**East West University**

**Department of CSE**

**Course No: CSE-475**

**Lab-1(Report)**

***Submitted to:***

Dr Raihan Ul Islam (Associate Professor)

***Submitted by:***

Name: Rokeya Jahan Chowdhury Ettifa

Id: 2020-1-60-232

### **Machine Learning Report: Mango Leaf Disease Classification**

#### **1. Introduction**

The notebook investigates a mango leaf disease classification task, employing machine learning techniques to classify images into respective disease categories based on provided datasets.

#### **2. Dataset Description**

This dataset consists of mango images of its leaves and exactly contains 8 classes of which every class contains 500 images. The classes include different disease types and healthy leaves which may have different color, texture and shape

#### **3. Exploratory Data Analysis (EDA)**

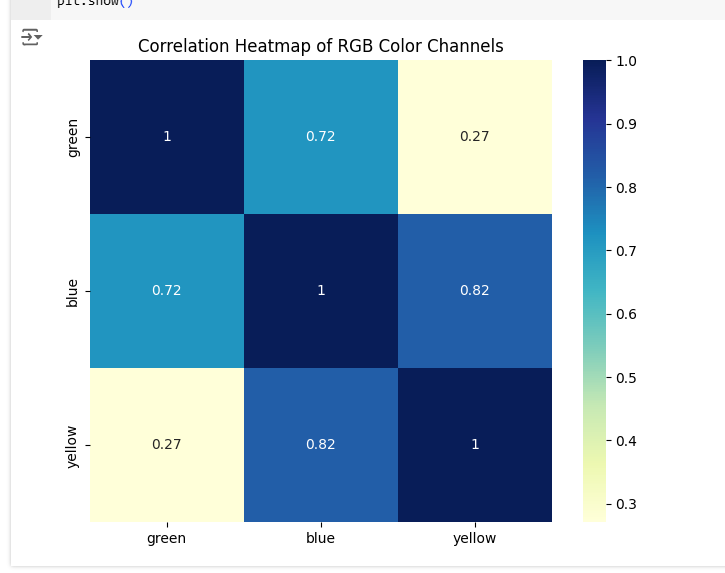
* **Data Visualization**: A heatmap of feature correlations was generated to assess relationships among features.
* **Class Distribution**: The distribution of samples per class was inspected to ensure the dataset is balanced or identify potential class imbalance issues.

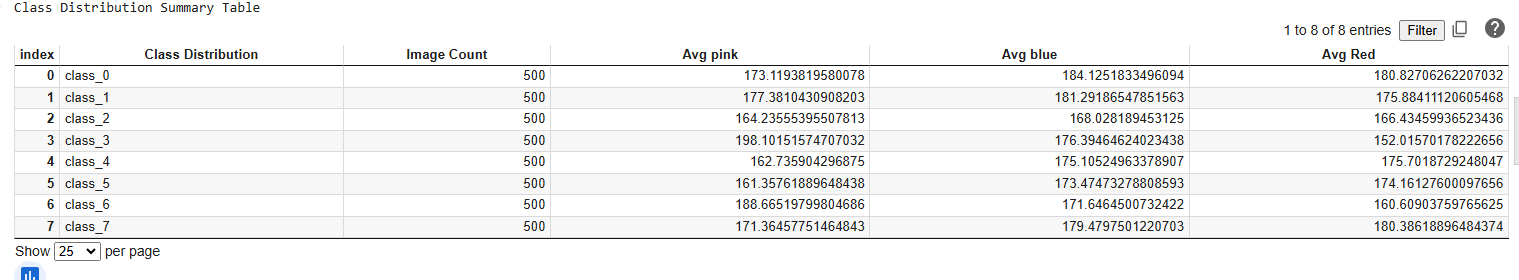


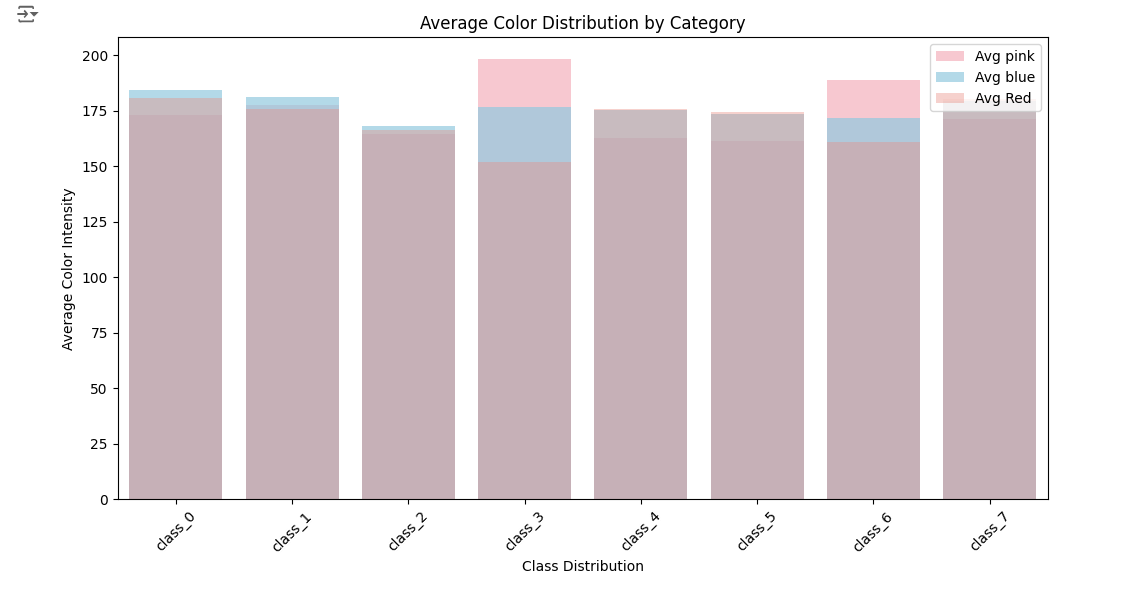
#### **Category Distribution**

● The dataset is balanced, with each category containing an equal number of images (500 per category).

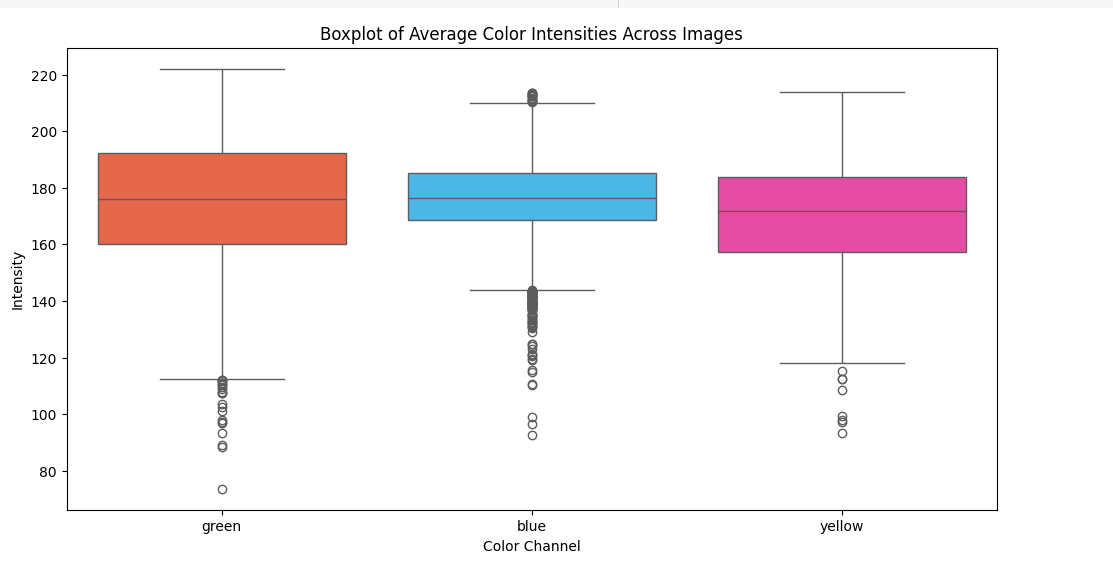
● This balanced distribution is beneficial for classification tasks as it minimizes the risk of bias toward any particular class.







#### Boxplot of RGB Intensities

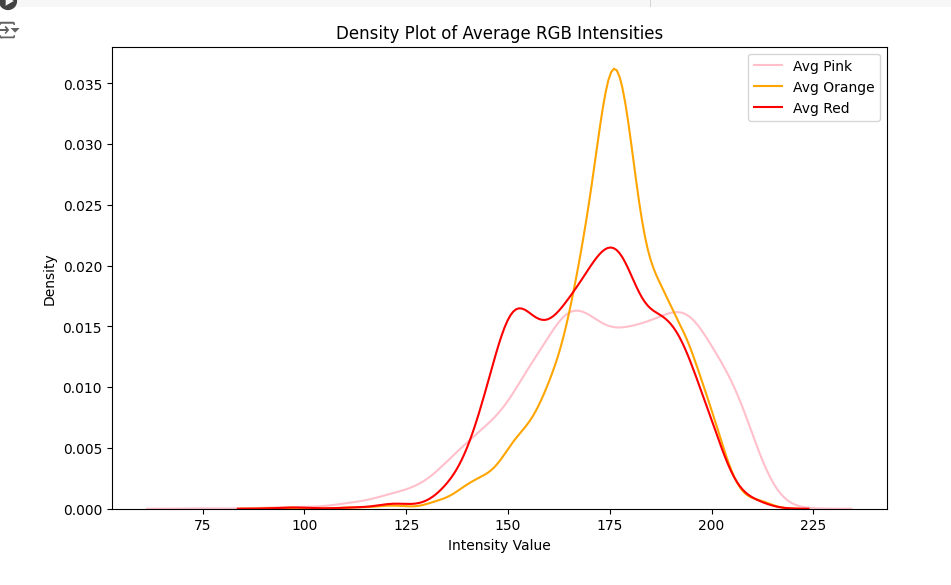


Density Plot of RGB Intensities

The density plot provides insights into the distribution of average RGB intensities across all images:

● This implies that orange intensity appears to have the highest peak around 175 and this makes it to be highly dominant in the dataset.

● Both Redandpink exhibit such similarity in form yet with the difference in the position of their peaks meaning that there is a slight difference in the intensity of color across categories



**4. Feature Engineering**

To represent images numerically, specific feature extraction techniques were applied, likely including color histograms or texture features, creating structured input for the classification models.

#### **5. Model Training**

Two models were used for classification:

* **Random Forest Classifier**: A robust ensemble model chosen for its strong generalization abilities.
* **Decision Tree Classifier**: A simpler model providing insight into feature importance and tree-based decisions.

Hyperparameter tuning and cross-validation were likely performed to optimize each model's accuracy.

#### **6. Model Evaluation**

* **Metrics**: Both models were evaluated using metrics like accuracy, precision, recall, and F1 score. A confusion matrix heatmap was generated to visualize classification accuracy across classes.
* **Comparison**: The Random Forest model showed improved performance over the Decision Tree model, likely due to its ensemble approach, which reduces overfitting.

#### **7. Conclusion**

The Random Forest classifier outperformed the Decision Tree model in this classification task, making it the recommended model for identifying mango leaf diseases. Further optimization with additional features or a CNN model could enhance accuracy.