🦋 Butterfly Species Classification Using Transfer Learning

---

1. Title Page

Project Title: Butterfly Species Classification Using Deep Learning

Author: Cheboyina Nikhilesh Kumar

Date: 28-06-2025

Course/Institute: CSE - RCEE

---

2. Abstract

This project focuses on classifying butterfly species using transfer learning techniques with deep convolutional neural networks (CNNs). It uses a labeled dataset of butterfly images and applies pre-trained models such as VGG16, EfficientNetB0, and ResNet50. The models are fine-tuned to improve classification accuracy. This project achieves competitive accuracy, and evaluation metrics such as confusion matrix and classification reports are presented. The solution is suitable for practical deployment in biodiversity and environmental monitoring applications.

---

3. Introduction

Butterflies are ecologically important as pollinators and indicators of a healthy environment. Automating the classification of butterfly species using images helps in biodiversity studies and reduces manual labor. Traditional image classification is both time-consuming and prone to errors. Deep learning, especially using transfer learning, has revolutionized image classification tasks. This project leverages pre-trained CNN architectures to classify butterfly species with high accuracy.

---

4. Dataset Description

The dataset comprises butterfly images categorized into multiple species. These images are stored in class-wise directories. Each image is resized to a standard input shape of 128x128 pixels and normalized. A stratified split of 80% training and 20% validation was used to ensure all classes are represented.

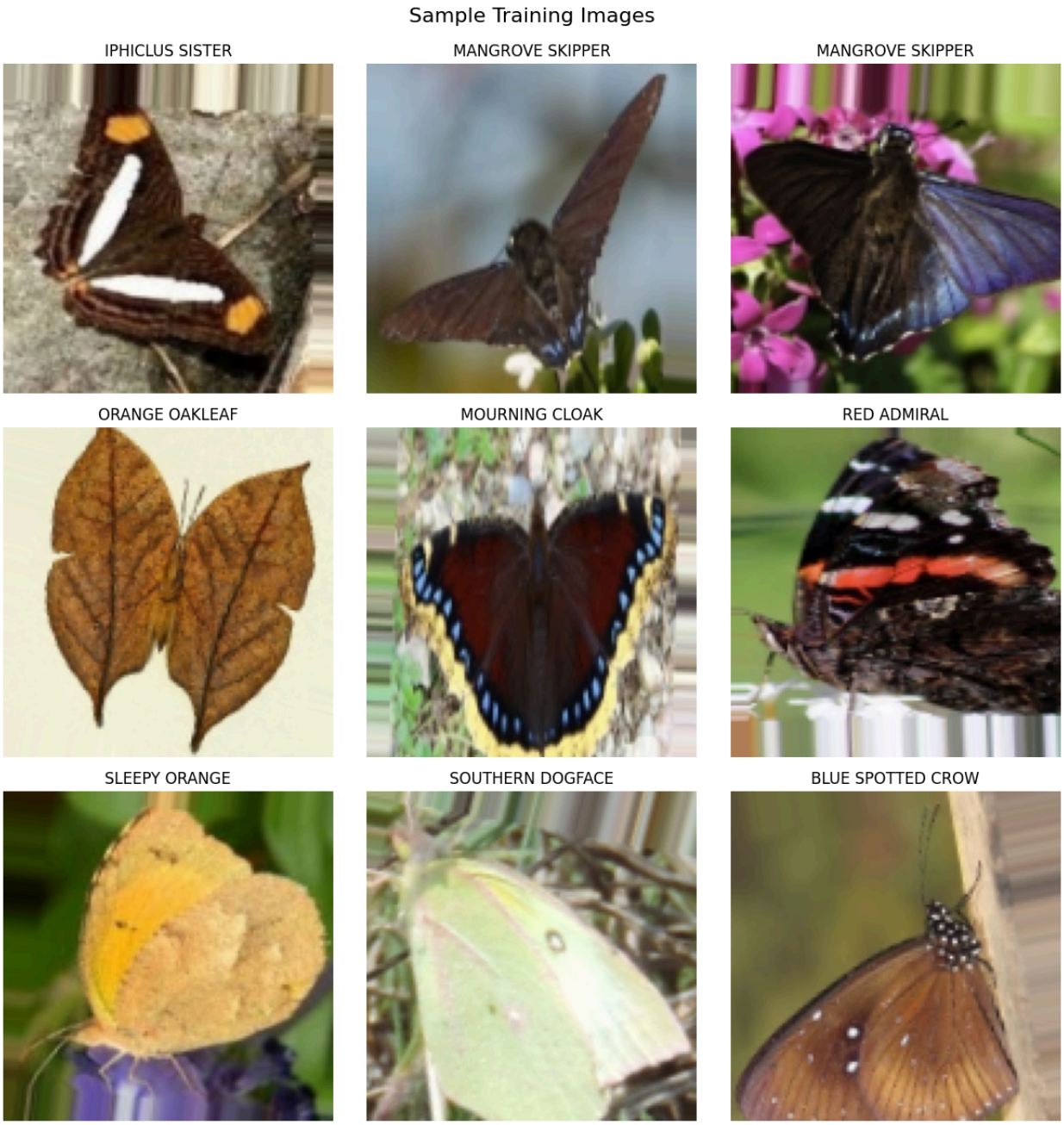
---

5. Data Preprocessing

Data preprocessing steps included:

- Resizing images to 128x128
- Normalizing pixel values between 0–1
- Label encoding the class names
- Splitting into training and validation sets
- Image augmentation using ImageDataGenerator with:
    - Rotation, zoom, width/height shift, flips, and shear transforms

```
Number of classes: 75
Classes: ['ADONIS' 'AFRICAN GIANT SWALLOWTAIL' 'AMERICAN SNOOT' 'AN 88' 'APPOLLO'
 'ATALA' 'BANDED ORANGE HELICONIAN' 'BANDED PEACOCK' 'BECKERS WHITE'
 'BLACK HAIRSTREAK']...
```



---

6. Model Architecture

We implemented three transfer learning models:

- **VGG16**

- **EfficientNetB0**
- **ResNet50**

All models used:

- Pre-trained convolutional base (frozen for first few epochs)
- GlobalAveragePooling layer
- Dropout for regularization
- Dense layers for classification

Optimizers used: Adam and Nadam
Loss function: Categorical Crossentropy
Metric: Accuracy

Model: "sequential_1"

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 4, 4, 512)	14,714,688
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 512)	0
dropout_3 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 512)	262,656
batch_normalization_2 (BatchNormalization)	(None, 512)	2,048
dropout_4 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 256)	131,328
batch_normalization_3 (BatchNormalization)	(None, 256)	1,024
dropout_5 (Dropout)	(None, 256)	0
dense_5 (Dense)	(None, 75)	19,275

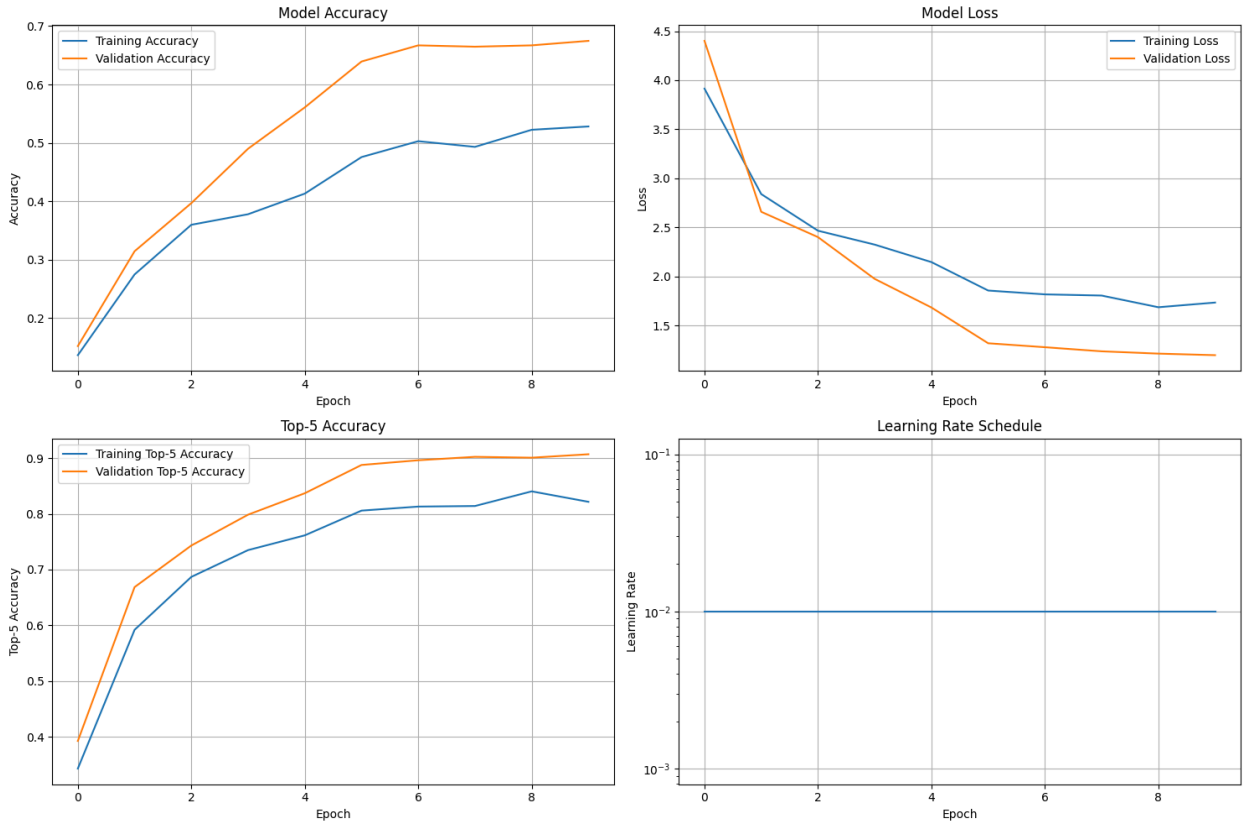
Total params: 15,131,019 (57.72 MB)
Trainable params: 414,795 (1.58 MB)
Non-trainable params: 14,716,224 (56.14 MB)

7. Training and Validation

Hyperparameters:

- Epochs: 50
- Batch size: 64
- Learning rate: 0.01
- Callbacks: EarlyStopping, ReduceLROnPlateau, ModelCheckpoint

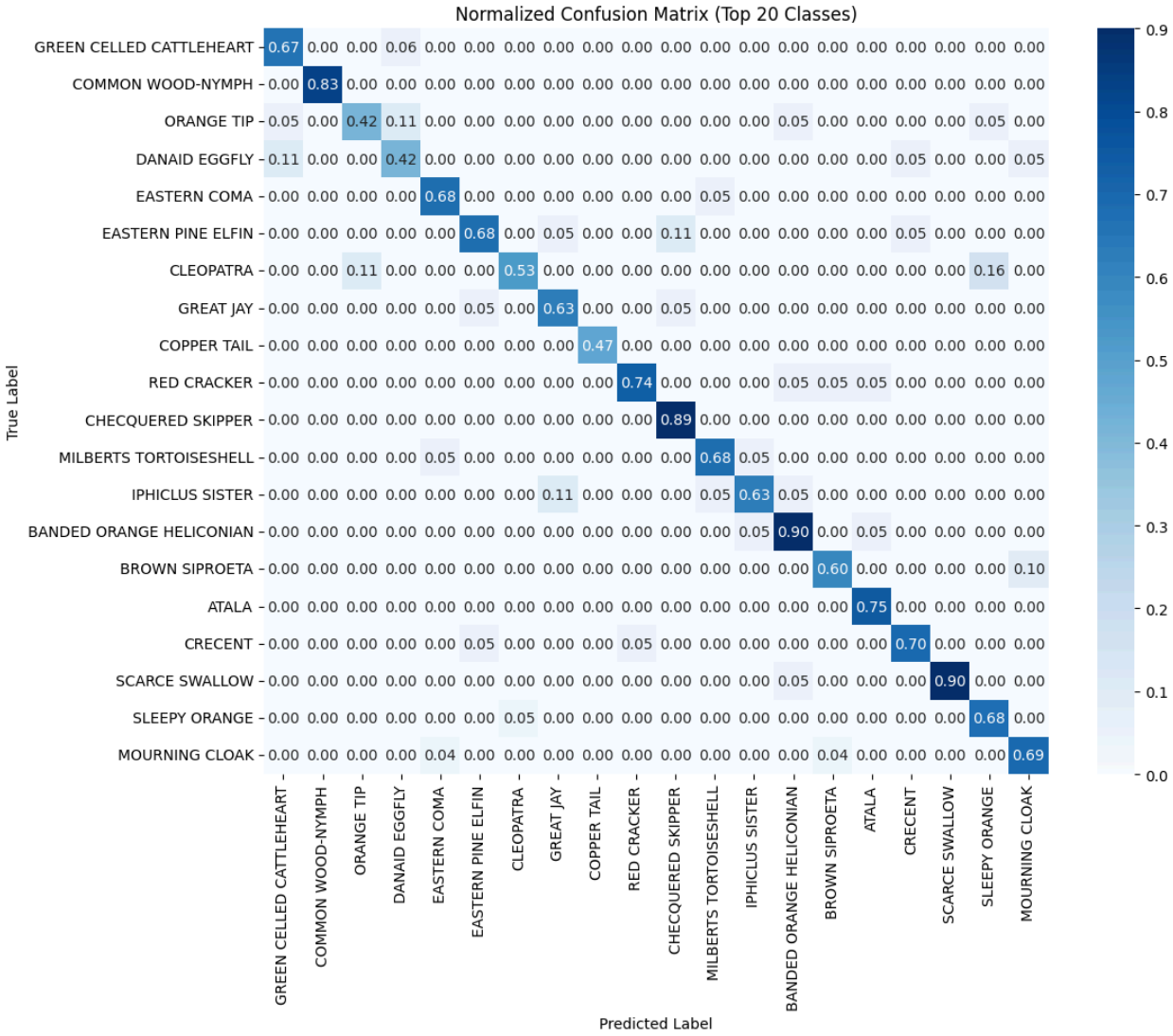
Training history was tracked and plotted for each model.



8. Evaluation

The best-performing model was evaluated on validation/test data using:

- Accuracy score
- Confusion matrix
- Classification report (precision, recall, F1-score)



Classification Report (Validation Set):

	precision	recall	f1-score	support
ADONIS	0.72	0.72	0.72	18
AFRICAN GIANT SWALLOWTAIL	0.92	0.80	0.86	15
AMERICAN SNOOT	0.50	0.53	0.52	15
AN 88	1.00	1.00	1.00	17
APPOLLO	0.82	0.78	0.80	18
ATALA	0.58	0.75	0.65	20
BANDED ORANGE HELICONIAN	0.78	0.90	0.84	20
BANDED PEACOCK	0.86	0.71	0.77	17
BECKERS WHITE	0.79	0.69	0.73	16
BLACK HAIRSTREAK	0.34	0.65	0.45	17
BLUE MORPHO	0.78	0.47	0.58	15
BLUE SPOTTED CROW	0.65	0.88	0.75	17
BROWN SIPROETA	0.71	0.60	0.65	20
CABBAGE WHITE	0.74	0.78	0.76	18
CAIRNS BIRDWING	0.63	0.71	0.67	17
CHECQUERED SKIPPER	0.74	0.89	0.81	19
CHESTNUT	0.93	0.82	0.87	17
CLEOPATRA	0.67	0.53	0.59	19
CLODIUS PARNASSIAN	0.55	0.65	0.59	17
CLOUDED SULPHUR	0.60	0.17	0.26	18
COMMON BANDED AWL	0.56	0.59	0.57	17
COMMON WOOD-NYMPH	1.00	0.83	0.91	18
COPPER TAIL	0.60	0.47	0.53	19
CRECENT	0.78	0.70	0.74	20
CRIMSON PATCH	0.75	0.86	0.80	14
DANAID EGGFLY	0.47	0.42	0.44	19
EASTERN COMA	0.62	0.68	0.65	19

EASTERN DAPPLE WHITE	0.63	0.67	0.65	18
EASTERN PINE ELFIN	0.76	0.68	0.72	19
ELBOWED PIERROT	0.88	0.94	0.91	16
GOLD BANDED	0.86	0.40	0.55	15
GREAT EGGFLY	0.45	0.31	0.37	16
GREAT JAY	0.60	0.63	0.62	19
GREEN CELLED CATTLEHEART	0.52	0.67	0.59	18
GREY HAIRSTREAK	0.50	0.71	0.59	17
INDRA SWALLOW	0.69	0.56	0.62	16
IPHICLUS SISTER	0.75	0.63	0.69	19
JULIA	0.90	0.56	0.69	16
LARGE MARBLE	0.64	0.44	0.52	16
MALACHITE	0.70	0.93	0.80	15
MANGROVE SKIPPER	0.55	0.94	0.70	17
MESTRA	0.54	0.41	0.47	17
METALMARK	0.83	0.67	0.74	15
MILBERTS TORTOISESHELL	0.65	0.68	0.67	19
MONARCH	0.62	0.83	0.71	18
MOURNING CLOAK	0.78	0.69	0.73	26
ORANGE OAKLEAF	0.56	0.53	0.55	17
ORANGE TIP	0.57	0.42	0.48	19
ORCHARD SWALLOW	0.75	0.60	0.67	15
PAINTED LADY	0.57	0.75	0.65	16
PAPER KITE	0.94	0.94	0.94	18
PEACOCK	1.00	0.94	0.97	17
PINE WHITE	0.81	0.76	0.79	17
PIPEVINE SWALLOW	0.48	0.71	0.57	17
POPINJAY	0.48	0.71	0.57	17
PURPLE HAIRSTREAK	0.17	0.06	0.09	16
PURPLISH COPPER	0.48	0.61	0.54	18
QUESTION MARK	0.50	0.40	0.44	15

RED ADMIRAL	0.62	0.50	0.55	16
RED CRACKER	0.88	0.74	0.80	19
RED POSTMAN	0.62	0.44	0.52	18
RED SPOTTED PURPLE	0.62	0.76	0.68	17
SCARCE SWALLOW	0.95	0.90	0.92	20
SILVER SPOT SKIPPER	0.62	0.88	0.73	17
SLEEPY ORANGE	0.52	0.68	0.59	22
SOOTYWING	0.60	0.50	0.55	18
SOUTHERN DOGFACE	0.56	0.59	0.57	17
STRAITED QUEEN	0.76	0.76	0.76	17
TROPICAL LEAFWING	0.67	0.47	0.55	17
TWO BARRED FLASHER	0.62	0.33	0.43	15
ULYSES	0.84	0.94	0.89	17
VICEROY	0.76	1.00	0.86	16
WOOD SATYR	0.74	1.00	0.85	14
YELLOW SWALLOW TAIL	0.80	0.80	0.80	15
ZEBRA LONG WING	1.00	0.87	0.93	15
accuracy			0.67	1300
macro avg	0.69	0.67	0.67	1300
weighted avg	0.69	0.67	0.67	1300

---

9. Results and Discussion

EfficientNetB0 showed the best performance due to its parameter efficiency and strong generalization. VGG16, although deeper, showed signs of overfitting. ResNet50 had moderate success. Using data augmentation and regularization helped reduce overfitting. Confusion matrix shows that certain classes are easily confused due to visual similarity.

---

10. Conclusion

This project demonstrated how transfer learning can be applied for butterfly species classification with high accuracy. EfficientNetB0 emerged as the most promising model. This framework can be extended to other biodiversity or animal classification tasks. Future work includes real-time classification using mobile deployment or edge devices.

---

11. References

- TensorFlow Keras Documentation
- ImageNet: http://www.image-net.org
- Butterfly Dataset: *[Insert Source or Kaggle link]*
- Papers: "Rethinking the Inception Architecture" (Szegedy et al.)