Plant Disease Classification Project Report

Title Page

Project Name: Plant Disease Classification Using CNN

Student Names & IDs:

- Amr Ahmed Mohamed – 2227238
- Karim Waheed Saber – 212728
- Ibrahim Alaa El-Sayed Atiya – 2227013
- Moamen Hany Afifi – 2227032
- Youssef Mohamed Samir – 2227032

Submission Date: 14 December 2025
Course Name & Instructor Names: CS417 – Dr. Mahmoud Esmat, Dr.Omar Morad

1. Introduction

Problem:
Plant diseases cause significant agricultural losses. Detecting them early is crucial to prevent crop damage and improve yield.

Importance:
Automatic disease detection using AI supports precision agriculture, enabling timely intervention, reducing losses, and decreasing the need for manual inspection.

2. Dataset

Source: New Plant Diseases Dataset – Kaggle
(https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset)

Number of Classes: 38

Number of Images per Class:

- Training: 70,295 images
- Validation: 14,069 images
- Test: 3,803 images

Preprocessing Steps:

- Resize images to **128×128**
- Convert to RGB
- Normalize pixel values to **[0,1]**
- Apply **CLAHE** for leaf detail enhancement

Data Augmentation (Training Only):

- Rotation
- Zoom
- Horizontal flip
- Width & height shifts
- Brightness adjustment

Class Weights: Calculated to handle imbalance

Visualization: Pie charts confirm dataset balance

---

3. Methodology

Model Architecture (CNN):

- **Feature Extraction:**
  - Conv2D (32 filters) + MaxPooling
  - Conv2D (64 filters) + MaxPooling
  - Conv2D (128 filters) + MaxPooling
- **Classification:**
  - Flatten → Dense(256, ReLU) → Dropout(0.5) → Dense(38, Softmax)

Training Procedure:

- Optimizer: **Adam**
- Loss Function: **Categorical Crossentropy**
- Metrics: **Accuracy**

- Epochs: up to 40
- Callbacks: EarlyStopping (patience=5), ModelCheckpoint (best val loss)
- Class weights applied

Hyperparameters:

- Batch size: 32
- Learning rate: default Adam

---

4. Results

Test Performance:

Metric	Value
Test Accuracy	~91%
Test Loss	~0.32

Confusion Matrix: Strong diagonal dominance; minor confusion among visually similar classes (e.g., Tomato diseases)

Training Curves: Accuracy and loss curves indicate steady learning without significant overfitting

F1-scores: Computed per class (add table/plot in full report)

---

5. Discussion

What Worked Well:

- CNN successfully captured leaf disease features
- Data augmentation improved generalization
- Class weighting balanced underrepresented classes

What Failed / Challenges:

- Minor misclassification among visually similar classes
- Dataset quality variations may affect model accuracy

Limitations:

- Some visually similar diseases are difficult to distinguish
- Model trained only on 38 classes; new diseases would require retraining

Future Work:

- Explore deeper architectures (ResNet, EfficientNet)
- Deploy as a web or mobile application for real-time detection
- Expand dataset for more plant species and diseases

---

6. Conclusion

The project successfully implemented a CNN for plant disease detection with **~91% test accuracy**.
The system can classify new leaf images and support early intervention in agriculture.

---

7. References

- New Plant Diseases Dataset – Kaggle (https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset)
- Google Colab Notebook (https://colab.research.google.com/drive/19mwVk_xIheZnvVrh1t8xtaccOT-0vx4m?authuser=0#scrollTo=zfCR2I65nFm2) – Used for training and testing
- Relevant CNN literature and tutorials
