SPRING 2025

# Butterfly Image Classification With ResNet50

**By**
Juana Wong

**CSc 44700 P**
Introduction to Machine Learning

# Agenda

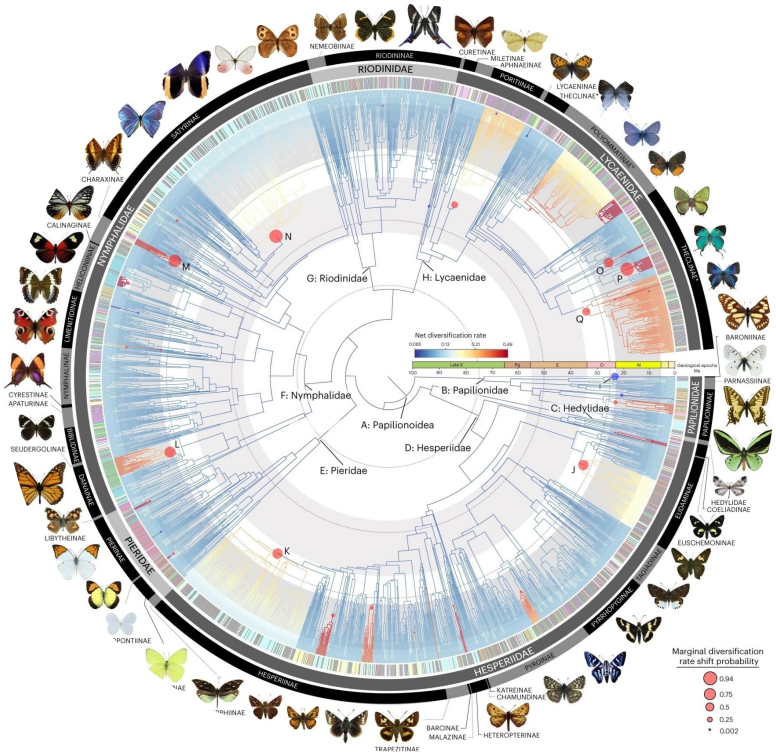
01 Problem & Dataset

02 Data Pipeline & Augmentation

03 Model Development & Training

04 Results, Evaluation & Discussion

# 01 Identify, Classify, Conserve, Support — Butterfly Populations




- approx. 17,500 to 20,000 butterfly species worldwide
- population crash!
  - U.S.A.: “107 species **declined** by **more than 50%**, and 22 species declined by **more than 90%**” from 2000–2020

Edwards, Collin B., et al. “Rapid Butterfly Declines across the United States during the 21st Century.” *Science*, vol. 387, no. 6738, 7 Mar. 2025, pp. 1090 – 1094, https://doi.org/10.1126/science.adp4671.

Rapidly and accurately identify butterflies to support large-scale **monitoring** of population trends & **conservation** initiatives.

Kawahara, Akito Y., et al. “A Global Phylogeny of Butterflies Reveals Their Evolutionary History, Ancestral Hosts and Biogeographic Origins.” *Nature Ecology & Evolution*, vol. 7, no. 7, 15 May 2023, pp. 1 – 11, https://doi.org/10.1038/s41559-023-02041-9.

# 01 “Butterfly Image Classification” (Version 2) Dataset

**DePie**

phucthaiv02 (he/him)

**Data Explorer**

Version 2 (237.31 MB)

- test
- train
- Testing_set.csv
- Training_set.csv

Thái Văn, Phúc (2024). Butterfly Image Classification (Version 2) [Dataset]. Kaggle. https://www.kaggle.com/datasets/phucthaiv02/butterfly-image-classification/data

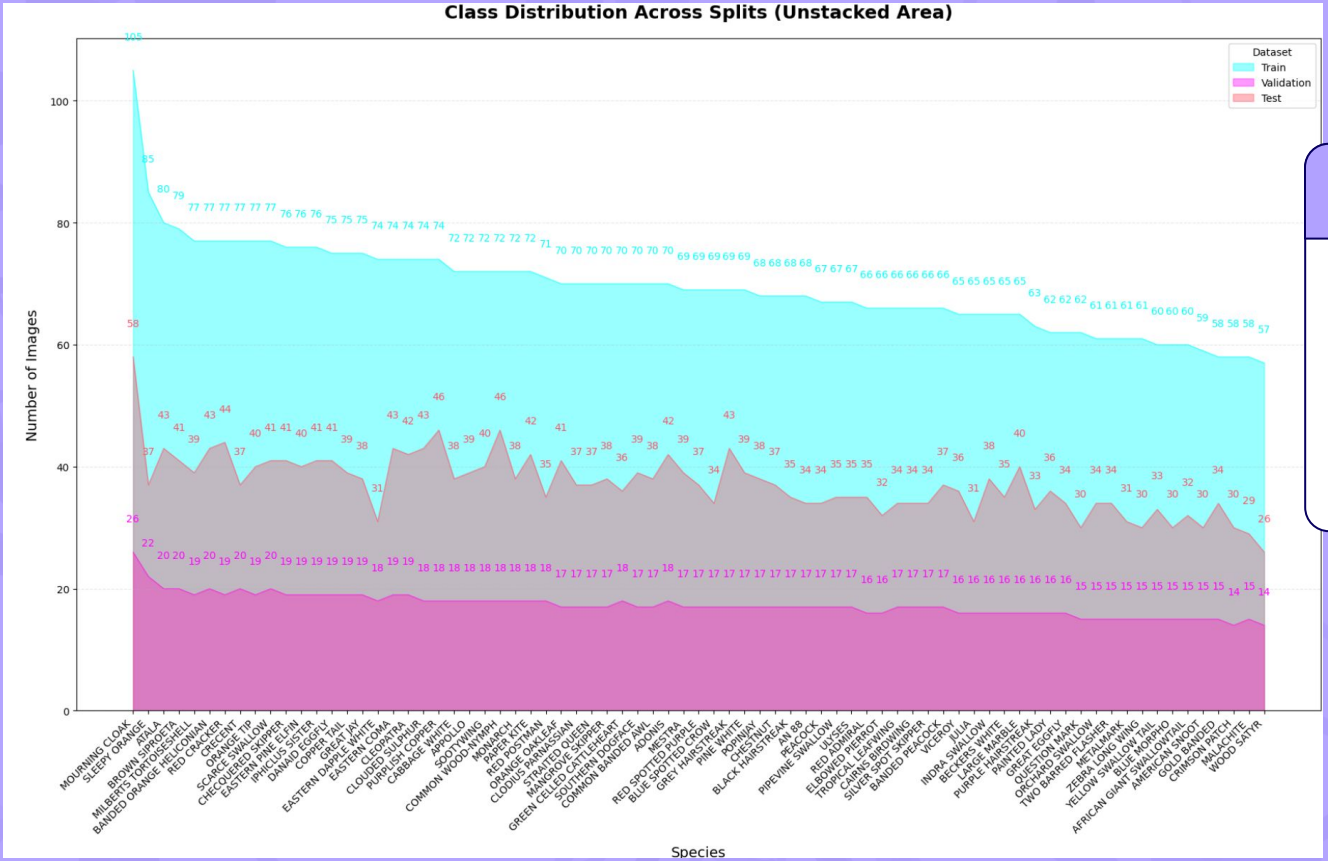
**Description**

- 9,000+ sample images
- 75 predefined classes/labels
- ‘train’ set – labeled (.csv)
- ‘test’ set – unlabeled

original ‘train’ dataset:
(‘train’, ‘val’): 80–20 split

train: 5,199 images
val: 1,300 images
test: 2,786 images

# 01 “Butterfly Image Classification” (Version 2) Dataset

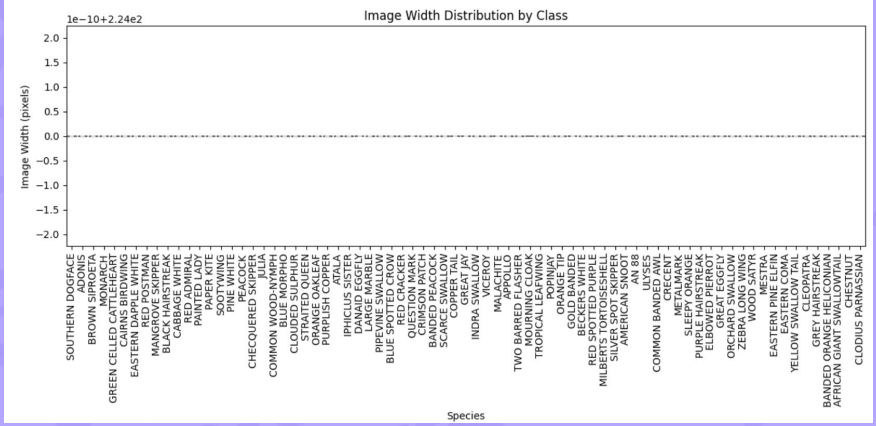


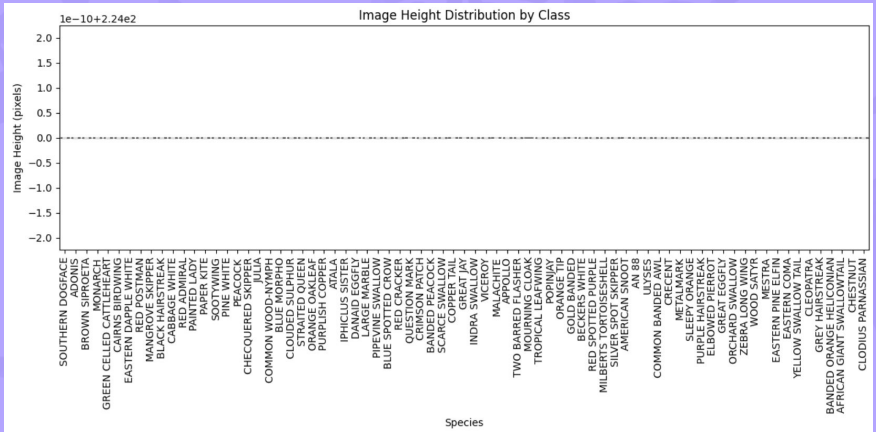

**Number of Images per Split/Set**

**Train**: 57–105 (avg. 86)
**Validation**: 14–26 (avg. 17)

**Test***: no way to verify if *predicted classes* are *true*, since data was originally unlabeled; 26–58 (avg. 37)

# 01 “Butterfly Image Classification” (Version 2) Dataset



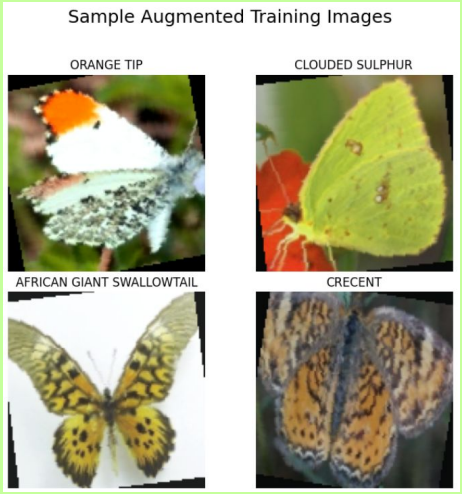

**Distribution of Image Dimensions in Train Dataset**

Quality check: No variance, no distribution in all original images.

Can proceed to **data preprocessing** and **data augmentation** phases.

02

# Data Pipeline & Augmentation




| Raw Images (JPG) | Data Loading | Data Preprocessing | Data Augmentation (Train) | DataLoader | Model Training |
|---|---|---|---|---|---|
| | read filepaths, csv | | | | |
| | split to train/val/test | | | | |
| | | Resize (128x128) | | | |
| | | Normalize using ImageNet mean/std | | | |
| | | | RandomResizedCrop | | |
| | | | RandomHorizontalFlip | | |
| | | | RandomRotation (15°) | | |
| | | | ColorJitter | | |
| | | | | Batch, shuffle, load | |
| | | | | | Simple Butterfly CNN |
| | | | | | ResNet50 |

To increase data diversity, improve generalization.
To overcome limitations of a small dataset.

# 03 Baseline Model & Transfer Learning Approach

## SimpleButterflyCNN

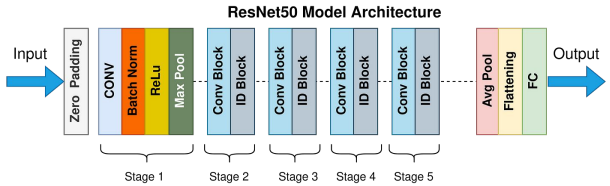
- CNN (4 conv layers, ReLU, max pooling), achieved ~73% val accuracy at 16/20 epochs
- Limitation: plateau from limited model capacity, underfitting

--- Final Metrics (SimpleButterflyCNN) ---
Final Training Loss: 0.8904
Final Validation Loss: 0.9954
Final Training Accuracy: 72.88%
Final Validation Accuracy: 73.54%
Best Validation Accuracy: 73.54% at epoch 19

## ResNet50

- Pretrained model on ImageNet (50 layers, 16 residual blocks)
- Final layer adapted for 75 classes
- Train/validation gap is small (96.7% vs. 91.2%), indicating strong generalization and minimal overfitting

--- Final Metrics (ResNet50) ---
Final Training Loss: 0.0926
Final Validation Loss: 0.4293
Final Training Accuracy: 96.69%
Final Validation Accuracy: 91.15%
Best Validation Accuracy: 91.77% at epoch 18
Best Validation Loss: 0.4074 at epoch 10




=== Training Summary (ResNet50) ===
Model: ResNet50 (pretrained on ImageNet, final layer for 75 classes)
Epochs: 20
Optimizer: Adam
Batch size: 32
Learning Rate Scheduler: StepLR (step_size=7, gamma=0.1)
Initial Learning rate: 0.001
Final Learning rate after 20 epochs: 0.00001 (1e-05)
Loss function: CrossEntropyLoss
Data augmentation: RandomResizedCrop, Flip, Rotation, ColorJitter
Device: CPU

# 04 Baseline Model vs. ResNet50 – Training and Validation Progress

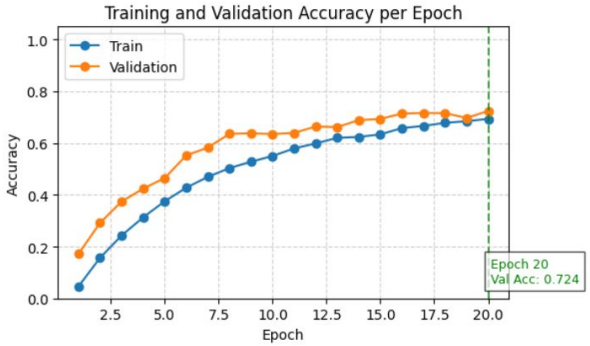
## Accuracy & Loss

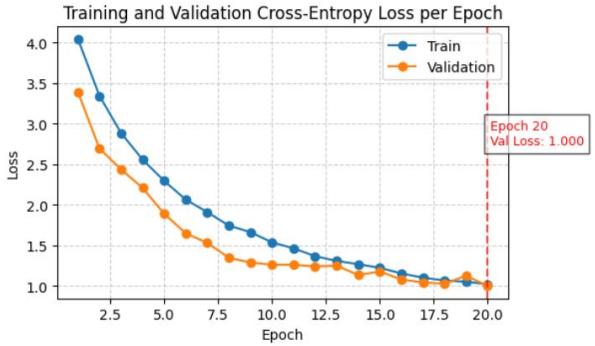
The SimpleButterflyCNN (baseline model) plateaued around 72% accuracy at epoch 20.

The ResNet50 model's accuracy and loss improved rapidly during the first 8–10 epochs, reaching over a 90% validation accuracy by epoch 10.

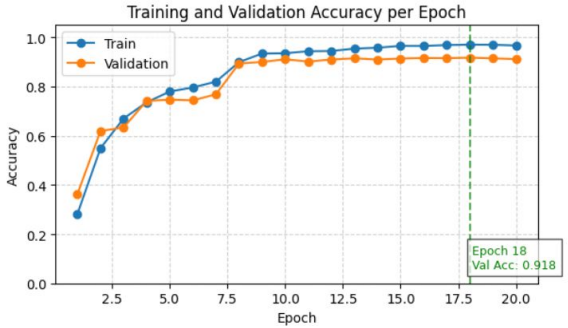
After that, both training and validation performance plateaued. Training beyond 10 epochs offered little benefit, so stopping earlier can be done to prevent overfitting.

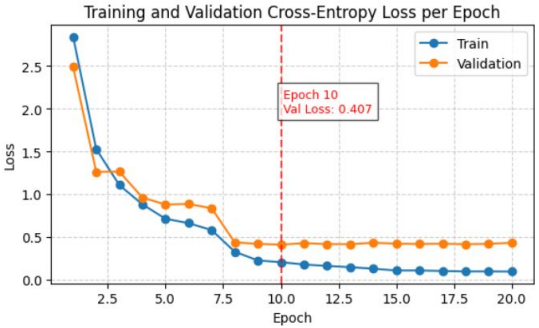
SimpleButterflyCNN




ResNet50



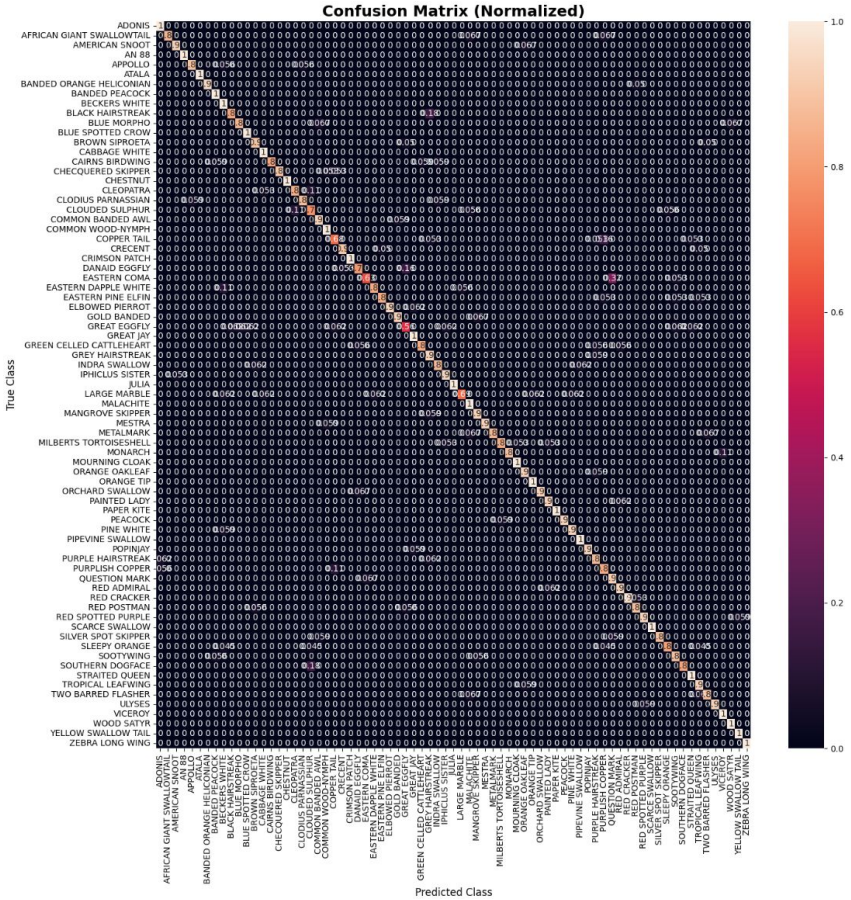
# 04 Classification Report – Heatmap (Precision, Recall, F1-Score, Support)

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| ADONIS | 0.900000 | 1.000000 | 0.947368 | 18 |
| AFRICAN GIANT SWALLOWTAIL | 1.000000 | 0.866667 | 0.928571 | 15 |
| AMERICAN SNOOT | 0.933333 | 0.933333 | 0.933333 | 15 |
| AN 88 | 1.000000 | 1.000000 | 1.000000 | 17 |
| APPOLLO | 0.941176 | 0.888889 | 0.914286 | 18 |
| ATALA | 1.000000 | 1.000000 | 1.000000 | 20 |
| BANDED ORANGE HELICONIAN | 1.000000 | 0.950000 | 0.974359 | 20 |
| BANDED PEACOCK | 0.894737 | 1.000000 | 0.944444 | 17 |
| BECKERS WHITE | 0.727273 | 1.000000 | 0.842105 | 16 |
| BLACK HAIRSTREAK | 0.933333 | 0.823529 | 0.875000 | 17 |
| BLUE MORPHO | 0.928571 | 0.866667 | 0.896552 | 15 |
| BLUE SPOTTED CROW | 0.944444 | 1.000000 | 0.971429 | 17 |
| BROWN SIPROETA | 0.900000 | 0.900000 | 0.900000 | 20 |
| CABBAGE WHITE | 0.900000 | 1.000000 | 0.947368 | 18 |
| CAIRNS BIRDWING | 1.000000 | 0.823529 | 0.903226 | 17 |
| CHECQUERED SKIPPER | 1.000000 | 0.894737 | 0.944444 | 19 |
| CHESTNUT | 1.000000 | 1.000000 | 1.000000 | 17 |
| CLEOPATRA | 0.888889 | 0.842105 | 0.864865 | 19 |
| CLODIUS PARNASSIAN | 0.937500 | 0.882353 | 0.909091 | 17 |
| CLOUDED SULPHUR | 0.700000 | 0.777778 | 0.736842 | 18 |
| COMMON BANDED AWL | 0.888889 | 0.941176 | 0.914286 | 17 |
| COMMON WOOD-NYMPH | 0.900000 | 1.000000 | 0.947368 | 18 |
| COPPER TAIL | 0.764706 | 0.684211 | 0.722222 | 19 |
| CRECENT | 0.947368 | 0.900000 | 0.923077 | 20 |
| CRIMSON PATCH | 1.000000 | 1.000000 | 1.000000 | 14 |
| DANAID EGGFLY | 0.882353 | 0.789474 | 0.833333 | 19 |
| EASTERN COMA | 0.923077 | 0.631579 | 0.750000 | 19 |
| EASTERN DAPPLE WHITE | 0.937500 | 0.833333 | 0.882353 | 18 |
| EASTERN PINE ELFIN | 0.941176 | 0.842105 | 0.888889 | 19 |
| ELBOWED PIERROT | 1.000000 | 0.937500 | 0.967742 | 16 |
| GOLD BANDED | 0.933333 | 0.933333 | 0.933333 | 15 |
| GREAT EGGFLY | 0.642857 | 0.562500 | 0.600000 | 16 |
| GREAT JAY | 0.904762 | 1.000000 | 0.950000 | 19 |
| GREEN CELLED CATTLEHEART | 0.937500 | 0.833333 | 0.882353 | 18 |
| GREY HAIRSTREAK | 0.727273 | 0.941176 | 0.820513 | 17 |
| INDRA SWALLOW | 0.875000 | 0.875000 | 0.875000 | 16 |
| IPHICLUS SISTER | 0.900000 | 0.947368 | 0.923077 | 19 |
| JULIA | 1.000000 | 1.000000 | 1.000000 | 16 |
| LARGE MARBLE | 0.916667 | 0.687500 | 0.785714 | 16 |
| MALACHITE | 0.789474 | 1.000000 | 0.882353 | 15 |
| MANGROVE SKIPPER | 0.888889 | 0.941176 | 0.914286 | 17 |
| MESTRA | 1.000000 | 0.941176 | 0.969697 | 17 |
| METALMARK | 1.000000 | 0.866667 | 0.928571 | 15 |
| MILBERTS TORTOISESHELL | 0.941176 | 0.842105 | 0.888889 | 19 |
| MONARCH | 1.000000 | 0.888889 | 0.941176 | 18 |
| MOURNING CLOAK | 0.962963 | 1.000000 | 0.981132 | 26 |
| ORANGE OAKLEAF | 0.888889 | 0.941176 | 0.914286 | 17 |
| ORANGE TIP | 0.950000 | 1.000000 | 0.974359 | 19 |
| ORCHARD SWALLOW | 1.000000 | 0.933333 | 0.965517 | 15 |
| PAINTED LADY | 0.882353 | 0.937500 | 0.909091 | 16 |
| PAPER KITE | 1.000000 | 1.000000 | 1.000000 | 18 |
| PEACOCK | 1.000000 | 0.941176 | 0.969697 | 17 |
| PINE WHITE | 0.941176 | 0.941176 | 0.941176 | 17 |
| PIPEVINE SWALLOW | 0.944444 | 1.000000 | 0.971429 | 17 |
| POPINJAY | 1.000000 | 0.941176 | 0.969697 | 17 |
| PURPLE HAIRSTREAK | 0.777778 | 0.875000 | 0.823529 | 16 |
| PURPLISH COPPER | 0.714286 | 0.833333 | 0.769231 | 18 |
| QUESTION MARK | 0.666667 | 0.933333 | 0.777778 | 15 |
| RED ADMIRAL | 0.882353 | 0.937500 | 0.909091 | 16 |
| RED CRACKER | 1.000000 | 0.947368 | 0.972973 | 19 |
| RED POSTMAN | 0.888889 | 0.888889 | 0.888889 | 18 |
| RED SPOTTED PURPLE | 0.941176 | 0.941176 | 0.941176 | 17 |
| SCARCE SWALLOW | 1.000000 | 1.000000 | 1.000000 | 20 |
| SILVER SPOT SKIPPER | 1.000000 | 0.882353 | 0.937500 | 17 |
| SLEEPY ORANGE | 0.947368 | 0.818182 | 0.878049 | 22 |
| SOOTYWING | 0.842105 | 0.888889 | 0.864865 | 18 |
| SOUTHERN DOGFACE | 1.000000 | 0.823529 | 0.903226 | 17 |
| STRAITED QUEEN | 0.894737 | 1.000000 | 0.944444 | 17 |
| TROPICAL LEAFWING | 0.800000 | 0.941176 | 0.864865 | 17 |
| TWO BARRED FLASHER | 0.866667 | 0.866667 | 0.866667 | 15 |
| ULYSES | 1.000000 | 0.941176 | 0.969697 | 17 |
| VICEROY | 0.888889 | 1.000000 | 0.941176 | 16 |
| WOOD SATYR | 0.933333 | 1.000000 | 0.965517 | 14 |
| YELLOW SWALLOW TAIL | 0.937500 | 1.000000 | 0.967742 | 15 |
| ZEBRA LONG WING | 1.000000 | 1.000000 | 1.000000 | 15 |

## Classification Report – Summary

Overall, **precision**, **recall**, and **F1-score** indicate high robust performance and good generalization by the **ResNet50**. Some specific classes struggle.

# 04 Normalized Confusion Matrix (Recall)



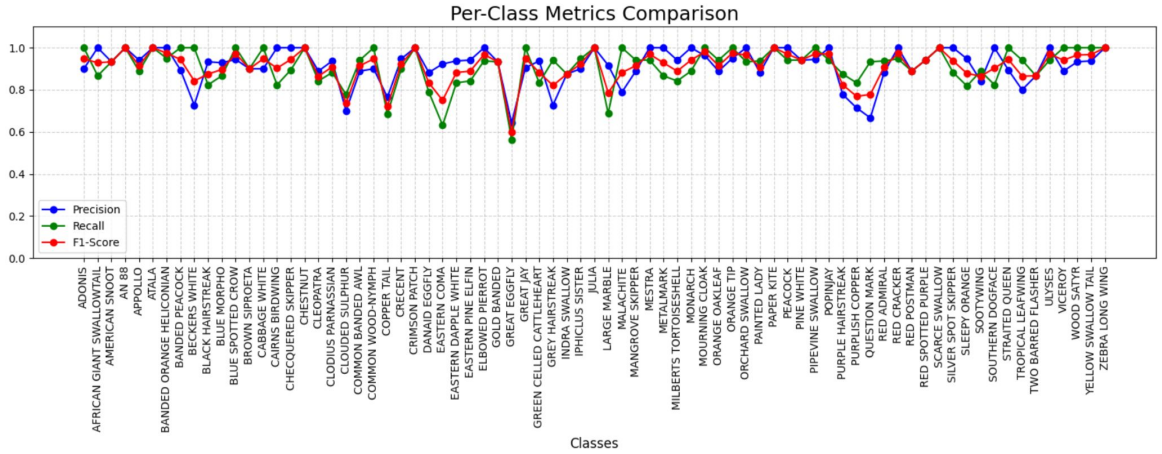

## Normalized Confusion Matrix

Most butterfly classes show high recall (0.9+) – the model correctly identifies the majority of samples for most classes.

Several classes achieved perfect recall (1.0) – no misclassifications for those categories.

A few low-performing classes (<0.8) – the model struggles to distinguish these classes from others (misclassification).

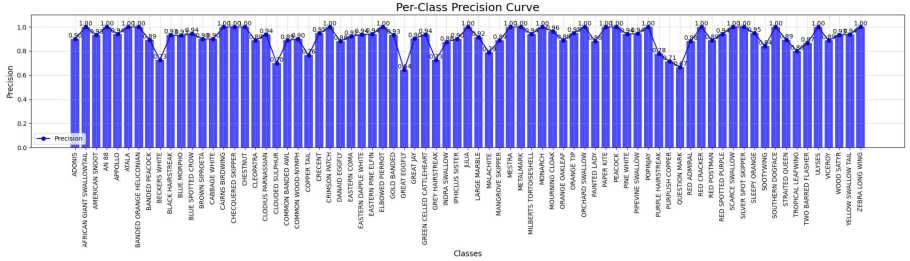
# 04 Per-Class Performance Metrics – Precision, Recall, F1-Score

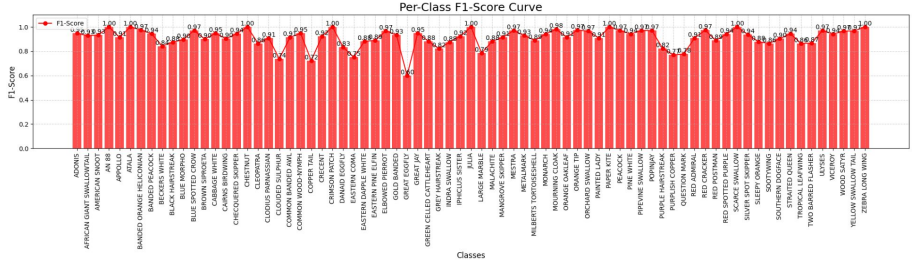


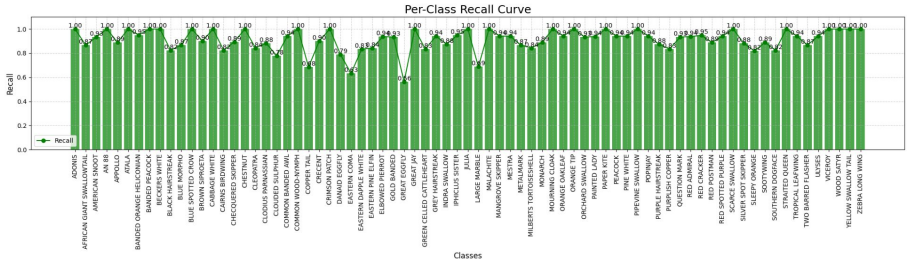

**ResNet50 Per-Metric Comparison (F1-Score)**

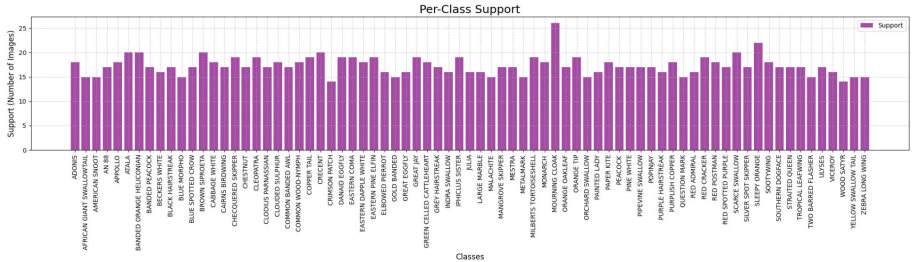
**Overall Accuracy**: 0.91 (91%)
**Highest Performing Classes** (8): 1.0 (100%)
**Lowest Performing Class** (1): 0.6 (60%)
**Average Classes**: Between 0.85 and 0.97 – indicates strong and consistent performance across most categories.

# 04 Per-Class Performance Metrics – Precision, Recall, F1-Score



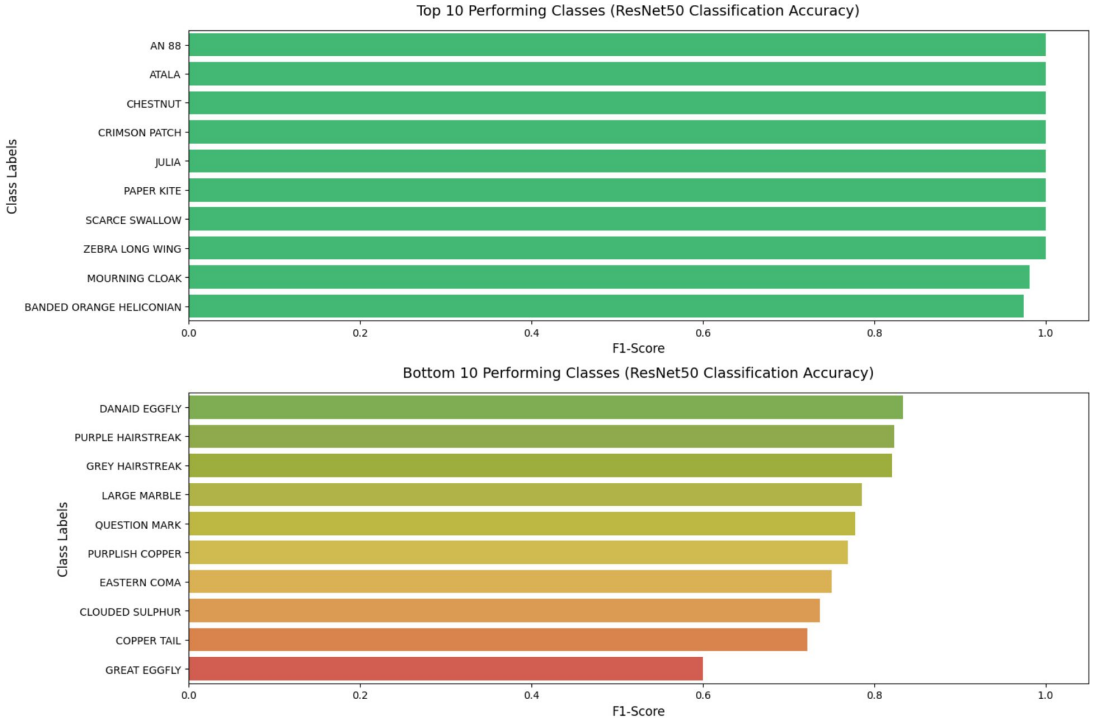




**ResNet50 Per-Metric Comparison**

Viewed altogether, or by each per-metric measure independently, **most classes exhibit high accuracy performances**–with *consistent* measures in all graphs. Specific classes struggle in *all three* measures.

# 04 Classification Performance/Accuracy Rankings




## Top 10 & Bottom 10

**Train**: avg. 86 images per class

**Validation**: avg. 17 images per class

Some classes had perfect 100% classification accuracy.

Some classes had moderate accuracy when classifying species.

One class, the **Great Eggfly**, noticeably performed the worst.

# 04 Great Eggfly (0.6 F1–Score)

## Great Eggfly

**Precision**: 0.64 – 64% of images predicted as Great Eggfly was correct

**Recall**: 0.56 – of all Great Eggfly images, 56% were correctly predicted

**F1–Score**: 0.60 – harmonic mean of precision and recall; moderate/average score – can be improved

**Support**: 16 – small sample size, 16 true images in the set

The **GREAT EGGFLY** may closely resemble other species.

# 04 Great Eggfly (0.6 F1-Score)

## Great Eggfly

**Precision**: 0.64 **Recall**: 0.56 **F1-Score**: 0.60 **Support**: 16

## Incorrect Prediction

Sootywing

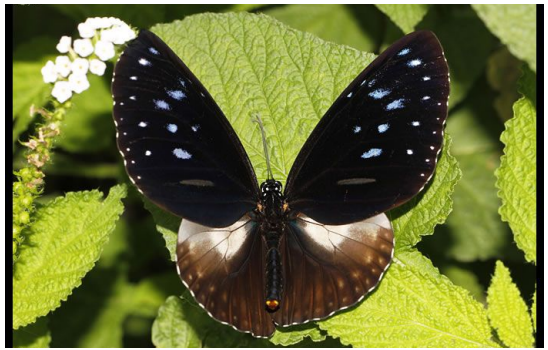
# 04 Great Eggfly (0.6 F1-Score)

## Great Eggfly

**Precision**: 0.64 **Recall**: 0.56 **F1-Score**: 0.60 **Support**: 16

## Incorrect Prediction

Blue Spotted Crow



# 04 Great Eggfly (0.6 F1-Score)

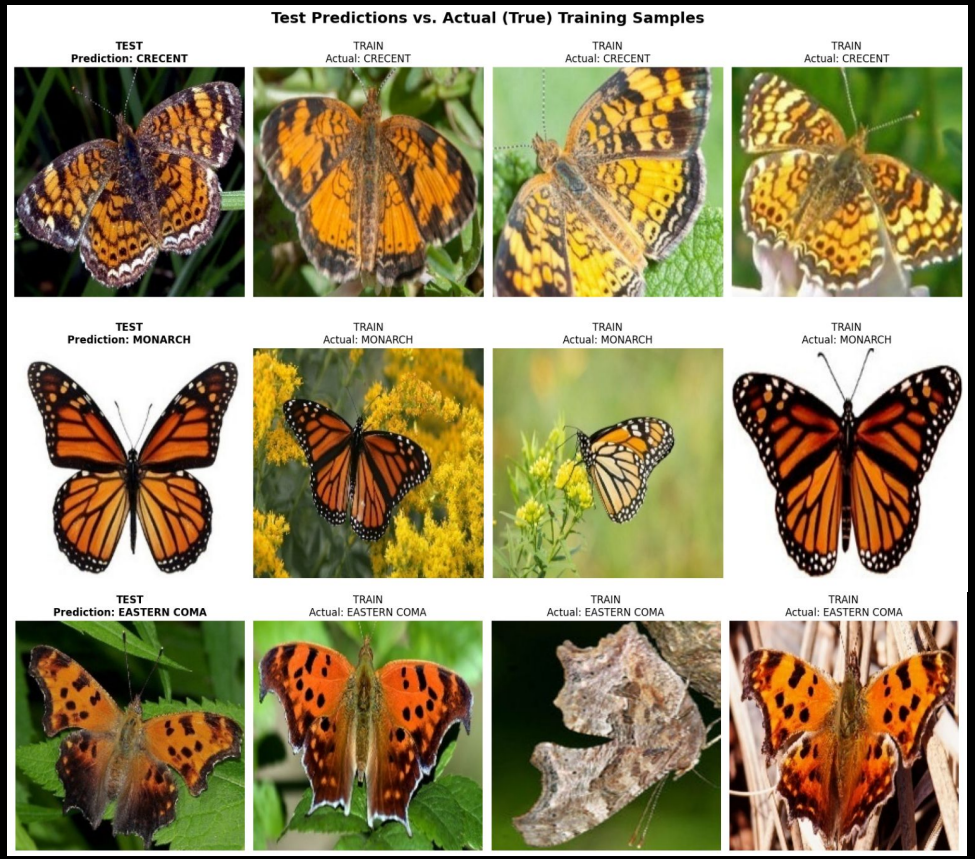
## Great Eggfly

**Precision**: 0.64 **Recall**: 0.56 **F1-Score**: 0.60 **Support**: 16

## Incorrect Prediction

Iphiclus Sister

# 04 Example – Test Predictions (Unverified/Unverifiable)




## Crecent

**Precision**: 0.95 **Recall**: 0.90
**F1-Score**: 0.92 **Support**: 20

## Monarch

**Precision**: 1.00 **Recall**: 0.89
**F1-Score**: 0.94 **Support**: 18

## Eastern Coma

**Precision**: 0.92 **Recall**: 0.63
**F1-Score**: 0.75 **Support**: 19

of actual EASTERN COMA butterfly images, it struggles to detect & identify

# 04 Discussion & Areas of Improvement



## Collect more data

Increase support, more diverse examples reduces false positives and false negatives

## Adjust model values

Eg. adjust model to favor precision over recall, or vice versa

## Data augmentation

Apply different transformations and adjustments to sample images, improve generalization

## Improve model architecture

Adjust depth and width of CNN or try different architectures, eg. custom CNN, ResNet50, etc.

## Hyperparameter tuning

Change learning rate, batch size, optimizer, number of epochs, etc.

Edwards, Collin B., et al. “Rapid Butterfly Declines across the United States during the 21st Century.” *Science*, vol. 387, no. 6738, 7 Mar. 2025, pp. 1090 – 1094, https://doi.org/10.1126/science.adp4671.

Kawahara, Akito Y., et al. “A Global Phylogeny of Butterflies Reveals Their Evolutionary History, Ancestral Hosts and Biogeographic Origins.” *Nature Ecology & Evolution*, vol. 7, no. 7, 15 May 2023, pp. 1 – 11, https://doi.org/10.1038/s41559-023-02041-9.

Thái Văn, Phúc (2024). Butterfly Image Classification (Version 2) [Dataset]. Kaggle. https://www.kaggle.com/datasets/phucthaiv02/butterfly-image-classification/data