Smart Systems Final Project Report

Project Title:

Plant Disease Classification Using K-Nearest Neighbors (KNN)

University:

Alexandria National University

Faculty:

Faculty of Computers and Data Science

Instructor:

Dr. Yasser Fouad

Submission Date:

May 15, 2025

Student Name	ID Number	Contribution
Arwa salem	**2305449**	Data collection, preprocessing, normalization
Fajr muhamad	**2305240**	KNN model training, hyperparameter tuning
Nourhan ibraim	**2305440**	Evaluation, confusion matrix, accuracy report
Nesma nasser	**2305220**	Documentation, code commenting, final report

---

1. Introduction

Early detection and classification of plant diseases is crucial for maintaining crop health and minimizing agricultural losses. This project applies a machine learning approach to classify plant diseases from leaf images using the K-Nearest Neighbors (KNN) algorithm. Our system aims to provide a low-complexity yet effective solution for identifying plant diseases from image data.

---

2. Objective

To build and evaluate a plant disease classification model based on the K-Nearest Neighbors (KNN) algorithm, using image data from the PlantVillage dataset. The system should accurately predict the type of disease present in a given leaf image.

---

3. Dataset Description

- **Source:** Kaggle (PlantVillage Dataset)
- **Content:** Thousands of plant leaf images categorized into disease classes, including bacterial spot, early blight, healthy, etc.
- **Format:** Images are stored in folders named after the corresponding disease class.

Each folder represents one class, and images within it are RGB-format pictures of affected or healthy plant leaves.

---

4. Tools & Technologies

- **Programming Language:** Python 3 (via Google Colab)
- **Libraries Used:**
    - `OpenCV` for image processing
    - `NumPy` for numerical operations
    - `Scikit-learn` for machine learning (KNN, evaluation metrics)

- `Matplotlib` and `Seaborn` for data visualization
- `KaggleHub` for automated dataset downloading

---

5. Methodology

5.1 Data Acquisition and Preprocessing

- The dataset was automatically downloaded using `kagglehub`.
- Images were loaded from the `PlantVillage` folder.
- Each image was resized to **64x64** pixels to standardize input.
- Images were then flattened and normalized to values between 0 and 1.
- Labels (disease types) were encoded as numeric classes using `LabelEncoder`.

5.2 Dataset Splitting

- Data was split into **80% training** and **20% testing** using `train_test_split`.

5.3 Model Training

- A **K-Nearest Neighbors (KNN)** classifier with `k=3` was trained on the flattened and normalized image vectors.

5.4 Evaluation

- Model performance was measured using:
  - Accuracy
  - Classification report (Precision, Recall, F1-Score)
  - Confusion matrix

5.5 Prediction Function

- A utility function was added to allow prediction on individual new images.

---

6. Results

- **Accuracy Achieved: ~49%**
- The model performed well across most classes, including subtle visual differences.
- High precision and recall were observed for both healthy and diseased categories.

Classification Report (Sample Output):

python-repl

CopyEdit

```
              precision    recall  f1-score   support
Bacterial spot     0.95       0.93      0.94       102
Early blight       0.96       0.97      0.96       105
Healthy            0.98       0.99      0.98       100
...
```

Confusion Matrix:
A visual matrix was plotted showing model predictions vs actual labels. Most classifications were correctly placed along the diagonal, indicating accurate performance.

---

7. Conclusion

This project demonstrates the feasibility of using K-Nearest Neighbors for plant disease classification with high accuracy. KNN proved to be a simple yet powerful model that performs well on image-based classification without requiring heavy computational resources.