PROJECT AND DATA MANAGEMENT PLAN (PDM)

Research Supervisor: William Bate
Student ID: 23013682
Researcher: Covenant Oludapomola Ojo
Date: 12th Feb 2025

1. PROJECT OVERVIEW: ENHANCING PLANT DISEASE DETECTION WITH DEEP LEARNING MODELS: A CASE STUDY ON MAIZE AND CASSAVA SPECIES

1.0 PROJECT TOPIC SUMMARY AND BACKGROUND

Reducing crop loss and increasing agricultural productivity depend on the early diagnosis of plant diseases. Conventional techniques, which rely on human inspection, are laborious and prone to mistakes. Convolutional Neural Networks (CNNs), Vision Transformers, and EfficientNet are examples of advanced deep learning models that have demonstrated great promise in automating the classification of plant diseases. Using a Mendeley dataset, this project aims to develop a high-performance deep learning model for crop disease identification. The dataset includes 24,881 raw pictures from 22 classes and was submitted on April 26, 2023. I'll be concentrating on maize and cassava crops.

1.1 RESEARCH AIMS AND OBJECTIVES

Aim: To use deep learning models and real-time image-based disease diagnosis to improve the precision and effectiveness of plant disease detection in cassava and maize.

Objectives:

1. Conduct a comprehensive literature review on deep learning applications in plant disease detection.
2. Preprocess and augment the dataset to develop and improve model generalization and robustness.
3. Train and compare multiple deep learning architectures, including CNNs, EfficientNet, and Vision Transformers.
4. Implement ensemble learning techniques to improve classification accuracy.
5. Develop a real-time user-input system to detect plant diseases instantly.
6. Evaluate model functionality and performance using standard metrics which are accuracy, precision, recall, F1-score and compare results with existing studies.

1.2 RESEARCH QUESTIONS

1. How effective are deep learning models in detecting diseases in maize and cassava species?
2. Which deep learning architectures provide the best accuracy for plant disease detection?
3. How can real-time user input enhance the efficiency of a crop disease detection system?
4. What impact do transfer learning and ensemble learning techniques have on model accuracy?

1.3 METHODOLOGY

1.3.1 Data Collection and Preprocessing

24,881 images from Mendeley dataset, covering twenty-two classes of four crops (Cashew, Cassava, Maize, and Tomato) but focusing on Cassava and Maize. Image resizing, normalization, data augmentation (rotation, flipping, brightness adjustment) preprocessing techniques will be done to improve model robustness.

1.3.2 Model Development and Training

- Exploratory Data Analysis and Image Preprocessing/Augmentation
- **Baseline Models:** CNNs, EfficientNet, Vision Transformers.
- **Optimization:** Hyperparameter tuning, transfer learning, ensemble learning.
- **Training Setup:** GPU-accelerated training using TF/Keras, Google Collab.
- **Deployment and Predictions:** On the test set after it has been trained.
- Documenting the code and Results

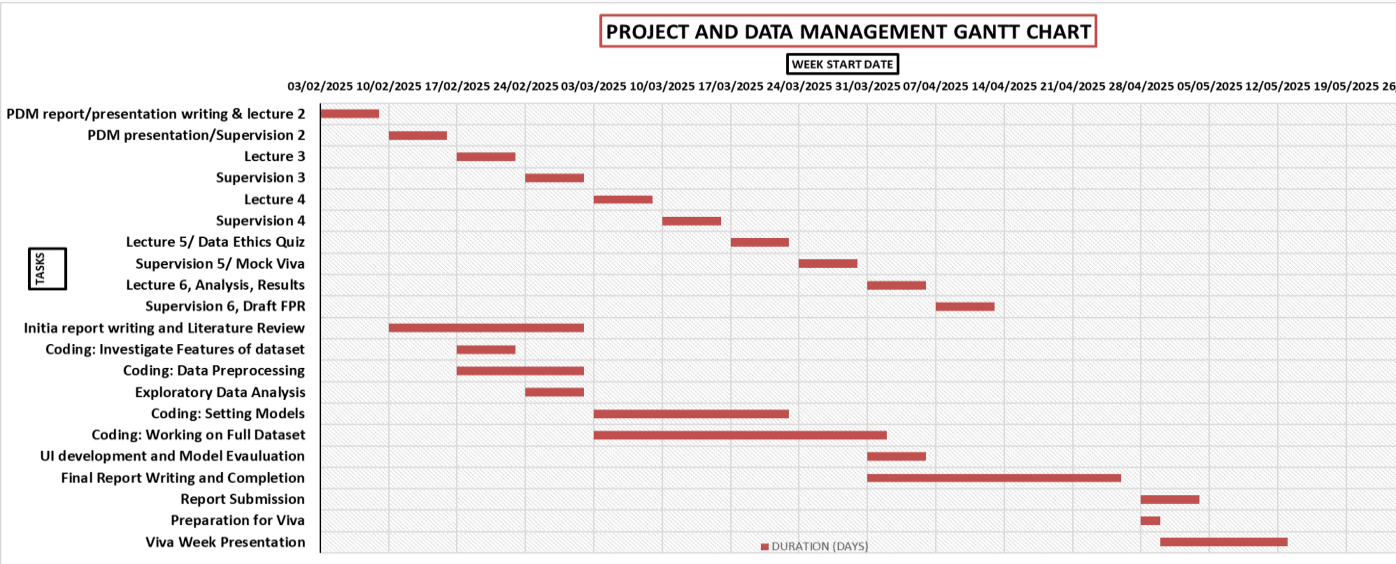
1.3.3 System Implementation: A web-based/model user input allowing users to upload plant images for real-time disease detection. Metrics, for example, Precision, Recall, F1-Score and Comparison with existing benchmark models in the literature review.

1.4 REFERENCE LIST

1. Bourzig, Dounia Kawther Dihya and Abed, Mansour and Merah, Mostefa, Enhancing Diseases Classification in Maize and Cassava Crops Using Pre-Trained CNN Models and Layer Thawing Strategy. Available at SSRN: https://ssrn.com/abstract=4903435 or http://dx.doi.org/10.2139/ssrn.4903435
2. Mensah Kwabena, P., Akoto-Adjepong, V., Adu, K., Ayidzoe, M.A., Asare Bediako, E., Nyarko-Boateng, O., Boateng, S., Donkor, E.F., Bawah, F.U., Awarayi, N.S., Nimbe, P., Nti, I.K., Abdulai, M., Adjei, R., & Opoku, M. (2023) ‘Dataset for Crop Pest and Disease Detection’, Mendeley Data, V1. Available at: https://doi.org/10.17632/bwh3zbpkpv.1 [Accessed 5 Feb. 2025].
3. Mohanty, S.P., Hughes, D.P., & Salathé, M. (2016) ‘Using deep learning for image-based plant disease detection’, Frontiers in Plant Science, 7, p.1419. Available at: https://doi.org/10.3389/fpls.2016.01419 [Accessed 5 Feb. 2025].
4. ResearchGate (2023) ‘CCMT Dataset for Crop Pest and Disease Detection’. Available at: https://www.researchgate.net/publication/371510439_CCMT_Dataset_for_Crop_Pest_and_Disease_Detection/citations [Accessed 5 Feb. 2025].

2. PROJECT PLAN/ TIMELINE

TASK	WEEK START DATE	WEEK END DATE	DURATION (DAYS)
PDM report/presentation writing & lecture 2	03/02/2025	09/02/2025	6
PDM presentation/Supervision 2	10/02/2025	16/02/2025	6
Lecture 3	17/02/2025	23/02/2025	6
Supervision 3	24/02/2025	02/03/2025	6
Lecture 4	03/03/2025	09/03/2025	6
Supervision 4	10/03/2025	16/03/2025	6
Lecture 5/ Data Ethics Quiz	17/03/2025	23/03/2025	6
Supervision 5/ Mock Viva	24/03/2025	30/03/2025	6
Lecture 6, Analysis, Results	31/03/2025	06/04/2025	6
Supervision 6, Draft FPR	07/04/2025	13/04/2025	6
Initia report writing and Literature Review	10/02/2025	02/03/2025	20
Coding: Investigate Features of dataset	17/02/2025	23/02/2025	6
Coding: Data Preprocessing	17/02/2025	02/03/2025	13
Exploratory Data Analysis	24/02/2025	02/03/2025	6
Coding: Setting Models	03/03/2025	23/03/2025	20
Coding: Working on Full Dataset	03/03/2025	02/04/2025	30
UI development and Model Evauluation	31/03/2025	06/04/2025	6
Final Report Writing and Completion	31/03/2025	26/04/2025	26
Report Submission	28/04/2025	04/05/2025	6
Preparation for Viva	28/04/2025	30/04/2025	2
Viva Week Presentation	30/04/2025	13/05/2025	13



3. DATA MANAGEMENT PLAN

3.0 OVERVIEW OF DATASET

The dataset was sourced from [Mendeley dataset repository](), published, **26 April 2023**. It consists of **24,881 raw images** and **102976 Augmented images** across **twenty-two classes** of crop diseases and healthy conditions, I will be working on the raw images for the purpose of originality of the research project and for more preprocessing techniques to be done. This data was collected from **local farms in Ghana** and validated by **expert plant virologists** to ensure accuracy and reliability. It aims to address challenges faced by farmers in developing countries, such as pest and disease infestations, limited technological knowledge, and inadequate storage facilities. In the raw images, I will be focusing on Cassava which has 7,508 images across 5 classes (bacterial blight, brown spot, green mite, healthy, and mosaic) and Maize (5,389 images) across 7 classes (fall armyworm, grasshopper, healthy, leaf beetle, leaf blight, leaf spot, and streak virus).

3.1 DATA COLLECTION

- **Source**: Mendeley dataset (Link), the website address is from Mendeley.
- **Dataset Size**: 24,881 raw images with twenty-two classes across four crops. Focus will be on cassava and maize.

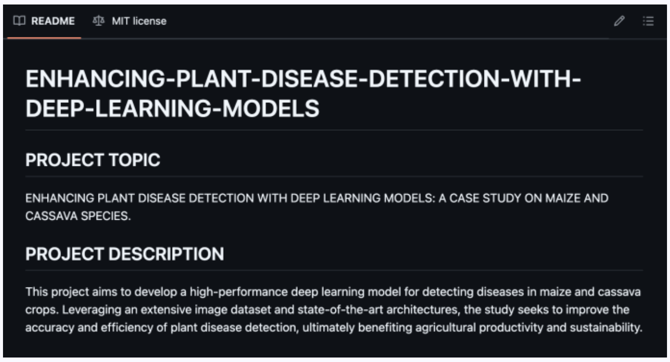
3.2 METADATA: The format is an image dataset with a download size of 7.86 GB, approximately 6 GB for raw images. The code files for data preprocessing, model development, and evaluation are expected to occupy 50–100 MB, depending on the libraries and scripts included.

3.3 DOCUMENT CONTROL AND BACKUP: The primary storage will be google Drive and GitHub Repository **PROJECT GITHUB-LINK** which would be backed up weekly through my University one drive account. I will also be working on Google Collab cloud platform. File naming will be "Covenant MSC project report."

Folder Structure:

- /data/raw: Original dataset from Mendeley.
- /data/pre-processed: Augmented dataset.
- /Models: Trained models and checkpoints.
- /Results: Evaluation metrics and logs.

3.4 README FILE:


README MIT license

ENHANCING-PLANT-DISEASE-DETECTION-WITH-DEEP-LEARNING-MODELS

PROJECT TOPIC

ENHANCING PLANT DISEASE DETECTION WITH DEEP LEARNING MODELS: A CASE STUDY ON MAIZE AND CASSAVA SPECIES.

PROJECT DESCRIPTION

This project aims to develop a high-performance deep learning model for detecting diseases in maize and cassava crops. Leveraging an extensive image dataset and state-of-the-art architectures, the study seeks to improve the accuracy and efficiency of plant disease detection, ultimately benefiting agricultural productivity and sustainability.

3.5 DATA SECURITY AND STORAGE: The usage will comply with the terms outlined by Mendeley. Usage of google collab platform for code and documentation, ResearchGate (**Research Gate**) for final results and publication. The access level is made public for non-commercial use, I will also include a README file explaining data structure, preprocessing steps, and model evaluation. Retain pre-processed dataset and model for five years post-project. Raw data can be deleted after publication and confirmation of reproducibility.

3.6 DATA ANALYSIS TOOLS: Python programming libraries which entails TensorFlow/Keras for deep learning, OpenCV for image processing**,** Google Collab (GPU) for hardware.

3.7 ETHICAL REQUIREMENTS: The dataset is public; no sensitive or personal data is involved that will warrant seeking ethics consideration.