CMD
Healthy

Confusion Matrix (Iterative Pruning + Quantization)


Iterative Pruning - No Quantization
10% Sparsity
30% Sparsity
50% Sparsity
80% Sparsity
Iterative Pruning - 8-Bit Quantization
10% Sparsity
30% Sparsity
50% Sparsity
80% Sparsity
Iterative Pruning - 16-Bit Quantization
10% Sparsity
30% Sparsity
50% Sparsity
80% Sparsity
True label
Predicted label
CBB
CBSD
CGM
CMD
Healthy

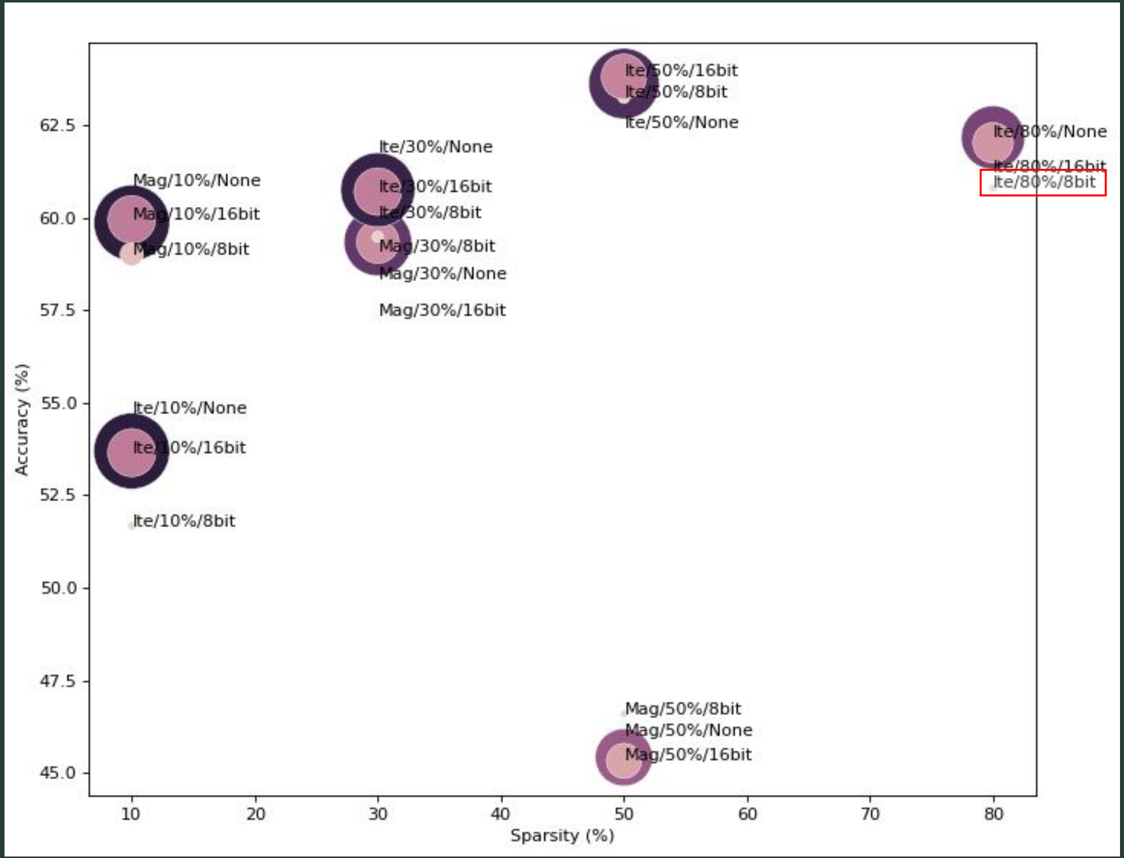
Distribution of Data Samples

62%	12%	11%	10%	5%



Cassava Mosaic Disease (CMD)	Healthy	Cassava Green Mottle (CGM)	Cassava Brown Streak Disease (CBSD)	Cassava Bacterial Blight (CBB)
13158 samples	2577 samples	2386 samples	2189 samples	1087 samples

Accuracy Trade-Off with Pruning & Model Size

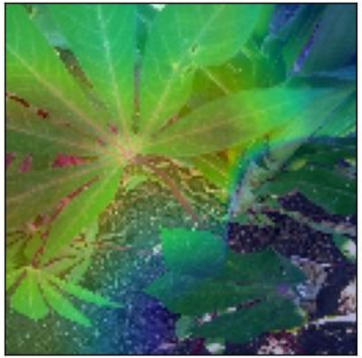


06.

Grad-CAM

(A visual interpretation technique to understand which aspects of the image the ML model is using to make its classification decision)

Grad-CAM - How It Works



Thank you for listening!

Questions?