EAST WEST UNIVERSITY

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.

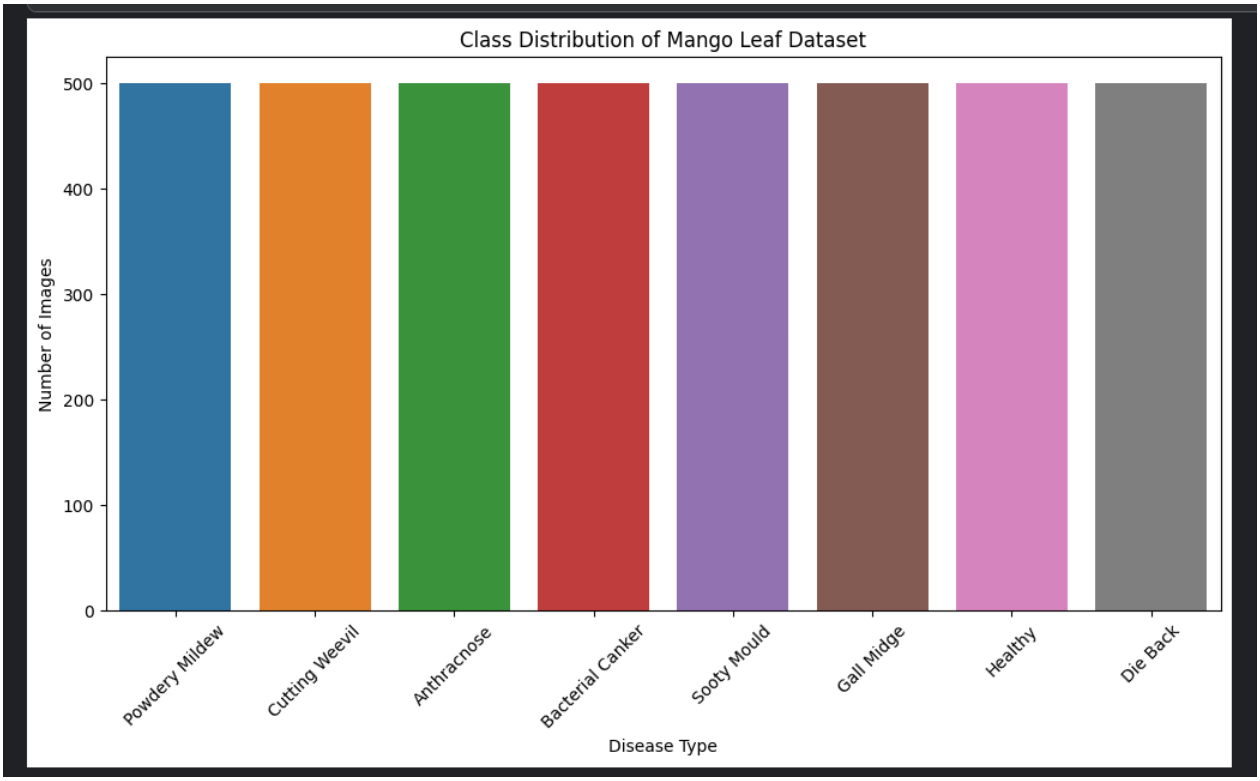

Figure: Class Distribution

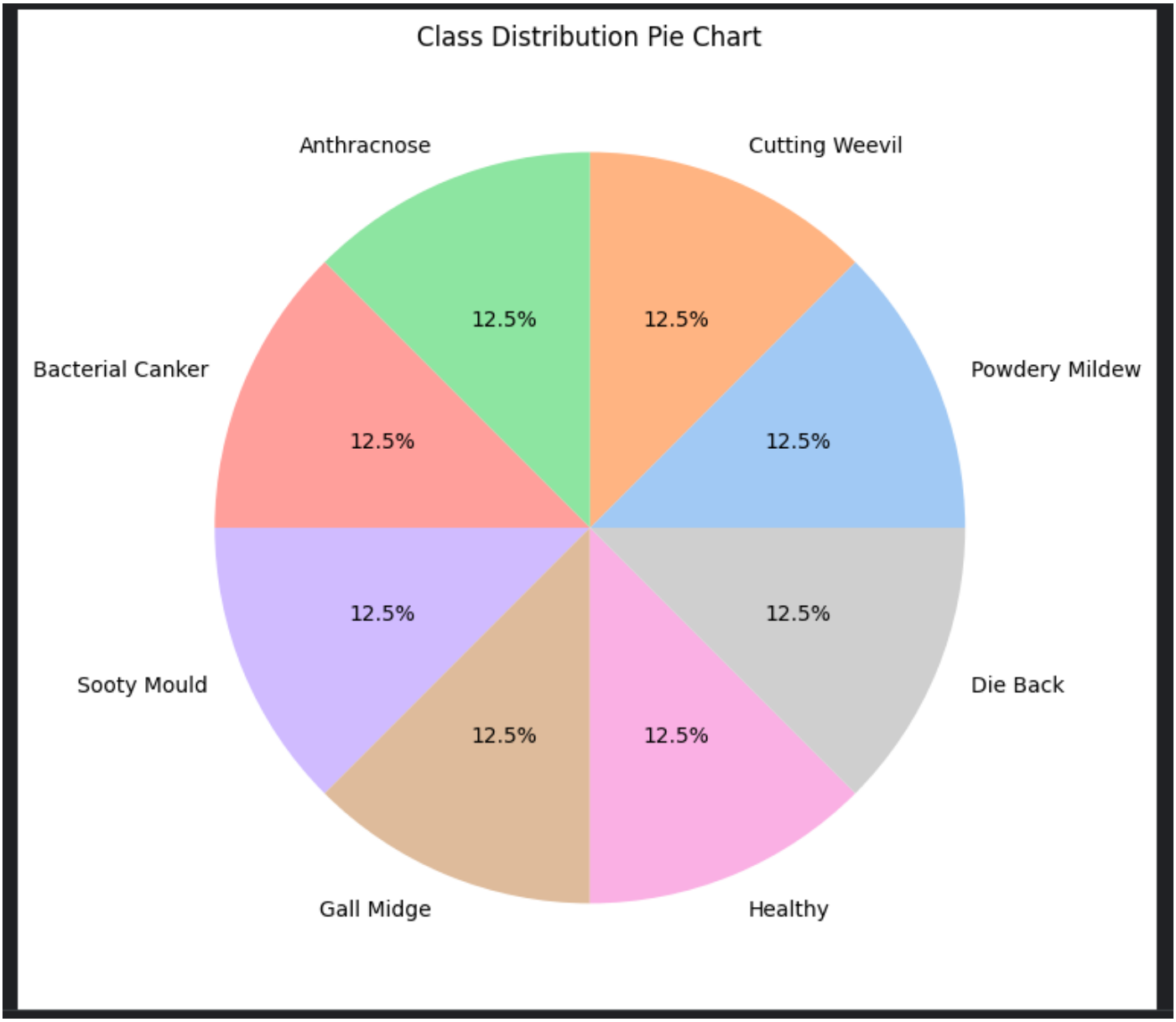

Figure: Pie Chart class distribution

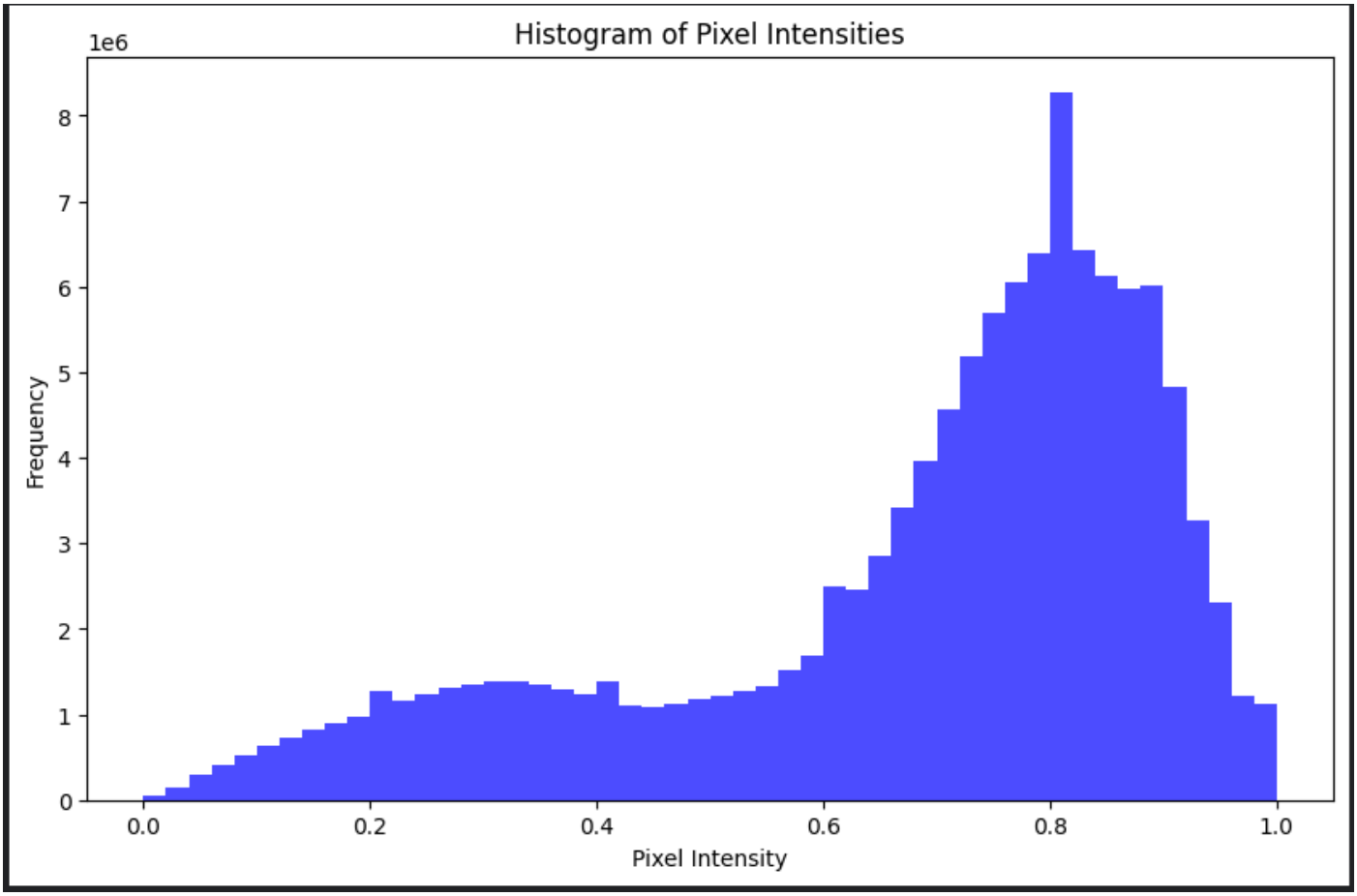

Figure: Pixel Intensity histogram

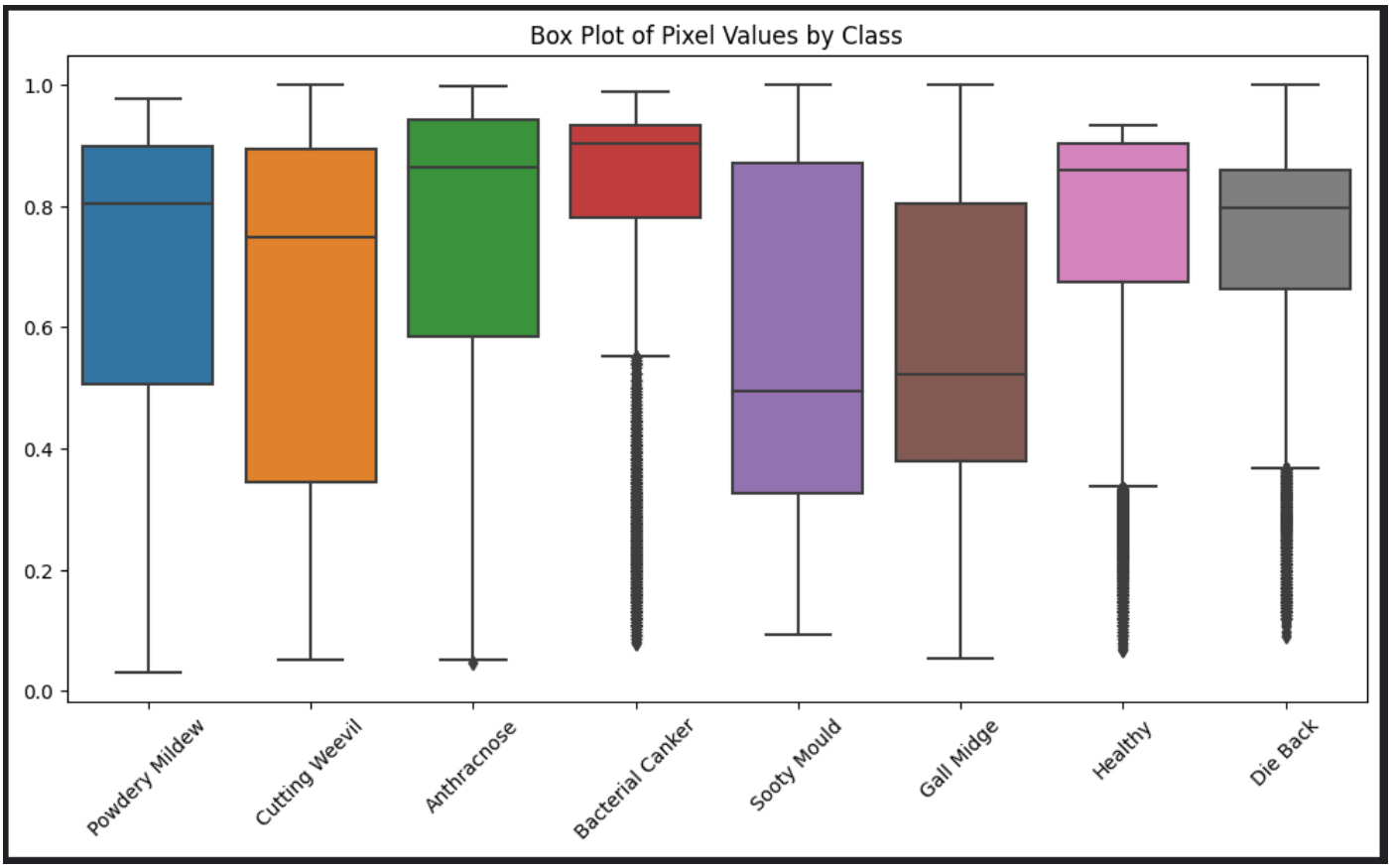

Figure: Box Plot of each class

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.

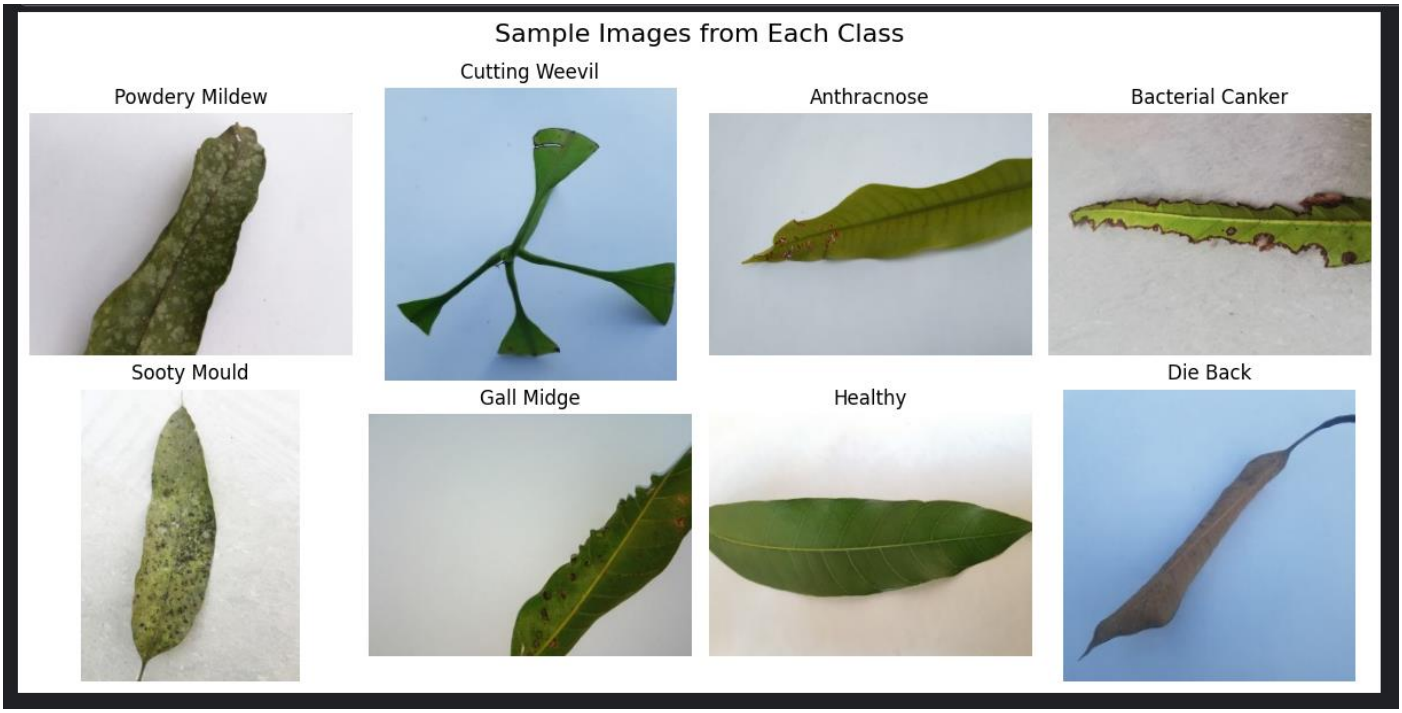

Figure: Sample image from each Class

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.

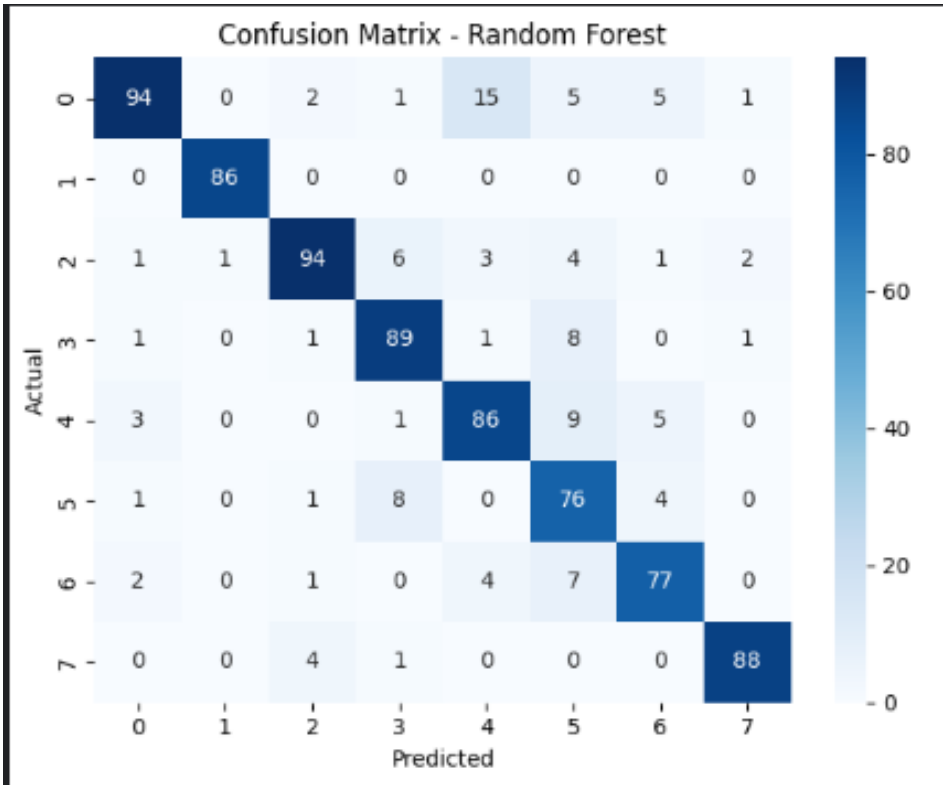

Figure: Confusion matrix of Random forest

3.2 Decision Tree Classifier

Achieved an accuracy of 67%.

Confusion matrix and classification report were generated.

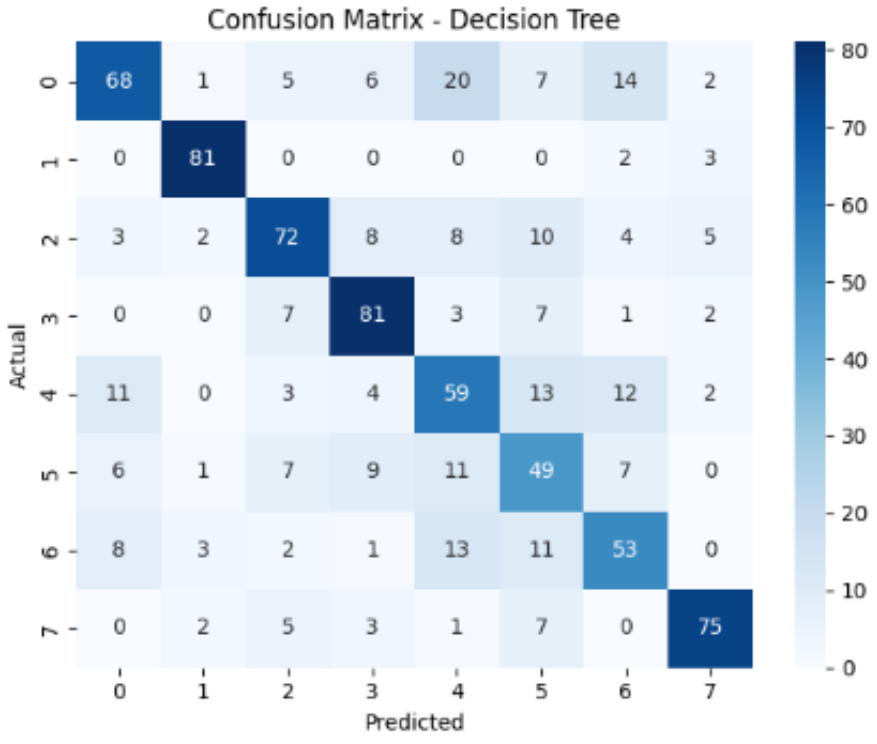

Figure: Confusion matrix of Decision Tree

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.

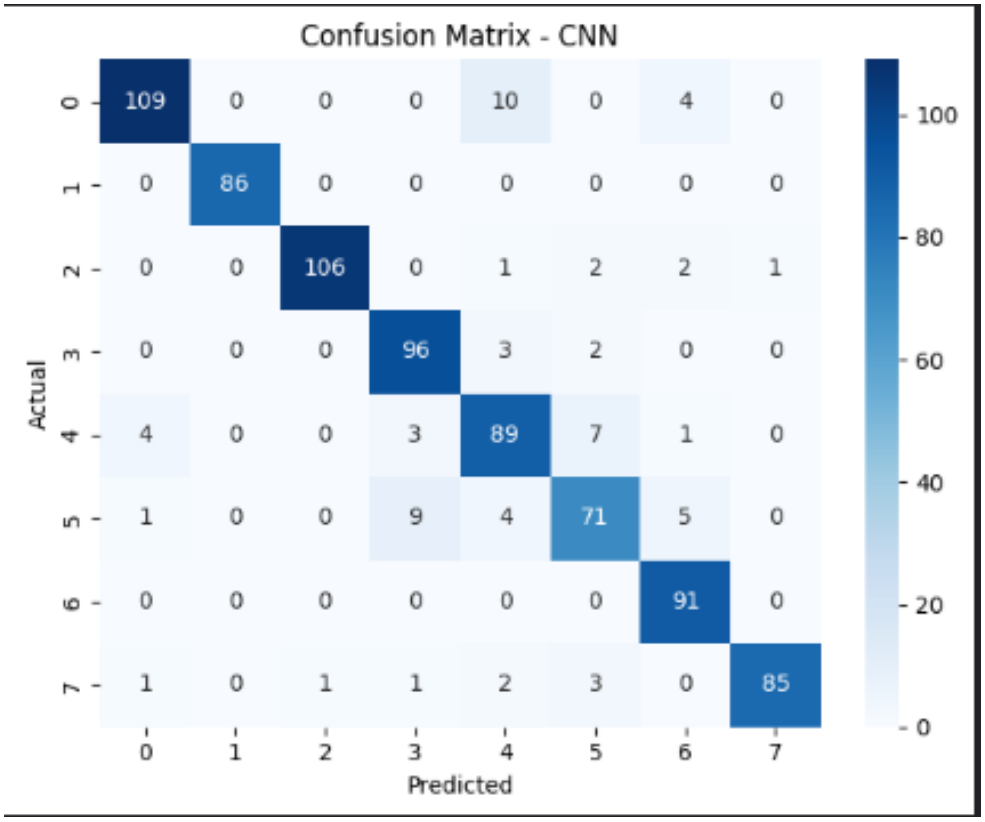

Figure: Confusion matrix of CNN

Comparative Analysis

I. The CNN model performed the best among the three models, showing higher accuracy and better generalization.
II. The Random Forest model performed better than the Decision Tree model in terms of accuracy and generalization.
III. The confusion matrices showed that CNN had the least misclassifications, followed by Random Forest and Decision Tree.
IV. 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")
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 443ms/step
Image: /kaggle/input/mangoleaf/MangoLeafBD Dataset/Cutting Weevil/20211011_170254 (Custom) (Custom).jpg
Random Forest Prediction: Cutting Weevil
Decision Tree Prediction: Cutting Weevil
CNN Prediction: Cutting Weevil

1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 21ms/step
Image: /kaggle/input/mangoleaf/MangoLeafBD Dataset/Healthy/20211231_162439 (Custom).jpg
Random Forest Prediction: Sooty Mould
Decision Tree Prediction: Gall Midge
CNN Prediction: Healthy

1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 16ms/step
Image: /kaggle/input/mangoleaf/MangoLeafBD Dataset/Cutting Weevil/20211011_162428 (Custom) (Custom).jpg
Random Forest Prediction: Cutting Weevil
Decision Tree Prediction: Cutting Weevil
CNN Prediction: Cutting Weevil

1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 16ms/step
Image: /kaggle/input/mangoleaf/MangoLeafBD Dataset/Die Back/20211129_160920 (Custom).jpg
Random Forest Prediction: Die Back
Decision Tree Prediction: Die Back
CNN Prediction: Die Back

1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 15ms/step
Image: /kaggle/input/mangoleaf/MangoLeafBD Dataset/Gall Midge/IMG_20211106_164555 (Custom).jpg
Random Forest Prediction: Gall Midge
Decision Tree Prediction: Gall Midge
CNN Prediction: Powdery Mildew

1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 16ms/step
Image: /kaggle/input/mangoleaf/MangoLeafBD Dataset/Powdery Mildew/IMG_20211107_123809 (Custom).jpg
Random Forest Prediction: Sooty Mould
Decision Tree Prediction: Sooty Mould
CNN Prediction: Powdery Mildew

1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 16ms/step
Image: /kaggle/input/mangoleaf/MangoLeafBD Dataset/Bacterial Canker/IMG_20211106_141106 (Custom).jpg
Random Forest Prediction: Bacterial Canker
Decision Tree Prediction: Bacterial Canker
CNN Prediction: Bacterial Canker
```

Figure: Predictions

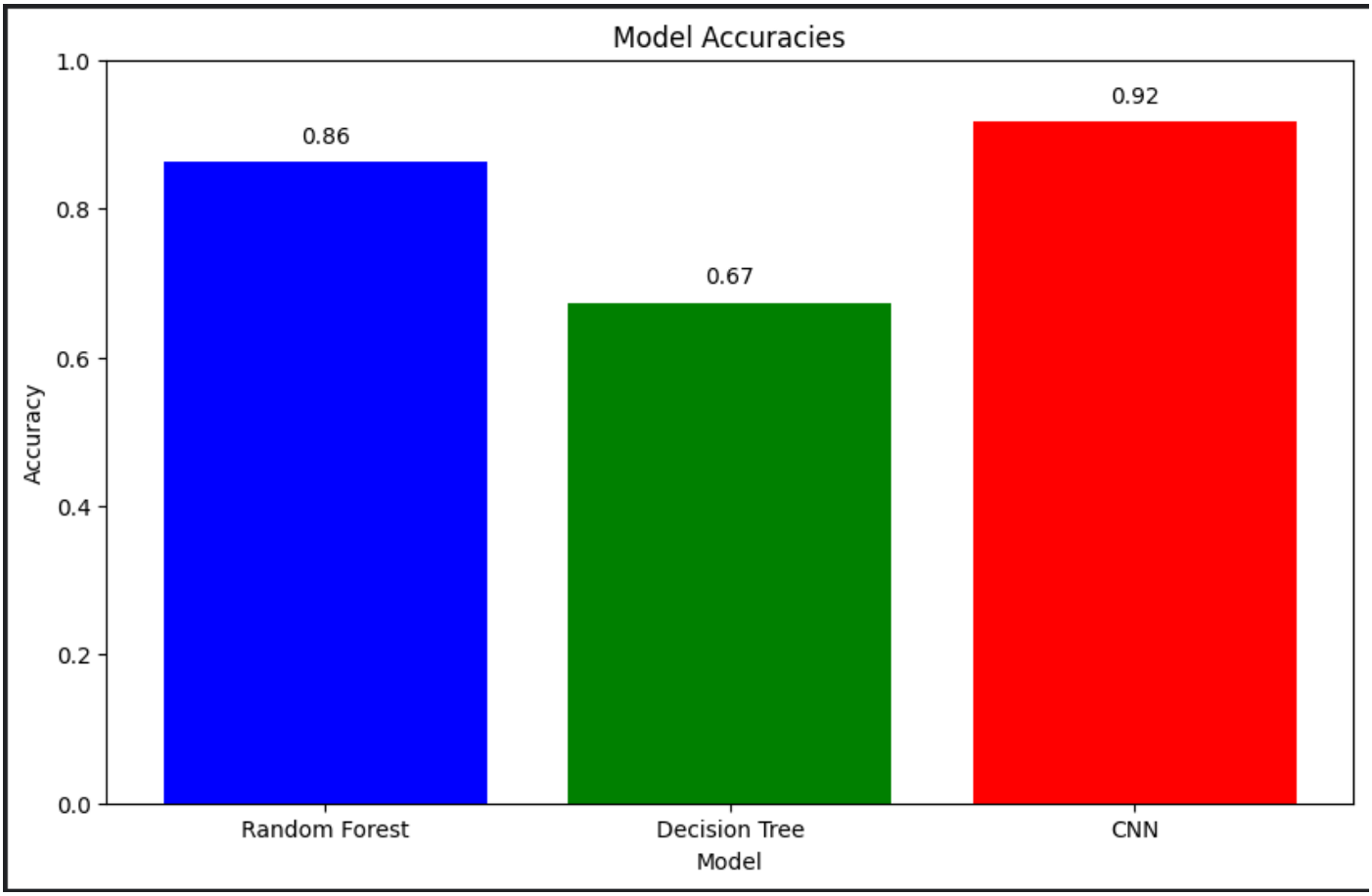

Figure: Model Accuracy

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.