Text

Description automatically generated

**Department of Computer Science & Engineering**

**Lab Report**

**“LAB 01”**

**Machine Learning**

**Course Code: CSE 475**

**Section: 2**

**Submitted To:**

**Dr Raihan Ul Islam**

Associate Professor

Department of Computer Science and Engineering

**Submitted By:**

**Name: MD. Sakibur Rahman**

**ID: 2021-3-60-057**

**Submission Date: 4 March 2025**

**Introduction**

The goal of this project is to classify mango leaf diseases using machine learning techniques, specifically Random Forest and Decision Tree classifiers. Additionally, we performed Exploratory Data Analysis (EDA) to understand the dataset and assess its distribution as well as I have implemented a custom CNN model for comparison of both random forest and decision tree accuracy.

**Exploratory Data Analysis (EDA)**

EDA was conducted to gain insights into the dataset's structure and characteristics. The key findings include:

2.1 Class Distribution

• The dataset consists of 8 classes of mango leaf diseases.

• A bar chart and a pie chart were used to visualize the class distribution.

• The harmonic mean of class distribution was calculated to analyze dataset balance.

A screenshot of a graph

AI-generated content may be incorrect.

Figure: Class Distribution

A pie chart with different colored circles

AI-generated content may be incorrect.

Figure: Pie Chart class distribution

A graph of a person

AI-generated content may be incorrect.

Figure: Pixel Intensity histogram

A chart of different colored rectangular shapes

AI-generated content may be incorrect.

Figure: Box Plot of each class

2.2 Image Analysis

• Histogram of pixel intensities was generated to observe pixel value distribution.

• Box plots were used to compare pixel intensity ranges among different classes.

Also given an vision what actually the image looks in every class.

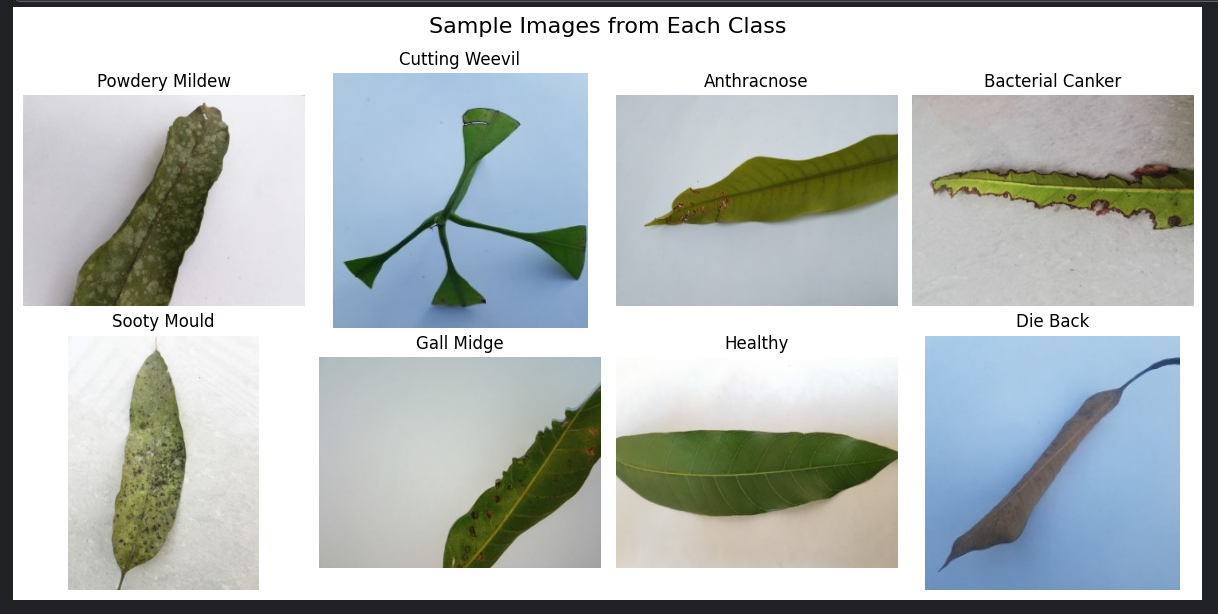


Figure: Sample image from each Class

**Model Training and Performance Evaluation**

Two models, Random Forest and Decision Tree, were trained on the dataset. The dataset was split into training and test sets (80%-20%).

3.1 Random Forest Classifier

Trained using 100 estimators.

Achieved an accuracy of 86%.

Confusion matrix and classification report were generated.

A graph with numbers and a number in the center

AI-generated content may be incorrect.

Figure: Confusion matrix of Random forest

3.2 Decision Tree Classifier

Achieved an accuracy of 67%.

Confusion matrix and classification report were generated.

A screenshot of a graph

AI-generated content may be incorrect.

Figure: Confusion matrix of Decision Tree

3.3 CNN model

A CNN model was designed with multiple convolutional and pooling layers.

Trained using 10 epochs with Adam optimizer.

Achieved an accuracy of 92%.

CNN outperformed both Decision Tree and Random Forest in classification accuracy.

A graph with numbers and a number in the center

AI-generated content may be incorrect.

Figure: Confusion matrix of CNN

**Comparative Analysis**

1. The CNN model performed the best among the three models, showing higher accuracy and better generalization.
2. The Random Forest model performed better than the Decision Tree model in terms of accuracy and generalization.
3. The confusion matrices showed that CNN had the least misclassifications, followed by Random Forest and Decision Tree.
4. Both traditional ML models struggled with certain leaf disease categories due to possible class imbalances.

**Testing on Selected Images**

A test was conducted using a manually selected image. Predictions were obtained from all three models, and results were compared. CNN showed the highest prediction consistency, followed by Random Forest, while Decision Tree had the highest misclassification rate.

# Test with selected 12 images

test\_images = test\_paths[:12] # Select first 12 images from test set

for img\_path in test\_images:

rf\_prediction = predict\_image(img\_path, rf\_model)

dt\_prediction = predict\_image(img\_path, dt\_model)

cnn\_prediction = predict\_image(img\_path, cnn\_model, is\_cnn=True)

print(f"Image: {img\_path}")

print(f"Random Forest Prediction: {rf\_prediction}")

print(f"Decision Tree Prediction: {dt\_prediction}")

print(f"CNN Prediction: {cnn\_prediction}\n")

A screenshot of a computer program

AI-generated content may be incorrect.

Figure: Predictions

A graph of different colored rectangles

AI-generated content may be incorrect.

Figure: Model Accuracy

**Conclusion**

This study demonstrated the effectiveness of machine learning in classifying mango leaf diseases. CNN outperformed both Random Forest and Decision Tree in terms of accuracy and reliability. Future work can involve:

* Enhancing CNN architecture with deeper layers for improved accuracy.
* Augmenting the dataset to address class imbalance.
* Exploring transfer learning techniques for better feature extraction and classification.