Smart Systems Final Project Report

\*\*Project Title:\*\* Plant Disease Classification using KNN

\*\*University:\*\* Alexandria National University

\*\*Faculty:\*\* Computers and Data Science

\*\*Instructor:\*\* Dr. Yasser Fouad

\*\*Team Members:\*\*: Adham Mahmoud Mahmoud - 2305367

Wael Ayman Youssery - 2305360

Mahmoud Hassan Mohamed - 2305356

Yahia Mohamed Eissa - 2305354

Youssef Mamdouh Mahmoud - 2305350

Alhassan Mohamed Hassan - 2305506

# 1. Objective

This project applies the K-Nearest Neighbors (KNN) algorithm to classify plant diseases from leaf images. It aims to assist in agricultural diagnostics by identifying diseases such as bacterial spot, early blight, or healthy status. The model helps prevent crop damage by enabling quick and automated detection.

# 2. Dataset

We used the PlantVillage dataset (https://www.kaggle.com/datasets/emmarex/plantdisease), which contains over 50,000 labeled images of diseased and healthy plant leaves across various crops.

# 3. Methodology

- Images are loaded from folders, each representing a disease class.  
- Images are resized to 64x64 pixels and kept in color (BGR).  
- Pixel arrays are flattened and used as feature vectors.  
- A KNN classifier (k=3) is trained on 80% of the data.  
- Model performance is evaluated using precision, recall, F1-score, and accuracy.

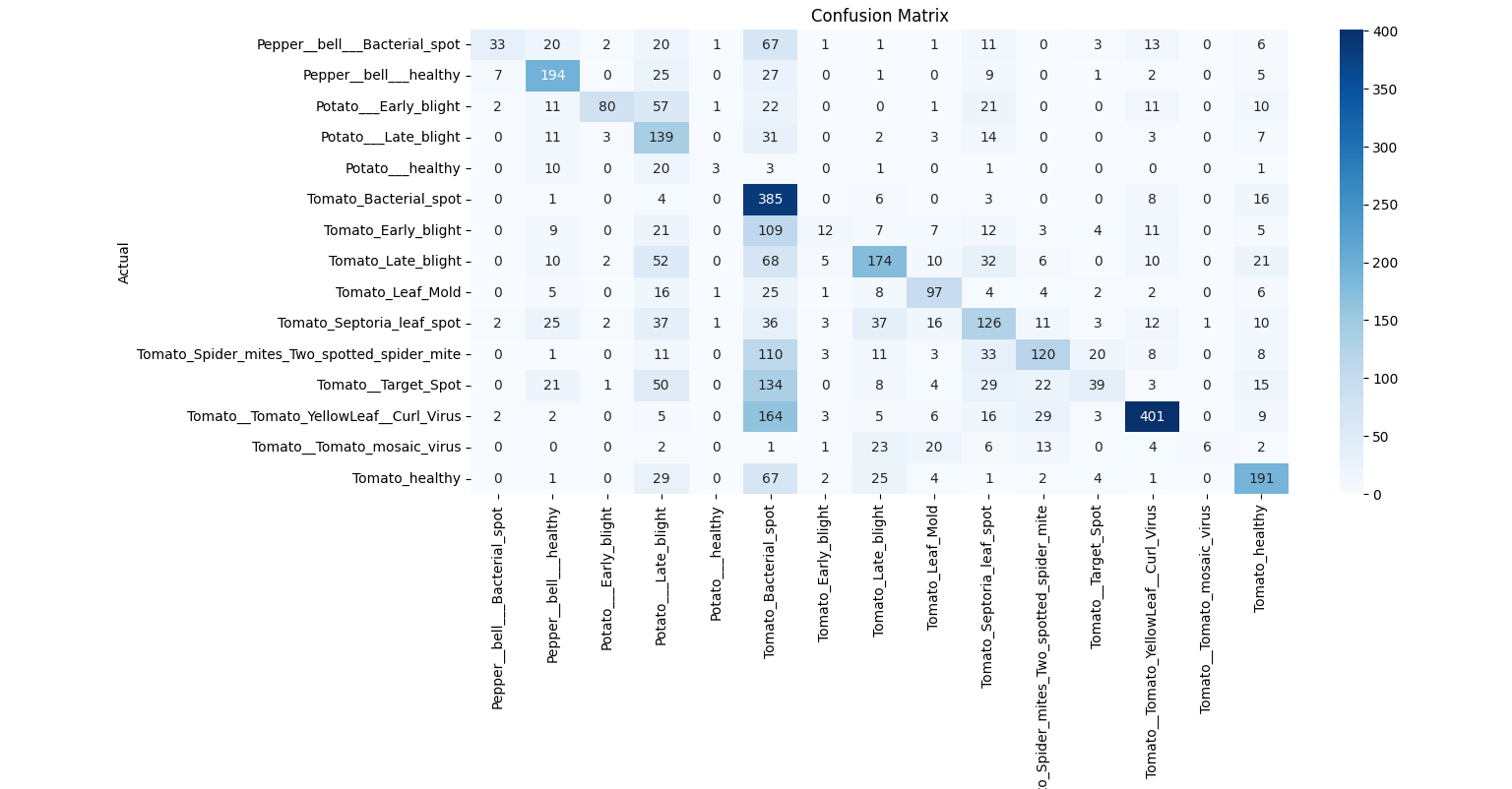
# 4. Results

The classification report below summarizes the testing performance:

precision recall f1-score support  
  
Pepper\_\_bell\_\_\_Bacterial\_spot 0.72 0.18 0.29 179  
Pepper\_\_bell\_\_\_healthy 0.60 0.72 0.66 271  
Potato\_\_\_Early\_blight 0.89 0.37 0.52 216  
Potato\_\_\_Late\_blight 0.28 0.65 0.40 213  
Potato\_\_\_healthy 0.43 0.08 0.13 39  
Tomato\_Bacterial\_spot 0.31 0.91 0.46 423  
Tomato\_Early\_blight 0.39 0.06 0.10 200  
Tomato\_Late\_blight 0.56 0.45 0.50 390  
Tomato\_Leaf\_Mold 0.56 0.57 0.57 171  
Tomato\_Septoria\_leaf\_spot 0.40 0.39 0.39 322  
Tomato\_Spider\_mites\_Two\_spotted\_spider\_mite 0.57 0.37 0.45 328  
Tomato\_\_Target\_Spot 0.49 0.12 0.19 326  
Tomato\_\_Tomato\_YellowLeaf\_\_Curl\_Virus 0.82 0.62 0.71 645  
Tomato\_\_Tomato\_mosaic\_virus 0.86 0.08 0.14 78  
Tomato\_healthy 0.61 0.58 0.60 327  
  
Accuracy 0.48 4128  
Macro avg 0.57 0.41 0.41 4128  
Weighted avg 0.57 0.48 0.47 4128

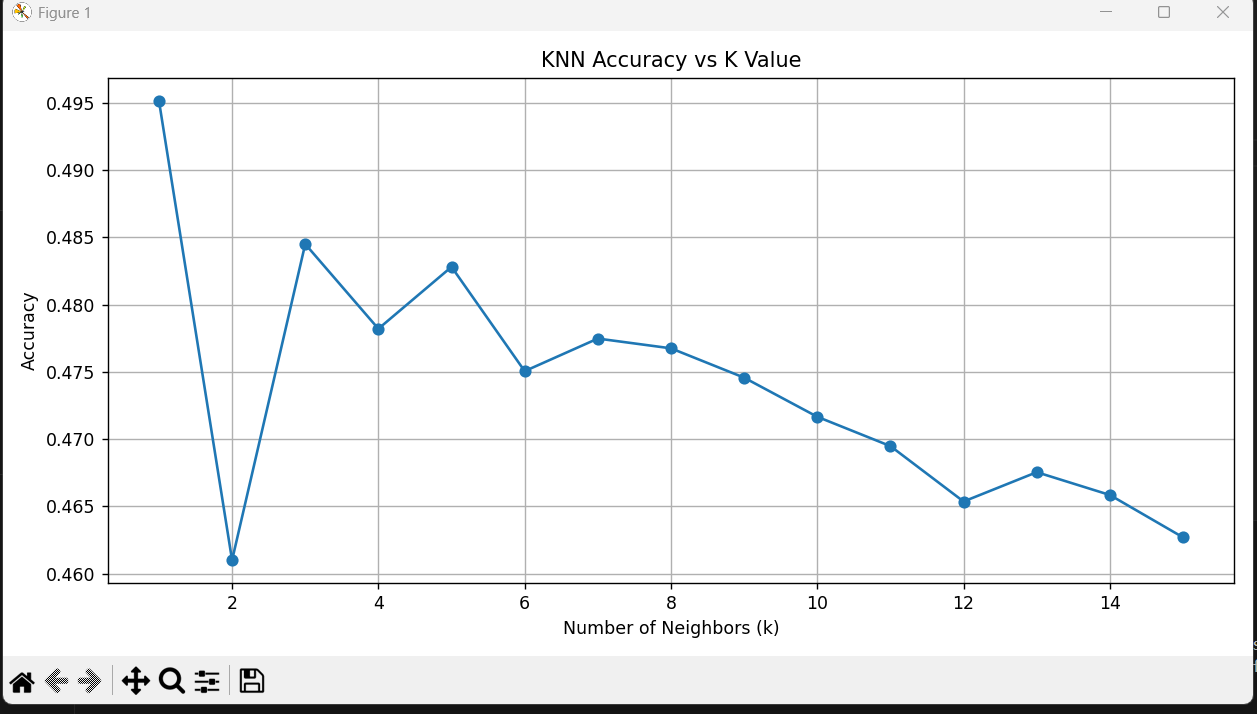
# 5. Confusion Matrix

The confusion matrix below shows the true labels versus predicted labels for all 15 disease classes:



# 6. KNN Tuning and Accuracy Plot

To find the optimal value of K, we trained the model using different values from 1 to 15 and plotted the accuracy for each value. The best performance was achieved at k = 1.



# 7. Conclusion

The KNN model achieved approximately 48% accuracy using color leaf images from the PlantVillage dataset. Color images improved performance significantly over grayscale. K = 1 gave the best results. The confusion matrix confirmed which diseases were often misclassified. While simple, KNN shows that plant disease classification is possible even with traditional models. Future improvements could include dimensionality reduction, feature extraction, or deep learning architectures for higher accuracy.