From Roots to Results: Classifying Plants Using Artificial Intelligence

† † Amy N. Njenga, Peter Kidiga, Queen Kibegi, Keysha Mideva ‡ †

‡United States International University

Introduction

Scenario: AI-Based Crop Monitoring for Small Farmers

Small-scale farmers across Africa face major challenges in identifying crop diseases early, often leading to reduced yields and food insecurity.

- By training a model on images of healthy and diseased crops, we demonstrate a low-cost, accessible solution that empowers farmers to monitor their crops using basic technology like smartphones.

Methodology

Data Collection

Images of healthy and unhealthy crops were collected from reliable online agricultural sources. The dataset included clear images showing visible signs of disease or normal growth.

Model Deployment

A lightweight image classification model was built using MobileNetV2 through Google Teachable Machine. The model was trained to classify images into two categories: Healthy and Unhealthy.

Training and testing

The dataset was split into training and testing sets and was tested on new images to evaluate its accuracy in distinguishing between healthy and unhealthy crops.

Evaluation and Usability

Prior Work

Previous research has shown that AI models like CNNs (e.g., VGG, ResNet, MobileNet) can accurately detect crop diseases from leaf images.

However, most of these solutions are designed for commercial farms and rely on large datasets and high-end tools, making them less practical for small-scale farmers. Our project builds on this work by focusing on a lightweight, accessible model suitable for low-resource settings.

Modelling

Model Selection and Architecture

We used MobileNetV2, a lightweight convolutional neural network pretrained on ImageNet, as the feature extractor due to its efficiency and strong performance on small datasets. A custom classification head was added, including global average pooling, a dense layer with ReLU activation, dropout, and a sigmoid output for binary classification.

Training Strategy

Training was done in two phases: initially freezing the base model to train the classification head, then fine-tuning the top 50 layers with a reduced learning rate. We used focal loss with label smoothing to handle class imbalance and improve model generalization.

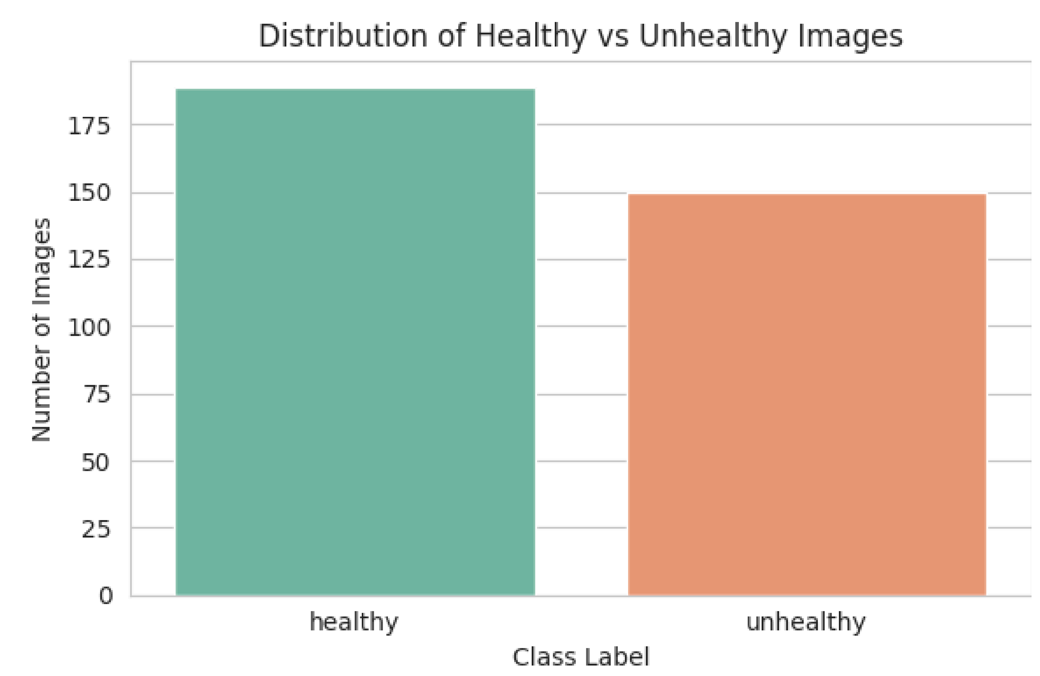
Regularization and Optimization

We applied data augmentation (rotation, shift, zoom, flip) to reduce overfitting and simulate a larger dataset. Additionally, EarlyStopping and ReduceLROnPlateau callbacks were used to prevent overfitting and dynamically adjust learning rates during training.

United States International University-Africa

Our Data

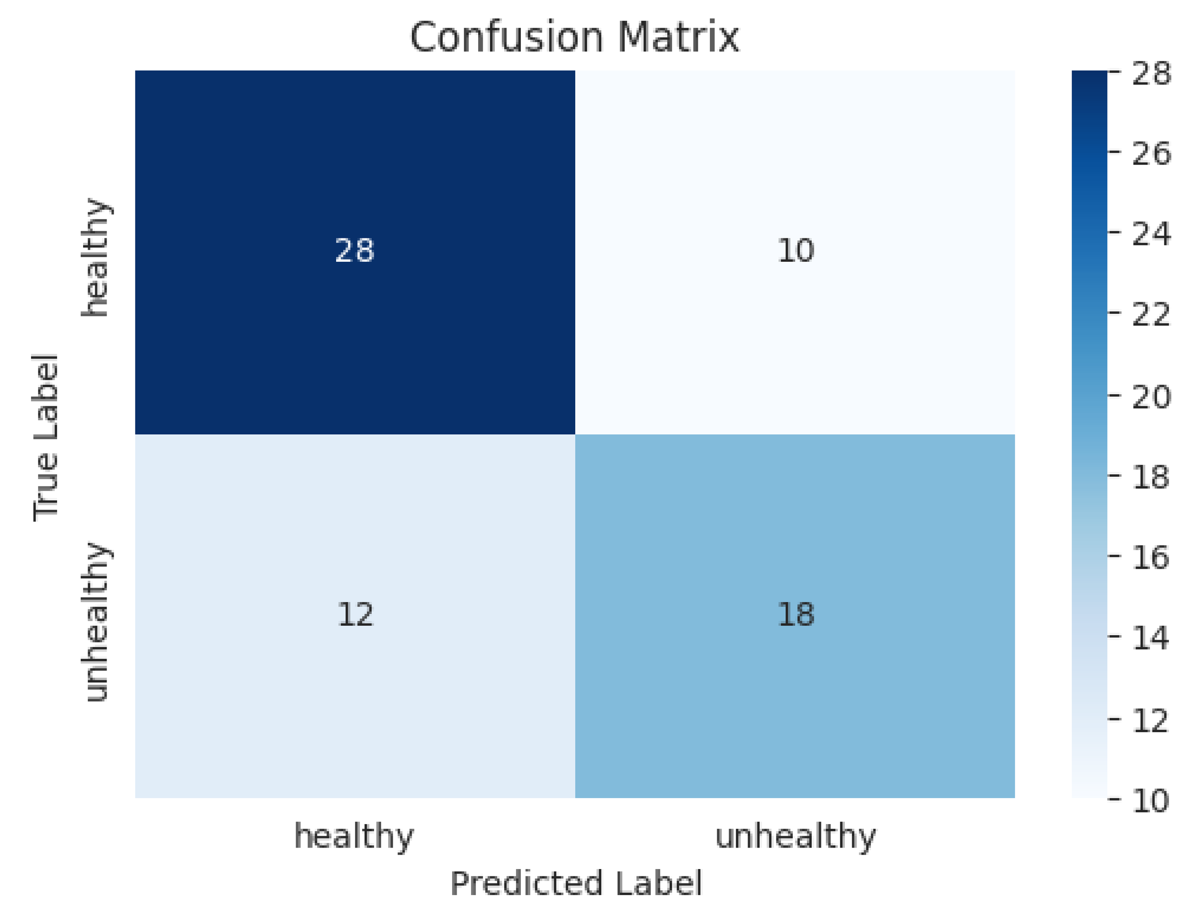
From our data collection, the healthy images were the majority



Distribution of collected images

Confusion Matrix

The confusion matrix shows that the model correctly identified 28 healthy plants **(true negatives)** and 18 unhealthy plants **(true positives)**. However, it also misclassified 10 healthy plants as unhealthy **(false positives)** and 12 unhealthy plants as healthy **(false negatives).** This means the model is slightly better at recognizing healthy plants, but still struggles to accurately detect all unhealthy cases.



Model Results

Challenges faced

- Anti-data mining policies: Many websites had restrictions in place to prevent scraping, which limited our ability to collect a diverse and large dataset. We had to look for alternative sources or manually download images.
- Bot detection: Some websites detected and blocked automated scraping tools. This interrupted the data collection process and required additional effort to bypass or switch methods.
- Unclear or low-quality images: A significant number of the images collected were blurry, poorly lit, or contained background noise, making them difficult for the model to learn from effectively.
- Incompatible formats: A few images were in unsupported or unusual formats (e.g., .webp, .tiff) that had to be converted to standard formats like .jpg or .png before being used in the model.

Conclusion

Despite challenges with data quality and size, the model achieved moderate accuracy and highlighted the potential of AI in agriculture. With a larger, cleaner dataset and further tuning, this approach could be scaled to support early detection of crop diseases for small-scale farmers.