**Butterfly Image Classification Using MobileNetV2 and Enhanced Deep Learning Architecture**

BDA-2206

Zhakhaev Madiyar, Gazizov Adilkhan

## 

## **1. Introduction**

This project focuses on creating a butterfly image classification system using machine learning and deep learning techniques. Butterflies are diverse and beautiful insects, with thousands of species found globally. Accurate classification of butterfly species is crucial for biodiversity monitoring, ecological research, and conservation efforts.

Using a pre-labeled dataset from Kaggle, the project aims to preprocess the data, apply transfer learning with MobileNetV2, and enhance the base model for improved classification accuracy. The output includes a robust deep learning model capable of identifying 75 butterfly species.

## **3. Dataset Description**

The dataset used in this project is sourced from Kaggle's **Butterfly Image Classification Dataset**. This dataset contains images of butterflies from 75 distinct species. The data is provided in a structured format, enabling efficient preprocessing and training.

### **Key Dataset Details:**

1. **Total Images:**
   * 6499 labeled images.
2. **Class Distribution:**
   * 75 butterfly species, with class names and numerical labels provided in the accompanying CSV file.
3. **Data Split:**
   * The data is split into **80% training data (5199 images)** and **20% validation data (1300 images)** to ensure adequate representation across all species.
4. **File Organization:**
   * The images are stored in a single directory. The CSV file contains the filenames and corresponding class labels, which are used for organizing images into class-specific directories during preprocessing.

### **Challenges in the Dataset:**

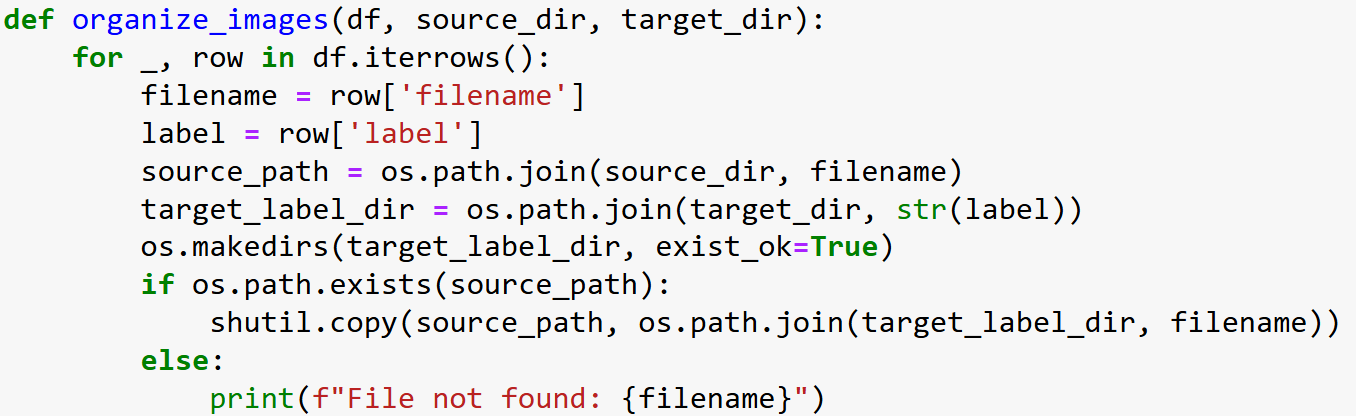
* **Class Imbalance:**Some butterfly species have more images than others, potentially biasing the model. Data augmentation is used to address this issue.
* **Variability in Image Quality:**Images vary in resolution, lighting, and background, requiring normalization and data augmentation to enhance generalization.

## **4. Methodology**

### **4.1 Data Preprocessing**

#### **Steps Involved:**

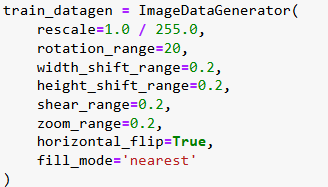
1. **Download Dataset:**The dataset is downloaded programmatically using the Kaggle API for reproducibility.
2. **Organize Files:**A script organizes images into class-specific folders for easier processing.
3. **Resizing and Normalization:**Images are resized to 224x224 pixels to match the input requirements of MobileNetV2 and normalized to scale pixel values between 0 and 1.
4. **Train-Validation Split:**A stratified split ensures that each class is proportionally represented in both the training and validation sets.



### **4.2 Data Augmentation**

Augmentation techniques are applied to artificially increase the size of the dataset and improve generalization. These include:

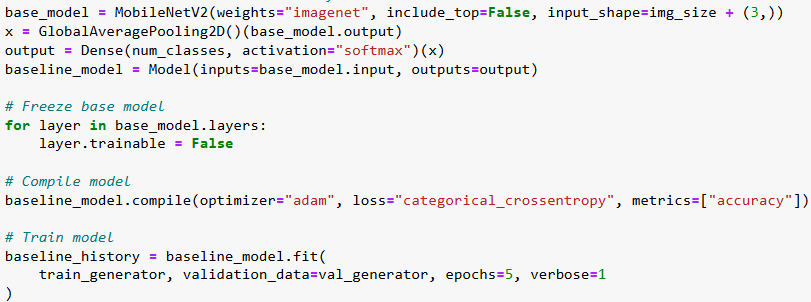
* Random Rotation
* Width and Height Shifting
* Shearing
* Zooming
* Horizontal Flipping



### **4.3 Model Implementation**

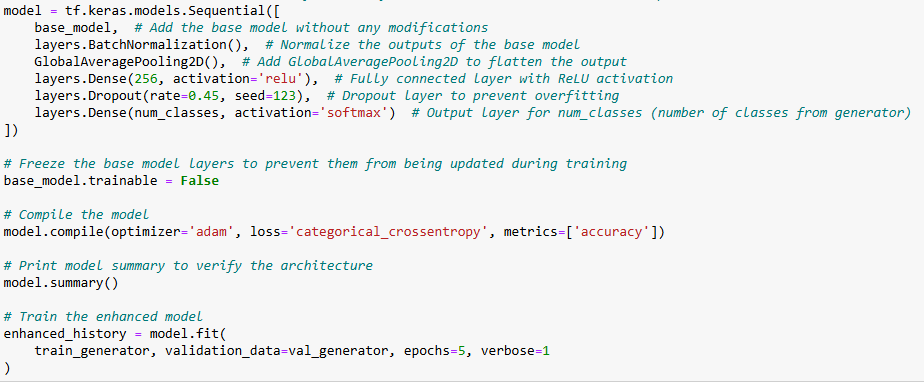
#### **Baseline Model (MobileNetV2)**

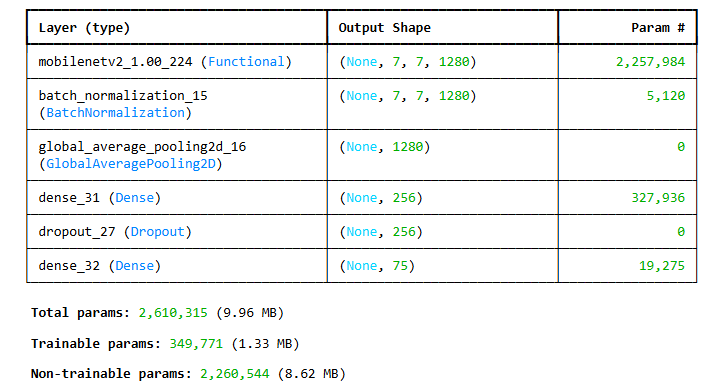
The base model uses MobileNetV2, pre-trained on the ImageNet dataset. A global average pooling layer and a dense output layer are added for classification.



#### **Enhanced Model:**

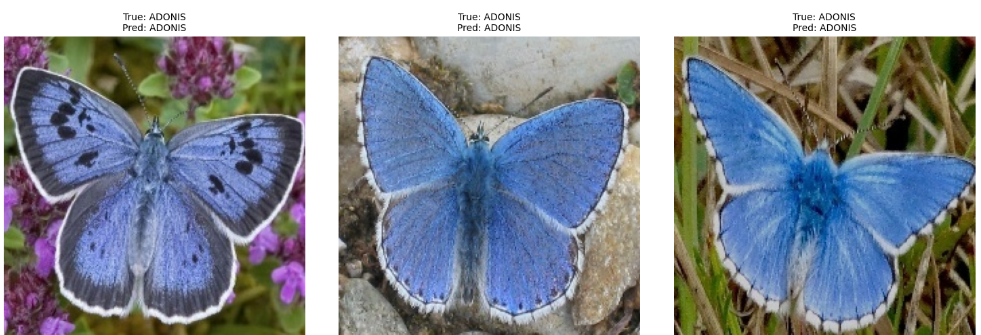
The enhanced architecture incorporates additional Dense and Dropout layers to reduce overfitting and improve accuracy.





### **4.4 Output Visualization**

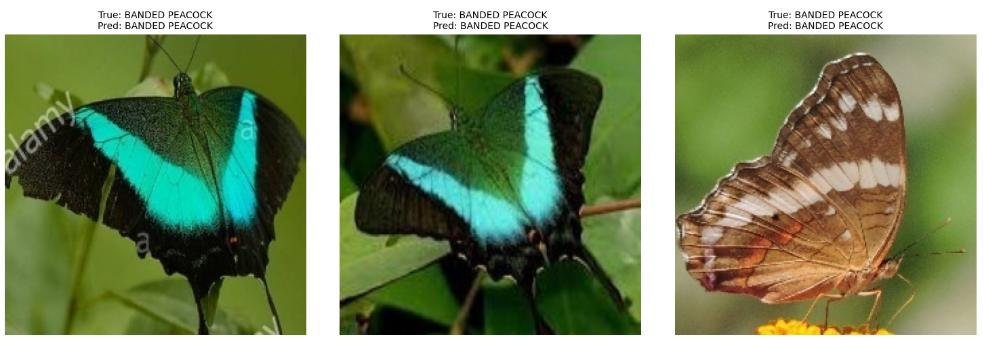
#### **Baseline Model (MobileNetV2)**

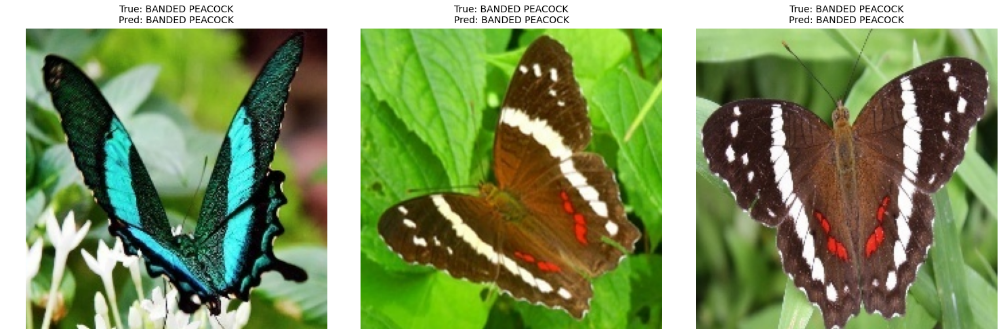






#### **Enhanced Model:**

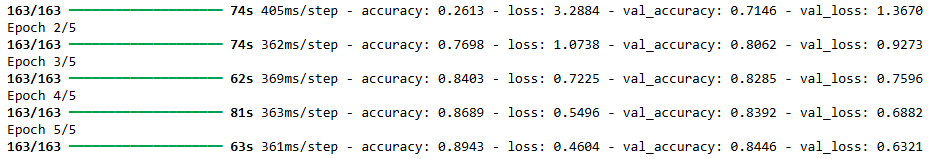




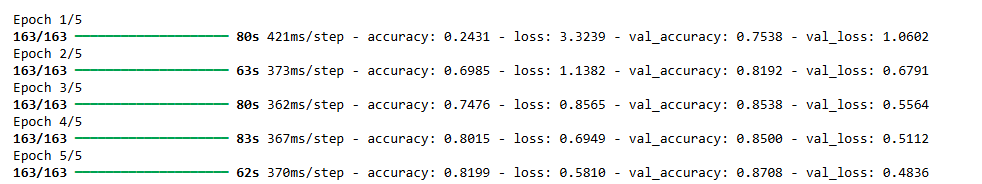


## **5. Results**

#### **Baseline Model (MobileNetV2)**



**Enhanced Model:**

****

## **6. Conclusion**

This project successfully classified 75 butterfly species using an enhanced deep learning model based on MobileNetV2. The enhancements improved accuracy while keeping training times manageable.

Screenshots of correctly and incorrectly classified images are attached to showcase the model's performance.