EAST WEST UNIVERSITY

Assignment-1

Mango Leaf Disease Classification

Course code: CSE 475

Course Title: Machine Learning

Section: 03

Fall 2024

Submitted by

Name: Nirzona Binta Badal

ID: 2021-2-60-051

Department of Computer Science & Engineering

Submitted To

Dr Raihan Ul Islam

Associate Professor

Department of Computer Science and Engineering

East West University

Date of Submission: 05 November, 2024

Objective

The objective of this project is to classify mango leaf diseases using image data and to evaluate the effectiveness of two machine learning models, Decision Tree and Random Forest, for this classification task.

Introduction

Accurate classification of plant diseases is essential in agriculture for early diagnosis and treatment. In this project, we focus on mango leaf diseases, aiming to categorize leaf images into disease classes using machine learning. The dataset consists of mango leaf images, each labeled with a disease class. The study compares the performance of Decision Tree and Random Forest classifiers in accurately identifying these classes.

Exploratory Data Analysis (EDA)

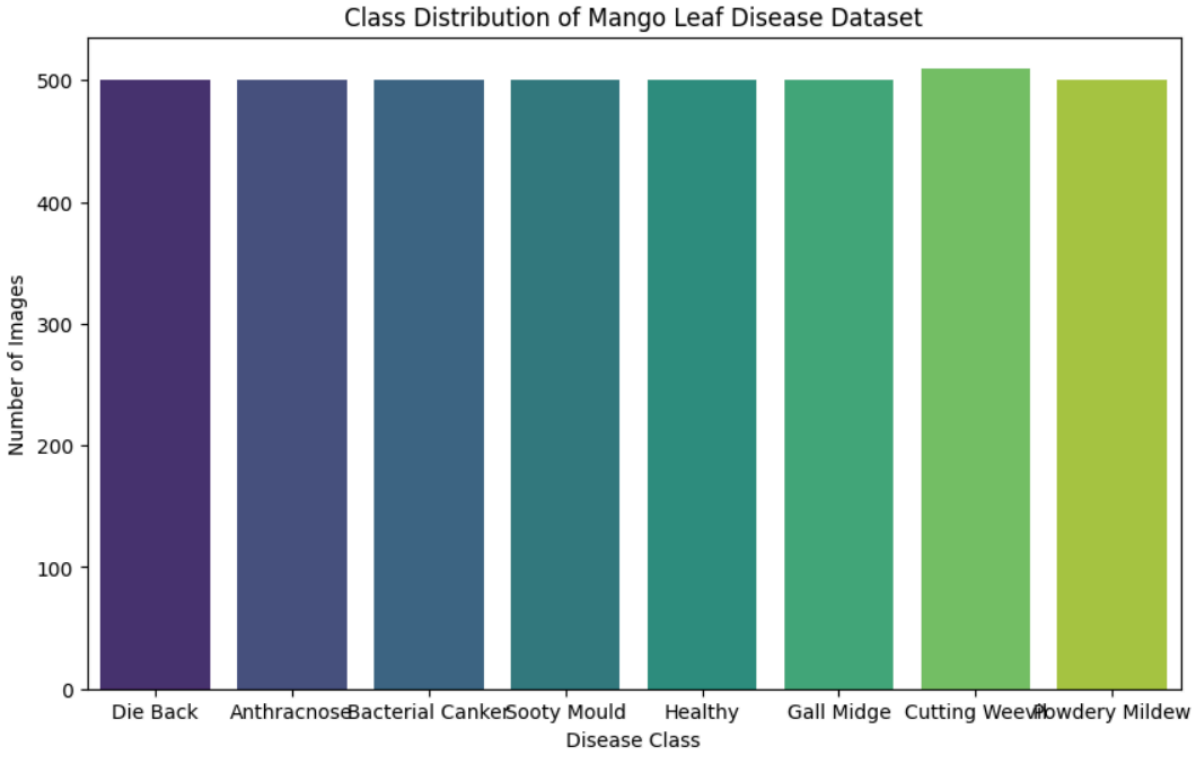
EDA provided insights into the structure and distribution of the dataset:

- **Class Distribution**: The dataset is composed of multiple classes, each corresponding to a different mango leaf disease, with varying numbers of images per class.
- **Sample Images**: Displaying sample images for each class offered a visual understanding of the dataset, showing the different patterns and textures indicative of each disease.

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
Dataset Description:
              Class  Number of Images
0          Die Back               500
1       Anthracnose               500
2  Bacterial Canker               500
3       Sooty Mould               500
4           Healthy               500
5        Gall Midge               500
6    Cutting Weevil               510
7    Powdery Mildew               500
<ipython-input-7-5bb33e0728cc>:32: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=F

  sns.barplot(data=class_counts_df, x='Class', y='Number of Images', palette='viridis')
```



- **Correlation Heatmap**: A heatmap was generated to observe correlations among the features (pixel values), offering insights into how feature relationships may vary across disease classes.

Data Preprocessing

The preprocessing steps included:

1. Loading and resizing each image to a fixed dimension to standardize input size.
2. Flattening images to transform them into feature arrays suitable for classification models.
3. Encoding disease labels as numerical values for compatibility with machine learning algorithms.
4. Splitting the data into training and testing sets, allowing for reliable performance evaluation.

Model Architecture

The project used two machine learning models:

1. **Decision Tree**: This model operates by recursively splitting the dataset based on feature values to form a tree-like structure, where each leaf represents a class label, and each node represents a decision point based on specific features. It's a highly interpretable model, often effective for smaller datasets or simpler classification tasks.

2. **Random Forest**: This ensemble model consists of multiple decision trees, each trained on random subsets of data and features. The final classification is determined by aggregating the predictions from each tree, making it more resilient to overfitting and more accurate, especially with complex datasets. This method is particularly well-suited for image classification tasks due to its ability to capture diverse patterns in the data.
1.

Training Process

Each model was trained using the standardized training dataset, with Random Forest leveraging an ensemble approach to reduce overfitting and improve generalization across different classes.

Evaluation Metrics

To evaluate model performance, accuracy and the confusion matrix were used:

- **Accuracy**: This metric measures the proportion of correct predictions across all classes.
- **Confusion Matrix**: Provided detailed insights into model performance on each disease class, showing the number of correct and incorrect predictions.

Results

The results indicate the following:

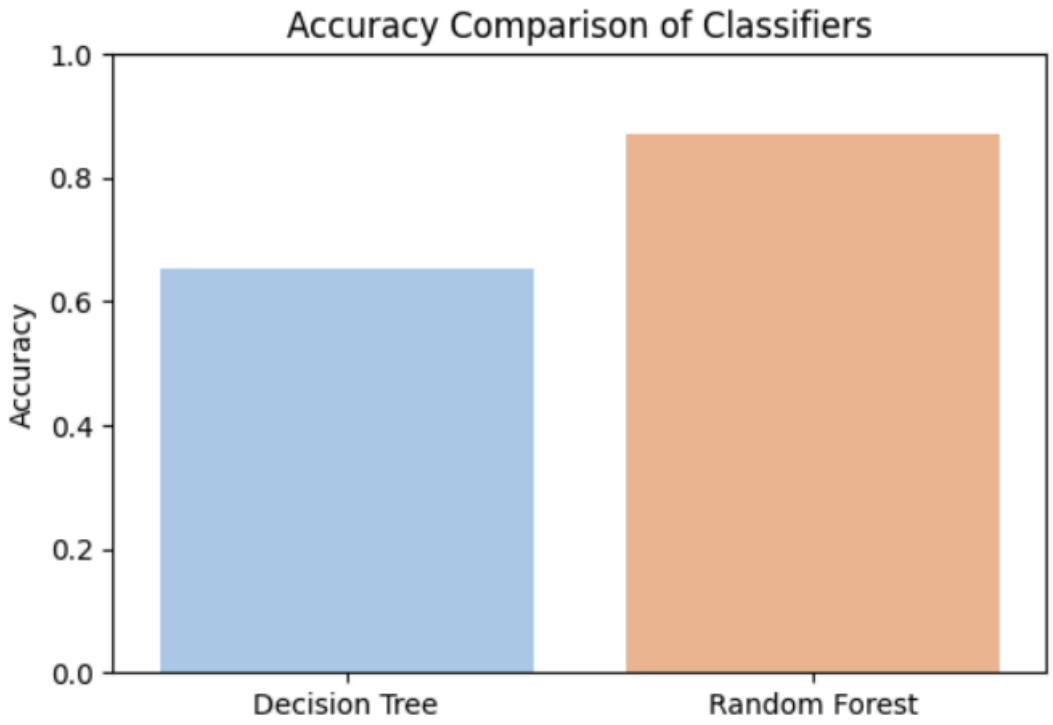
1. **Accuracy**:
    - Decision Tree achieved an accuracy of approximately 67%.
    - Random Forest achieved a higher accuracy of around 89%, indicating better performance in classifying the leaf diseases.

Decision Tree Classifier Report:

	precision	recall	f1-score	support
0	0.59	0.61	0.60	147
1	0.72	0.64	0.68	154
2	0.90	0.96	0.93	151
3	0.78	0.82	0.80	166
4	0.45	0.48	0.46	141
5	0.62	0.71	0.66	150
6	0.69	0.67	0.68	144
7	0.53	0.43	0.48	150
accuracy			0.67	1203
macro avg	0.66	0.66	0.66	1203
weighted avg	0.66	0.67	0.66	1203

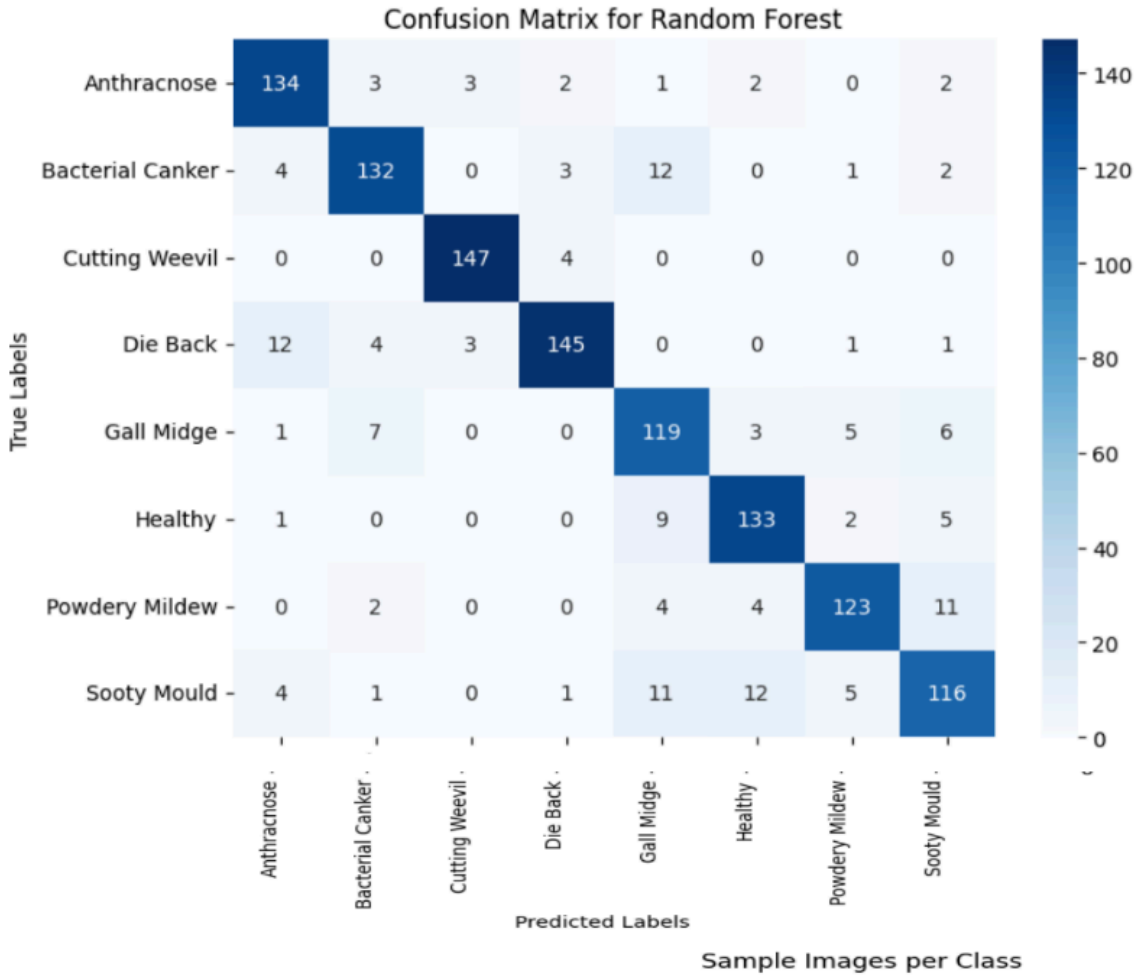
Random Forest Classifier Report:

	precision	recall	f1-score	support
0	0.94	0.93	0.93	147
1	0.88	0.90	0.89	154
2	0.98	0.99	0.98	151
3	0.95	0.94	0.94	166
4	0.76	0.79	0.77	141
5	0.85	0.91	0.88	150
6	0.94	0.87	0.90	144
7	0.83	0.81	0.82	150
accuracy			0.89	1203
macro avg	0.89	0.89	0.89	1203
weighted avg	0.89	0.89	0.89	1203




```
Summary of EDA Findings:
1. Dataset contains multiple classes of mango leaf diseases, with varying image counts per class.
2. Random Forest achieves higher accuracy than Decision Tree.
3. Confusion matrix shows classification performance per class with Random Forest.
```

2. **Confusion Matrix**:
   - The Random Forest model exhibited fewer misclassifications across classes compared to Decision Tree, highlighting its suitability for this task.



Conclusion

This project demonstrated the classification of mango leaf diseases using image data, with Random Forest outperforming Decision Tree in terms of accuracy and robustness. The findings suggest that Random Forest, with its ensemble nature, is a more effective model for high-dimensional data, such as images. The insights from this project can aid in developing automated systems for disease diagnosis in mango crops, contributing to better crop health management.