Plant Leaf Disease Classifier

1. Problem Statement and Applications

Today, farmers face huge losses in their crop produce due to diseases that are tedious to identify manually from individual plants. To mitigate this, we propose a system that accurately identifies the diseases present in the crops based on images of their plant parts. By swiftly recognizing viral diseases, bacterial blights, and fungal infections, the farmers are enabled to take prompt actions to promote recovery among the plants, thus maximizing product outcomes. However, choosing the appropriate model to arrive at the maximum accuracy without compromising the time taken to classify the disease of the plant requires the comparison of available models with varying numbers of train and test data. Apart from the models being trained from scratch, some models trained using transfer learning techniques will be included in the comparison to fortify the final models, hence emphasizing accuracy without compromising on the near real-time speeds of predictions.

2. Dataset

A thorough analysis was done on several datasets and from that We will be using 3 datasets as mentioned in below table:

Dataset details			
Dataset Name	Number of Images	Number of classes	Total Dataset Size
Plant Village [1]	61,000	38	5GB
Corn Maize Leaf Disease [2] [3] [4]	3,500	4	200MB
Tomato Leaf Disease [5]	16,000	10	1GB

3. Possible Methodology

With the dataset prepared, we will be implementing preprocessing methods on the datasets including: Limiting Classes - Classes from the Plant Village dataset are reduced from 38 to 20 classes. Data Augmentation - Rotation, Flipping, and scaling to increase the variability in the dataset. Image resizing - All images in the datasets resized to a uniform value to ensure input consistency. Noise reduction - since some images are noisy, denoising techniques are employed to reduce significant noise. Image normalization - Scaling pixels to [0,1] to facilitate faster convergence during training. Partitioning - All datasets are split into Train, Validate and Test subsets for further processing.

4. Model Architecture

For plant disease classification, we employ three advanced convolutional neural network (CNN) architectures: VGG or ResNet50, MobileNetV2, and InceptionV3 (also known as GoogLeNet). During training, specific convolution layers are chosen for each model for feature extraction. This step helps reduce dimensionality and focus on the most informative aspects of the data. Choosing the layers acts as a hyperparameter during the training process.

5. Training Process and Evaluation

The model training process for effective CNN architectures in plant disease classification starts with dataset preprocessing, including loading, augmenting, and normalizing pixel values. The dataset is then split into training, validation, and test sets to ensure unbiased evaluation. Selected CNN architectures—MobileNetV2, InceptionV3, and VGG (or ResNet50) —are initialized and the data is passed for feature extraction. Loss is calculated for the predicted value with respect to the true value and the parameters are optimized using the algorithms, such as the Adam optimizer. For every epoch, the model is evaluated with the unseen data (from validation set) and the loss and accuracy are noted. Our training process involves calculating both the training and the validation losses and accuracy for each epoch. This helps in finding the area at which the model performs better. Finally, after reviewing the plots, by doing some hyperparameter tuning, the best model is chosen.

References

[1] Plant Village Dataset from site https://data.mendeley.com/datasets/tywbtsjrjv/1 1

[2] Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N. (2020). PlantDoc: a dataset for visual plant disease detection. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD (pp. 249-253). 1

[3] J, Arun Pandian, Gopal, Geetharamani. (2019). Identification of Plant Leaf Diseases Using a 9-layer Deep Convolutional Neural Network. [Data set]. Mendeley Data. https://doi.org/10.17632/tywbtsjrjv.1 1

[4] Corn Maize leaf Disease Dataset from site https://www.kaggle.com/datasets/smaranjitghose/corn-or-maize-leaf-disease-dataset 1

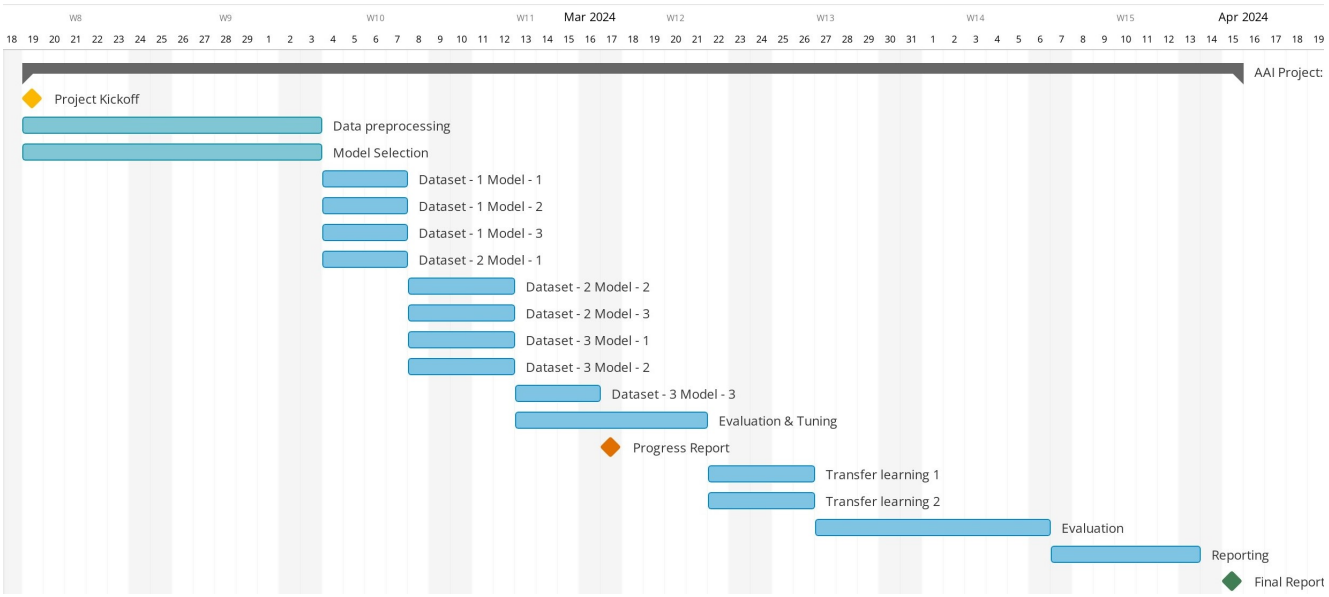

Figure 1. Project delivery timeline

[5] Tomato Leaf Disease Dataset from site https://www.kaggle.com/datasets/ashishmotwani/tomato 1