ENCHANTED WINGS

- MARVELS OF BUTTERFLY SPECIES

Prepared By

Udayagiri Poorna Akshaya

23B01A45B0

Shri Vishnu Engineering College for Women

2024 - 2025

Enchanted Wings: Marvels of Butterfly Species

---

1. Objective / Problem Statement

This project aims to develop a robust butterfly image classification model using transfer learning techniques. By leveraging a diverse dataset of butterfly species, the model can efficiently and accurately identify 75 different butterfly classes from images. The goal is to support biodiversity monitoring, ecological research, and citizen science initiatives by providing a fast, reliable tool for butterfly species identification.

---

2. Dataset Details

- **Total Images:** 6,499 butterfly images
- **Number of Classes:** 75 butterfly species
- **Dataset Split:** ~80% Training / ~20% Validation & Testing
- **Preprocessing:**
    - Resizing to 224×224 pixels
    - Normalization
    - Data augmentation (rotation, shifts, shear, zoom, brightness adjustment, horizontal flip)

---

3. Model Architecture and Training

- **Transfer Learning Base:** MobileNetV2 (pre-trained on ImageNet, frozen base)
- **Custom Classification Head:**
    - Global Average Pooling
    - Dense(256, ReLU) → Dropout
    - Dense(128, ReLU) → Dropout
    - Dense(75, Softmax)
- **Loss Function:** Categorical Cross-Entropy
- **Optimizer:** Adam
- **Training Strategy:**
    - 20 epochs with base frozen

- 30 epochs fine-tuning with class weights and callbacks (early stopping, LR scheduler)

---

4. Data Augmentation

- Augmentation done using ImageDataGenerator for the training set
- Techniques used: Rotation, zoom, shifts, shearing, brightness, and flips
- Validation data was only rescaled (no augmentation)
- Class weights were calculated and applied to handle imbalance

---

5. Streamlit Frontend

A user-friendly **Streamlit web app** was developed that allows:

- Uploading butterfly images
- Preprocessing and resizing
- Predicting species using the trained model
- Displaying actual label (from filename), prediction, and confidence %

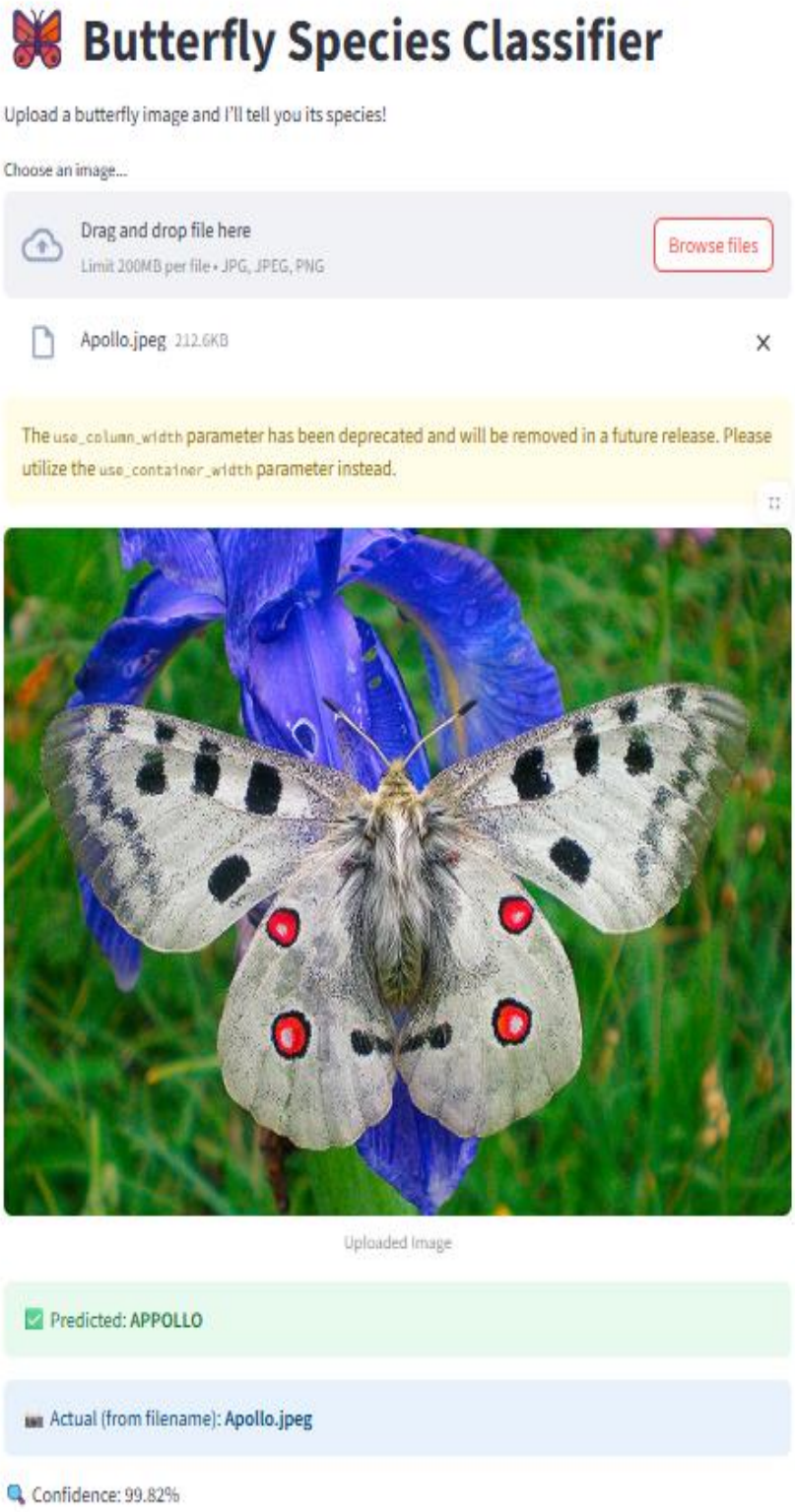
---

6. User Interface & Prediction Results

Image 1: Streamlit_Interface

Caption: *Figure 1 – Streamlit Upload Interface*

Image 2: APPOLLO Prediction

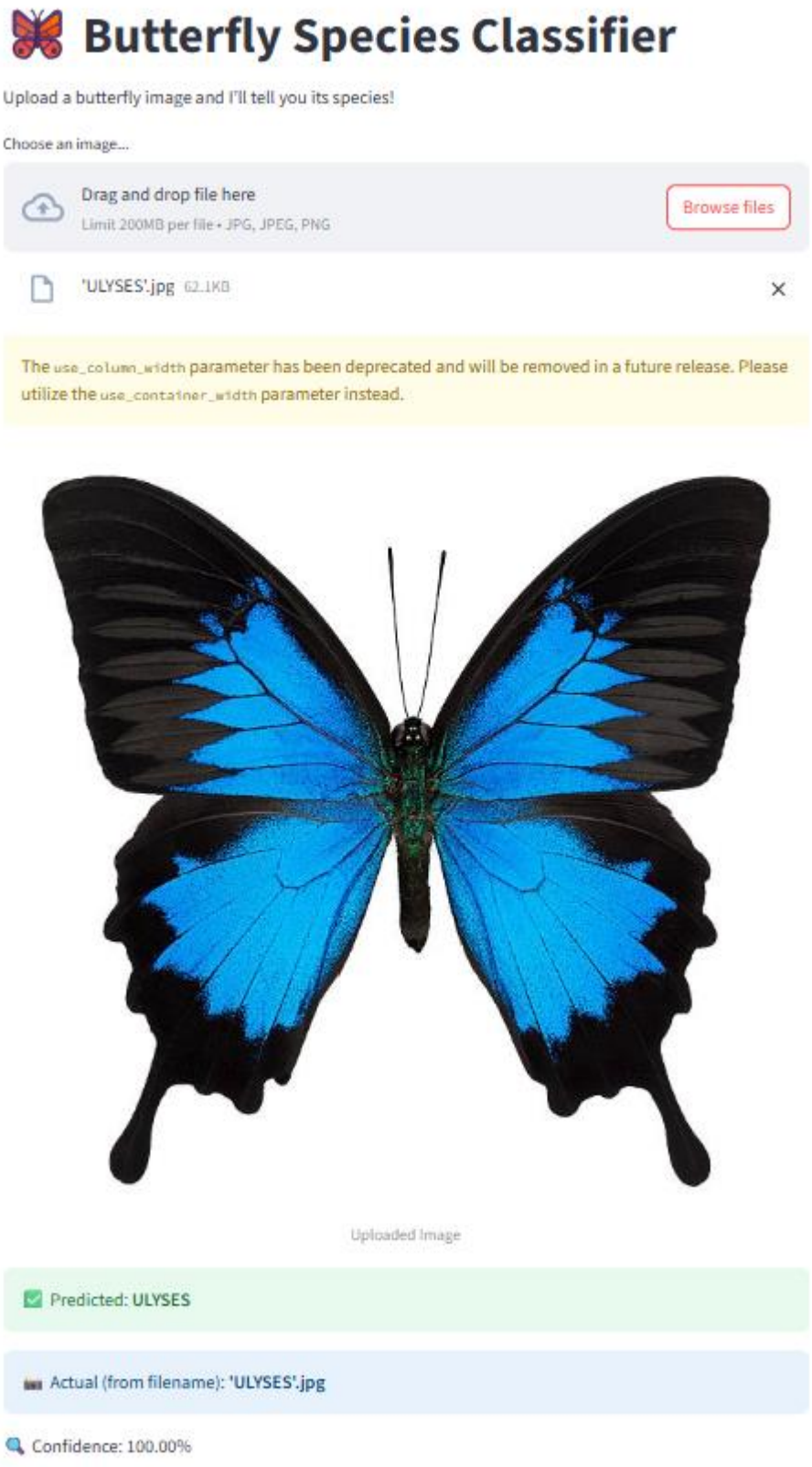


Caption: *Figure 2 – Predicted: APPOLLO (Confidence: 99.82%)*

Image 3: BROWN SIPROETA Prediction

Caption: *Figure 3 – Predicted: BROWN SIPROETA (Confidence: 99.93%)*

Image 4: ULYSES Prediction



Caption: *Figure 4 – Predicted: ULYSES (Confidence: 100.00%)*

7. Model Performance

📊 Training Metrics – MobileNetV2

Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
25	75.79%	0.8153	90.23%	0.3718
26	76.09%	0.7949	90.38%	0.3677
27	76.43%	0.7742	90.69%	0.3613
28	78.15%	0.7323	90.46%	0.3554
29	79.72%	0.7066	90.92%	0.3478
30	79.61%	0.6834	90.69%	0.3433

🐆 Training Metrics – EfficientNetB0

Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
15	18.17%	3.6662	43.08%	3.2458
16	21.15%	3.5740	38.15%	3.3022
17	24.46%	3.4608	47.38%	3.0858
18	24.39%	3.4062	49.85%	3.0413
19	27.81%	3.2866	54.08%	2.7980
20	29.47%	3.2038	57.08%	2.8518

📋 Model Comparison Summary

Model	Val Accuracy	Val Loss	Epoch Time	Remarks
MobileNetV2	90.69%	0.3433	~330 sec	Fast, highly accurate, deployable
EfficientNetB0	57.08%	2.8518	~900 sec	Underperformed, slower training time

---

8. Challenges & Future Work

- Class imbalance and visually similar species posed difficulties

- EfficientNetB0 underperformed due to overfitting and model complexity
- Future improvements include:
    - Expanding dataset and rare species representation
    - Testing advanced architectures (e.g., Vision Transformers)
    - Building a mobile-friendly version using TensorFlow Lite
    - Adding GPS/environmental data for ecological insight

---

9. Conclusion:

This project successfully demonstrates the use of transfer learning for butterfly species classification with high accuracy. MobileNetV2 outperformed more complex models like EfficientNetB0, making it ideal for real-world deployment. The project supports biodiversity tracking and has the potential to evolve into a powerful citizen science tool through further development.

---

10. Tools and Technologies Used:

- Python 3.10.0
- TensorFlow & Keras
- Streamlit
- Pandas & NumPy
- Jupyter Notebook