Project Title: Disease Detection in Crop Plant-'Cassava Plant' Using Image Classification

Summary

Cassava is an important staple crop in many tropical countries, providing millions of people with critical carbohydrates. However, a number of illnesses, most notably Cassava Mosaic Disease (CMD), Cassava Green Mottle (CGM), Cassava Brown Streak Disease (CBSD), and Cassava Bacterial Blight (CBB), threaten its output. Effective management and mitigation of many diseases depend on early and precise detection. Creating a machine learning model that can categorise photos of cassava plants into five groups—CBB, CBSD, CGM, CMD, and Healthy—is the goal of this project. Utilizing a labeled dataset of cassava images, the model will assist in the timely identification of diseases, thereby supporting farmers and agricultural stakeholders in disease management efforts.

Research Question

How can machine learning techniques be applied to accurately classify images of cassava plants into specific disease categories, and what is the most effective model architecture for this task?

Aim and Objective

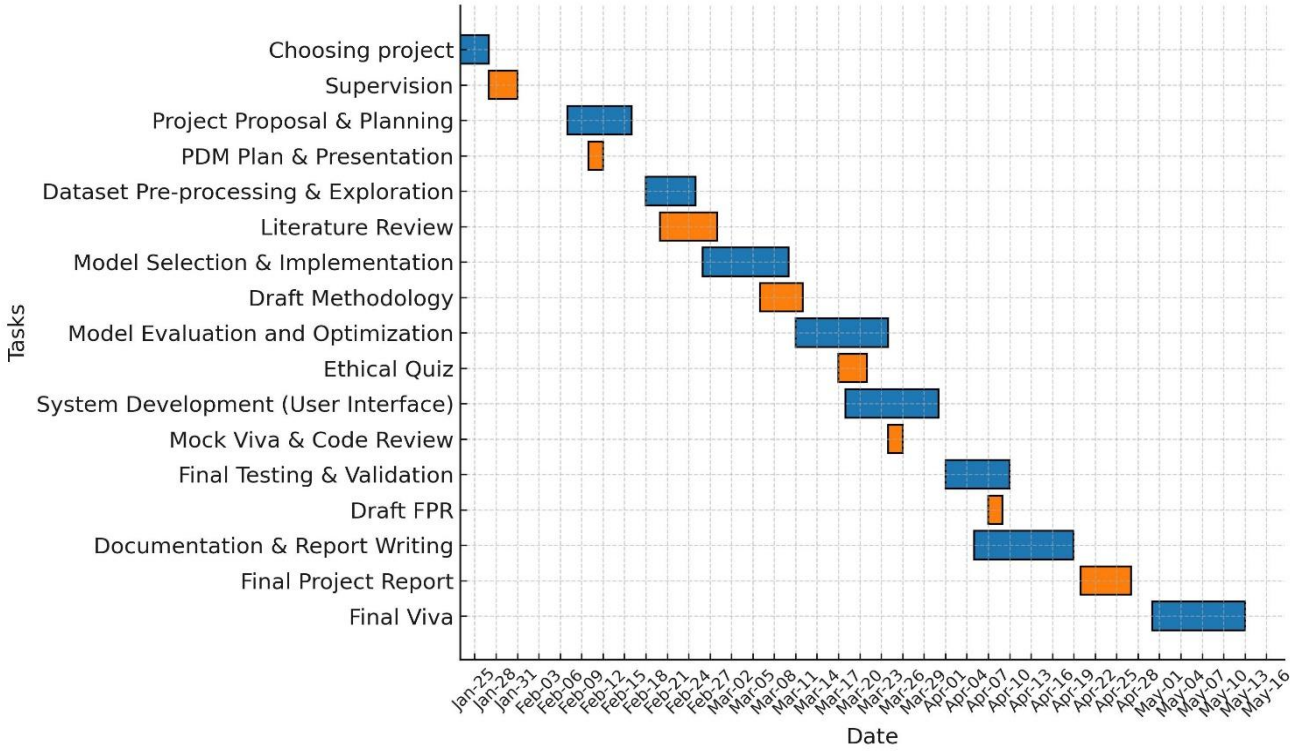
- To automatically classify cassava plant diseases using a machine learning model
- To comprehend class distributions and image quality, the dataset must be preprocessed and analysed.
- To evaluate various transfer learning techniques, such as ResNet or EfficientNet, and deep learning models (CNNs) for disease categorisation.
- To create an intuitive system for image uploads and plant disease detection.
- To use F1-score, precision, and recall metrics to assess the model's correctness.

Task List

Task	**Description**	**Start Date**	**End Date**
Choosing project	Select the project topic, define the problem statement, and outline the initial idea.	23-Jan	27-Jan
Supervision	Initial meeting with supervision to discuss project scope and expectation	27- Jan	31- Jan
Project Proposal & Planning	Develop a detailed project proposal, including objectives, research question, and methodology.	07-Feb	16-Feb
PDM Plan & Presentation	Prepare a Project Data Management (PDM) plan and present it for review.	10-Feb	12-Feb
Dataset Pre-processing & Exploration	Load the dataset, clean the data, analyze class distribution, and apply augmentation.	18-Feb	25-Feb
Literature Review	Study existing research papers on cassava disease detection and deep learning models.	20-Feb	28-Feb
Model Selection & Implementation	Train different deep learning models (CNNs, Transfer Learning, etc.) for disease classification.	26-Feb	10-Mar
Draft Methodology	Document the methodology used in the project,	06-Mar	12-Mar

	including dataset processing and model selection.		
Model Evaluation and Optimization	Tune hyperparameters, validate results, and optimize the model for better accuracy.	11-Mar	24-Mar
Ethical Quiz	Ethical quiz for research ethics and guidelines	17- Mar	21-Mar
System Development (User Interface)	Develop a simple web or desktop interface for image upload and disease classification.	18-Mar	31-Mar
Mock Viva & Code Review	Prepare for viva presentation and conduct a peer review of the code and findings.	24-Mar	26-Mar
Final Testing & Validation	Conduct real-world testing of the model and validate its performance with new images.	01-Apr	10-Apr
Draft FPR	Prepare a draft of the Final Project Report for initial review and feedback.	07-Apr	09-Apr
Documentation & Report Writing	Write technical documentation, project report, and ensure all details are well-documented.	05-Apr	19-Apr
Final Project Report	Finalize and submit the complete project report before the deadline.	20-Apr	27-Apr
Final Viva	Present the project findings and outcomes in the final viva assessment.	30-Apr	13-May

Gantt Chart



Overview of the dataset

This project's dataset, "Crop Diseases Classification Dataset," was obtained from Kaggle:https://www.kaggle.com/datasets/mexwell/crop-diseases-classification/data.Thousands of labelled photos of cassava leaves divided into five classes—CBB, CBSD, CGM, CMD, and Healthy—are included in the dataset. It comes with a JSON file with labels and a 'train_images' folder. With pictures in.jpg and.png formats and a CSV file that links image filenames to the appropriate disease diagnoses, the collection is roughly 5–10 GB in size.

This dataset is sourced from "PlantVillage," an open-access resource created to facilitate machine learning-based mobile disease diagnosis. More than 54,000 photos of both healthy and sick plant leaves from different kinds are included. https://www.tensorflow.org/datasets/catalog/plant_village

David P. Hughes and Marcel Salathe founded the PlantVillage project, which uses smartphones to assist farmers in lowering crop losses due to disease. In order to facilitate machine learning-based plant disease identification and promote global food security, it offers a sizable image archive.

In conclusion, the PlantVillage project, which was established to facilitate the development of mobile disease diagnostics and assist agricultural communities globally, is the source of the "Crop Diseases Classification" dataset on Kaggle.

- To maintain security, code and data will be backed up twice a week on One Drive and Github. For assessment purposes, only authorised personnel—including staff and markers—will be able to access the repository.

Ethical requirements

Since the dataset is open-source and freely accessible for research and teaching purposes, it is exempt from GDPR regulations because it does not contain any sensitive or personal data. The project conforms with ethical norms, guaranteeing the responsible use of data, and neither the collecting nor processing of data involves human people. The PlantVillage project is the source of the dataset, guaranteeing ethical data collecting. Use is permitted for educational and research purposes.

GitHub Link: https://github.com/Akshara-bs/Final-Project.git

ReadMe File Overview

The goal of the project, specifics of the dataset, and setup guidelines will all be briefly described in the GitHub ReadMe. It will outline the machine learning models that are employed, provide instructions for users to run and test the model, and compile evaluation metrics. It will also provide contact information for support and guidelines for contributions.

References

Ayu, H. R., Surtono, A., & Apriyanto, D. K. (2021). *Deep Learning for Detection Cassava Leaf Disease*. *Journal of Physics: Conference Series*, 1751(1), 012072.

Mwebaze, E., Gebru, T., Frome, A., Nsumba, S., & Tusubira, J. (2019). *iCassava 2019 Fine-Grained Visual Categorization Challenge*. *arXiv preprint arXiv:1908.02900*. https://arxiv.org/abs/1908.02900

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. (2017). *Deep Learning for Image-Based Cassava Disease Detection*. *Frontiers in Plant Science*, 8, 1852. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2017.01852/full