Pollen Grain Classification - Image Preprocessing Documentation

1. Project Objective:

   To classify pollen grains based on their microscopic image using machine learning.

2. Dataset Structure:

   - Images are stored with filenames indicating their class (e.g., 'anadenanthera_16.jpg').

   - Dataset contains various pollen types based on the filename prefix.

3. Preprocessing Steps:

   - Images are resized to 128x128 pixels to standardize input size.

   - Images are normalized to range [0,1] by dividing pixel values by 255.

   - Labels are extracted from filenames and encoded into numerical format.

   - Dataset is split into training and testing sets (80/20 split).

4. Data Augmentation (Optional):

   - Includes rotation, shifting, zooming, and flipping to improve model generalization.

5. Next Steps:

   - Build a Convolutional Neural Network (CNN) for classification.

   - Train the model using the augmented dataset.

   - Evaluate accuracy and refine model as needed.

This preprocessing pipeline ensures consistent and robust input for model training.