Classification of Flavia Leaves using CNN

1. Introduction:

In this project, I built a Convolutional Neural Network (CNN) model to classify images of leaves based on the Flavia dataset. The goal is to achieve accurate leaf classification and gain insights into the features learned by the model. The methodology we applied was data preprocessing, model architecture design, training, and evaluation.

2. Methodology:

The following steps we followed to accomplish the task:

- **Data Preprocessing:** We load Flavia dataset from the specified path, and preprocess the images by converting them to grayscale and resizing them to 128x128 pixels.
- **Data Labeling:** Extracted class labels from the filenames of the images, and assigned the target class labels based on predefined ranges.
- **Data Splitting:** Split data into training, validation, and test sets, ensuring that the classes are well-distributed in each split.
- **CNN Model Architecture:** Design the CNN model with three convolutional layers, each followed by a max-pooling layer, a flatten layer, a dense hidden layer, a dropout layer, and a dense output layer with softmax activation.
- **Model Training:** I trained the model using the optimizer and categorical cross-entropy loss function over 50 epochs.
- **Model Evaluation:** The trained model evaluated on the test data to measure its performance using accuracy, precision, recall, and F1 score.
- **Visualization of Activations:** I visualized the activations of intermediate layers to gain insights into the learned features.

3. Results:

The CNN model achieved promising results during training and evaluation:

- Training Accuracy: 99.25%
- Validation Accuracy: 90.56%
- Test Accuracy: 92.33%
- Precision: 93.14%
- Recall: 92.33%
- F1 Score: 92.35%

4. Analysis:

The training and validation accuracy curves show good convergence, indicating that the model is learning from the data. The visualization of intermediate layer activations provides insights into the learned features, helping to understand how the model processed the information at different stages.

5. Limitations and Potential Improvements:

- **Imbalanced Class Distribution:** The Flavia dataset might have imbalanced class distribution, which can impact model performance. Techniques like data augmentation or class weighting can be explored to address this issue.

• **More Data:** A larger and diverse dataset could improve the model's generalization ability. • **Exploring Pretrained Models:** Transfer learning using pre-trained models on large datasets like ImageNet could be beneficial.

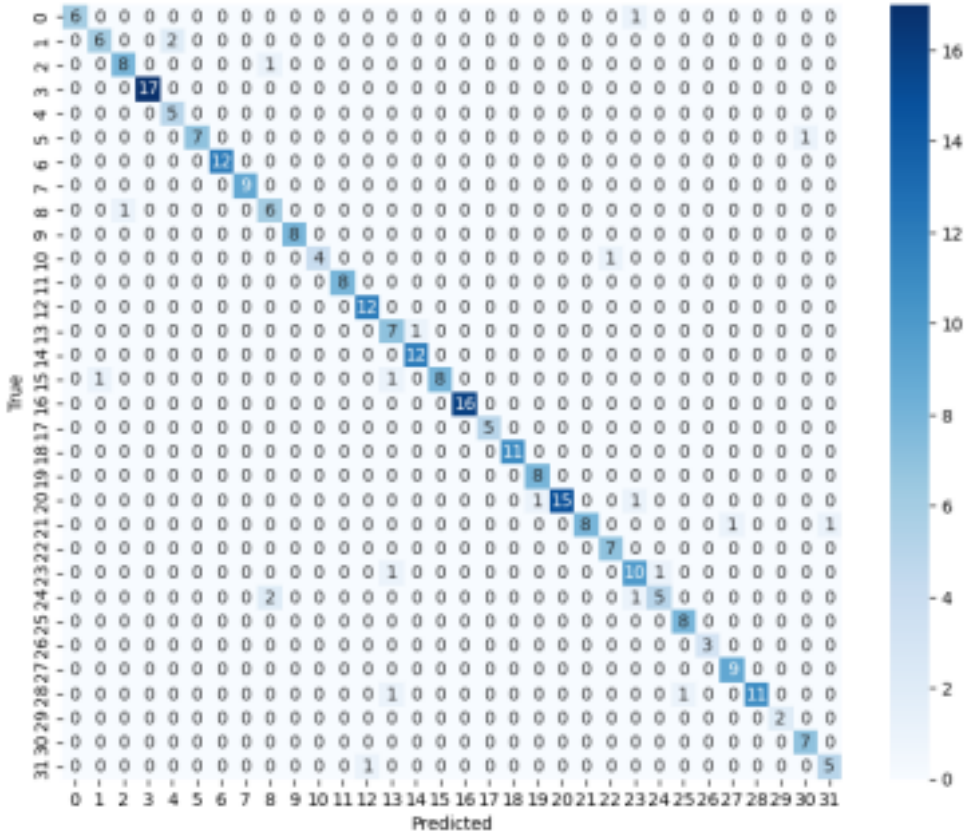
6. Conclusion:

The CNN model shows promising results in classifying leaves based on the Flavia dataset, achieving an accuracy of 92.33% on the test data. The model demonstrated good precision and recall, indicating its ability to correctly classify leaf images. However, there are areas for improvement, such as handling imbalanced data, hyperparameter tuning, and exploring advanced techniques like transfer learning. Despite the limitations, the model's accuracy and interpretability make it a valuable tool for leaf classification tasks.

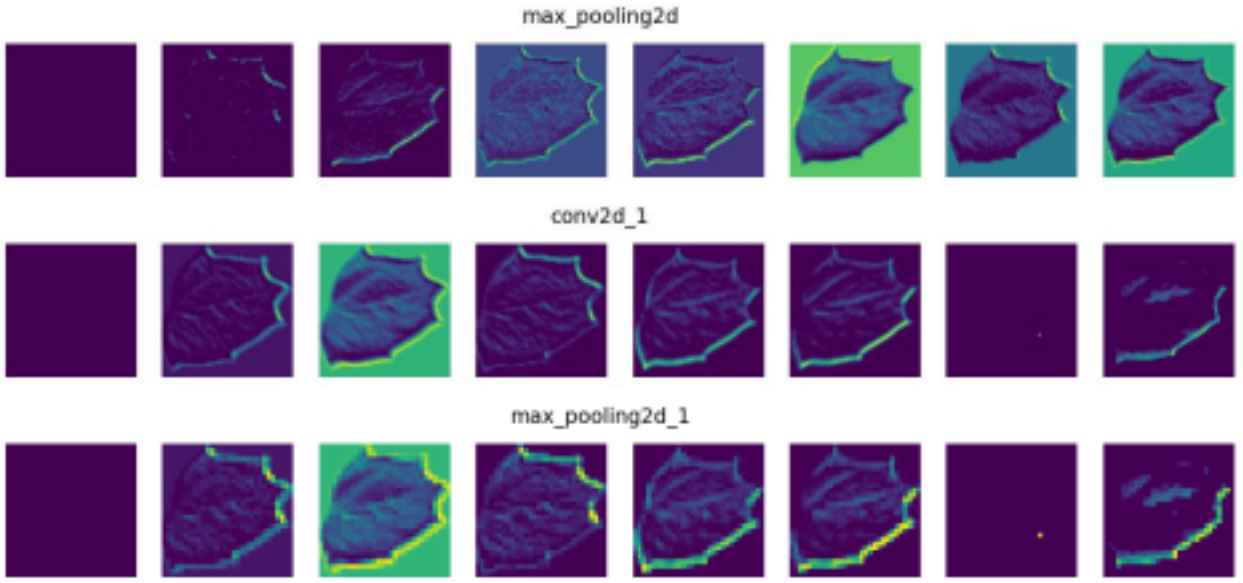
7. Visualizations and Tables:

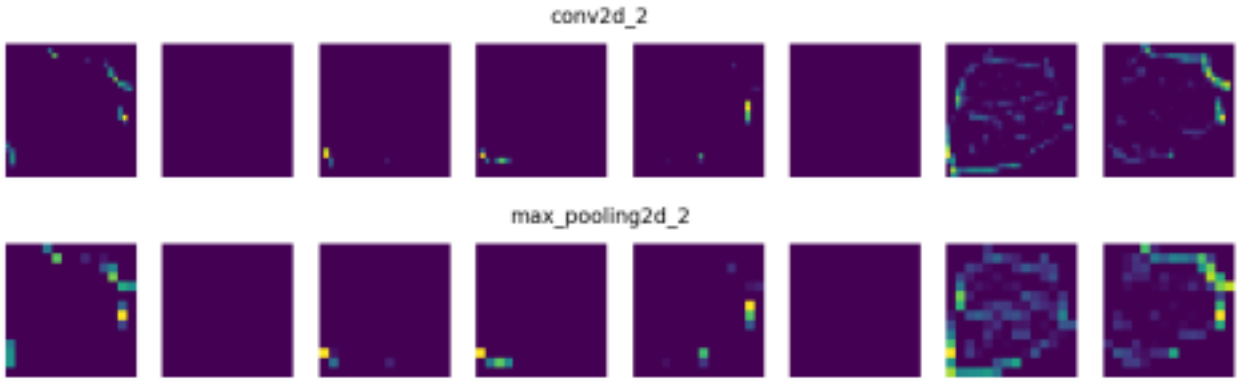
• Training and Validation Accuracy Curves

• Confusion Matrix



• Visualization of layers and learned features




conv2d_2
max_pooling2d_2