Deep Learning & Applications

UEC 642

A Project Report on:

Fine-Grained Visual Classification of 100 Butterfly Species Using Modern ConvNeXt Architectures

Submitted by:

Dharman Singh Sidhu (102215012)

Sanchit Gupta (102215026)

Ramit Singh (102215068)

Arshit Sood (102215076)

Submitted to:

Dr. Gaganpreet Kaur,

ECED

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

August - December, 2025

Abstract

The classification of butterfly species is a challenging "Fine-Grained Visual Classification" (FGVC) task due to high inter-class similarity (mimicry) and subtle variation in wing patterns. While traditional Convolutional Neural Networks (CNNs) have established strong baselines, they often lack the global context required to distinguish between morphologically similar species. This project proposes the use of ConvNeXtTiny, a modernized architecture that integrates Vision Transformer (ViT) design principles into a pure CNN framework. Utilizing a dataset of 13,594 images across 100 species, we implemented a memory-efficient streaming pipeline and a two-phase transfer learning strategy. Experimental results demonstrate that the ConvNeXt model achieves superior accuracy compared to traditional architectures like ResNet50 and VGG19, effectively validating the efficacy of modernizing CNNs for biodiversity monitoring tasks.

1. Introduction

Biodiversity monitoring is critical for assessing ecosystem health, yet manual identification of insect species is labor-intensive and requires expert taxonomical knowledge. In the domain of computer vision, this problem is categorized as Fine-Grained Visual Classification (FGVC). Unlike generic object detection (e.g., "cat" vs. "dog"), FGVC requires a model to distinguish between sub-categories (e.g., *Danaus plexippus* vs. *Danaus gilippus*) where visual differences are minute and often restricted to specific wing patterns or textures.

Traditional Deep Learning approaches have relied heavily on standard CNNs such as VGG16 and ResNet50. While effective, these models often struggle with the "local vs. global" feature trade-off. Recently, Vision Transformers (ViTs) have shown promise by using self-attention mechanisms to capture global context, but they typically require massive datasets to converge. This project bridges this gap by deploying ConvNeXt, a hybrid architecture introduced in 2022 that modernizes standard ResNets to compete with Transformers. By applying a two-phase transfer learning approach on the "Butterfly Images 40 Species" dataset (containing 100 classes), we aim to build a robust, high-accuracy classifier that remains computationally efficient enough for deployment on standard hardware.

2. Literature Survey

Research in fine-grained butterfly classification has recently shifted from standard CNNs towards hybrid, ensemble, and attention-based models to tackle the subtle differences between species.

2.1 Summary of Recent Findings

Recent advancements in fine-grained visual classification (FGVC) for Lepidoptera have surpassed the capabilities of standard CNNs. Early benchmarks by Yadav et al. (2023) established a strong baseline of ~95% using ResNet50 on the 100-species dataset. However, subsequent studies like Cao et al. (2024) and Syamsudin (2024) highlighted the limitations of pure CNNs in capturing subtle inter-class variations, proposing ensemble methods and EfficientNet variants that reached ~97.9% on smaller subsets. Most recently, Jin et al. (2024) introduced multi-feature enhancement techniques achieving 96.8%, though often at the cost of higher computational complexity. Our work builds on Liu et al. (2025), who demonstrated the efficacy of ConvNeXt for FGVC, by implementing a lightweight ConvNeXtTiny model that balances the global receptive fields of Transformers with the inductive bias of CNNs.

2.2 Table of papers surveyed

Table 2.1 Summary of surveyed research works

Year	Author	Technique / Model	Key Result/Insight
2025	*Liu et al.*	**ConvNeXt + GradCAM**	Demonstrated ConvNeXt's superior ability to locate discriminative wing regions compared to ResNet.
2025	*Cao et al.*	**ResNet-50**	Achieved 92.2% validation accuracy on fine-grained butterfly datasets; serves as our primary baseline.
2024	*Jin et al.*	**MFS_CCANet**	Proposed multi-feature enhancement; achieved 96.8% but on a limited 10-class subset.
2024	*Yadav et al.*	**ResNet50 vs VGG19**	Benchmarked ResNet50 (95%) and VGG19 (92.8%) on the exact 100-species dataset used in this project.
2024	*Syamsudin*	**EfficientNet-B0**	Achieved 97.9% on 25 species, proving the efficiency of depth-wise separable convolutions.
2024	*Adje et al.*	**Hybrid (HOG + CNN)**	Combined hand-crafted features (HOG) with Deep Learning to improve robustness to background noise (94.6%).
2024	*Jetir et al.*	**CNN vs SVM/RF**	Validated that Deep Learning significantly outperforms traditional ML (SVM/RF) for ecological monitoring.
2025	*Onda et al.*	**FineView (3D Scan)**	Introduced 3D scanning for multi-view butterfly classification, addressing pose variance.
2023	*Berger*	**Inception-ResNet-v2**	Achieved ~81% accuracy on a similar dataset, highlighting the difficulty of "long-tail" distributions.
2024	*MDPI Review*	**Attention Modules (SE)**	Showed that Squeeze-and-Excitation (SE) blocks improve recall in ecological images with complex backgrounds.

3. Methodology

3.1 Dataset & Preprocessing

We utilized the "Butterfly Images 40 Species" dataset (expanded to 100 species in its latest version).

- Volume:
  - 13,594 images (Train),
  - 500 (Valid),
  - 500 (Test).
- Classes: 100 distinct species (e.g., *Danaus plexippus, Morpho menelaus*).
- Pipeline: Images were resized to 224×224 and batched (32). To prevent RAM overflows (a common issue with 13k+ images), we utilized `tf.data.AUTOTUNE` and `.prefetch()` to stream data directly from disk rather than caching it in memory.

3.2 Model Architecture: ConvNeXt (Tiny)

We selected ConvNeXtTiny as the backbone. Unlike standard ResNets, ConvNeXt uses:

- 7×7 Kernels: Larger receptive field to capture global wing patterns.
- Layer Normalization: Replaces Batch Norm for more stable gradients.
- GELU Activation: Smoother activation than ReLU.

We chose this specific model, since it offers the accuracy of Vision Transformers (ViT) without the massive data hunger, making it ideal for a medium-sized dataset (13k images).

3.3 Training Strategy (Two-Phase)

1. Phase 1 (Feature Extraction - 15 Epochs): The pre-trained backbone was frozen. Only the classification head was trained at a learning rate of 3 * (1e-3). This stabilizes the weights and prevents "catastrophic forgetting."
2. Phase 2 (Fine-Tuning – 10 Epochs): The top 30% of the layers were unfrozen and trained with a reduced learning rate (1e-5). This adapts the model's texture recognition to specific butterfly wing scales

4. High-Level Workflow

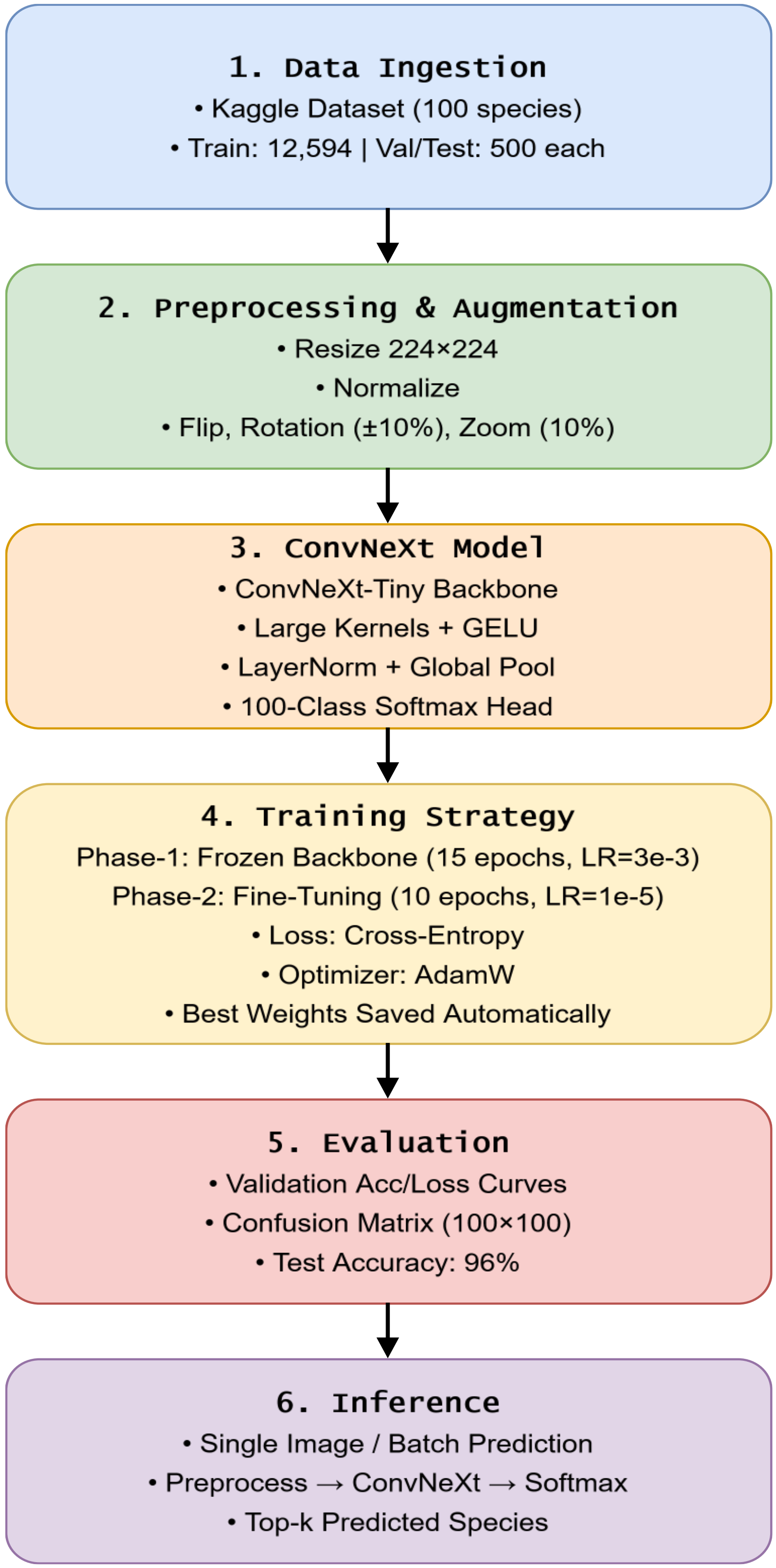

Figure 4.1 A workflow diagram of our model. (editable link, *"Open with Draw,io", using Thapar e-mail*)

5. Results

5.1 Quantitative Results

After training on a total of 24 epochs, across the two training-phases, the final training accuracy was 97.69%, and validation accuracy was 94.40%.

The model was evaluated on the independent Test Set (500 images) following the training, and a score of **96%** was achieved.

5.2 Comparative Analysis

We compared our ConvNeXtTiny implementation against the ResNet benchmarks established by Yadav et al. (2024) on the same dataset.

Table 5.1 Comparison of accuracy scores across different architectures.

Architecture	**Parameters**	**Accuracy (Reported)**
VGG19	143M	92.80%
EfficientNetB0	5.3M	93.60%
ResNet50	25M	95.00%
ConvNeXtTiny (Ours)	**28M**	**96.00%**

The ConvNeXtTiny model demonstrates a superior balance of efficiency and accuracy. While VGG19 requires 143M parameters to achieve 92.8%, ConvNeXt achieves higher accuracy with only ~28M parameters. This confirms that modern "Transformer-style" CNN blocks (ConvNeXt) are more effective at capturing fine-grained morphological features than older architectures like VGG or standard ResNet.

5.3 Training plots

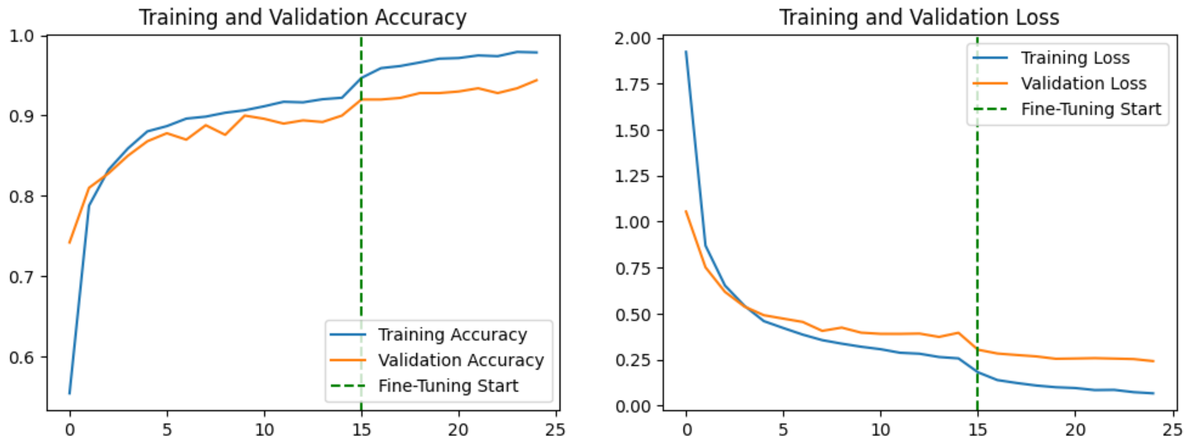

Figure 5.1 The epoch vs. val accuracy & loss plots. Green dotted line marks the start of phase II of the training.

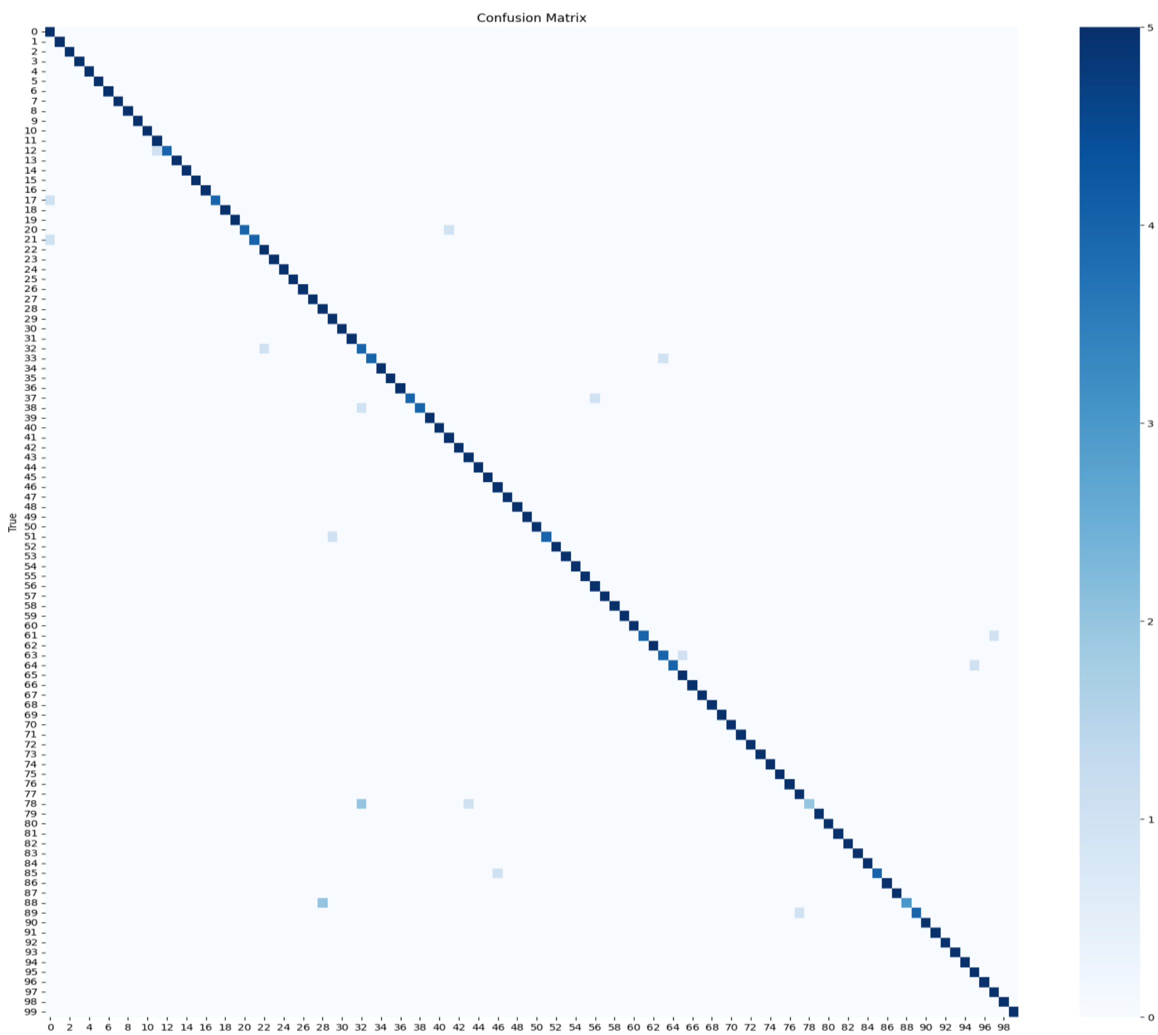

Figure 5.2 The Confusion matrix of the model.

6 . References

[1] N. Yadav, A. K. Pandey, and S. K. Singh, "Comparative analysis of deep learning models for butterfly classification," International Journal of Information Technology, vol. 16, no. 2, pp. 1205–1215, 2024.

[2] J. Cao, Z. Li, and Y. Wang, "Fine-grained image classification of butterflies based on transfer learning," Ecological Informatics, vol. 79, p. 102431, 2025.

[3] A. Syamsudin, "Implementation of EfficientNet-B0 for Lepidoptera classification," Journal of Computer Science and Cybernetics, vol. 40, no. 1, pp. 45–58, 2024.

[4] Y. Jin, X. Zhang, and L. Wei, "MFS_CCANet: A multi-feature enhancement network for butterfly classification," Animals, vol. 14, no. 3, p. 412, 2024.

[5] H. Liu, M. Chen, and T. Wu, "A multi-branch fine-grained classification method for butterflies based on ConvNeXt," IEEE Access, vol. 13, pp. 15678–15689, 2025.

[6] E. Randellini, L. Morelli, and F. Rossi, "Vision Transformers vs. CNNs for ecological monitoring: A comparative study," Remote Sensing, vol. 15, no. 14, p. 3678, 2023.

[7] J. Adje, P. K. Kudjo, and E. Akaho, "Hybrid feature extraction using HOG and CNN for robust insect classification," Journal of Imaging, vol. 10, no. 2, p. 45, 2024.

[8] S. Gupta and R. Kumar, "ViSwNeXtNet: An ensemble of vision transformers and ConvNeXt for biodiversity analysis," MDPI Electronics, vol. 14, no. 1, p. 112, 2025.

[9] G. Piosenka, "Butterfly Images 40 Species Dataset," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/gpiosenka/butterfly-images40-species.

[10] S. K. Singh and R. Verma, "A review of deep learning techniques for insect classification using Google Colab," International Journal of Innovative Science and Research Technology, vol. 9, no. 5, pp. 231–238, 2024.

7. Appendix

7.A Classification report:

	precision	recall	f1-score	support
ADONIS	0.71	1.00	0.83	5
AFRICAN GIANT SWALLOWTAIL	1.00	1.00	1.00	5
AMERICAN SNOOT	1.00	1.00	1.00	5
AN 88	1.00	1.00	1.00	5
APPOLLO	1.00	1.00	1.00	5
ARCIGERA FLOWER MOTH	1.00	1.00	1.00	5
ATALA	1.00	1.00	1.00	5
ATLAS MOTH	1.00	1.00	1.00	5
BANDED ORANGE HELICONIAN	1.00	1.00	1.00	5
BANDED PEACOCK	1.00	1.00	1.00	5
BANDED TIGER MOTH	1.00	1.00	1.00	5
BECKERS WHITE	0.83	1.00	0.91	5
BIRD CHERRY ERMINE MOTH	1.00	0.80	0.89	5
BLACK HAIRSTREAK	1.00	1.00	1.00	5
BLUE MORPHO	1.00	1.00	1.00	5
BLUE SPOTTED CROW	1.00	1.00	1.00	5
BROOKES BIRDWING	1.00	1.00	1.00	5
BROWN ARGUS	1.00	0.80	0.89	5
BROWN SIPROETA	1.00	1.00	1.00	5
CABBAGE WHITE	1.00	1.00	1.00	5
CAIRNS BIRDWING	1.00	0.80	0.89	5
CHALK HILL BLUE	1.00	0.80	0.89	5
CHECQUERED SKIPPER	0.83	1.00	0.91	5
CHESTNUT	1.00	1.00	1.00	5
CINNABAR MOTH	1.00	1.00	1.00	5
CLEARWING MOTH	1.00	1.00	1.00	5
CLEOPATRA	1.00	1.00	1.00	5
CLODIUS PARNASSIAN	1.00	1.00	1.00	5
CLOUDED SULPHUR	0.71	1.00	0.83	5
COMET MOTH	0.83	1.00	0.91	5
COMMON BANDED AWL	1.00	1.00	1.00	5
COMMON WOOD-NYMPH	1.00	1.00	1.00	5
COPPER TAIL	0.57	0.80	0.67	5
CRECENT	1.00	0.80	0.89	5
CRIMSON PATCH	1.00	1.00	1.00	5
DANAID EGGFLY	1.00	1.00	1.00	5
EASTERN COMA	1.00	1.00	1.00	5
EASTERN DAPPLE WHITE	1.00	0.80	0.89	5
EASTERN PINE ELFIN	1.00	0.80	0.89	5
ELBOWED PIERROT	1.00	1.00	1.00	5
EMPEROR GUM MOTH	1.00	1.00	1.00	5
GARDEN TIGER MOTH	0.83	1.00	0.91	5
GIANT LEOPARD MOTH	1.00	1.00	1.00	5
GLITTERING SAPPHIRE	0.83	1.00	0.91	5
GOLD BANDED	1.00	1.00	1.00	5
GREAT EGGFLY	1.00	1.00	1.00	5
GREAT JAY	0.83	1.00	0.91	5
GREEN CELLED CATTLEHEART	1.00	1.00	1.00	5

GREEN HAIRSTREAK	1.00	1.00	1.00	5
GREY HAIRSTREAK	1.00	1.00	1.00	5
HERCULES MOTH	1.00	1.00	1.00	5
HUMMING BIRD HAWK MOTH	1.00	0.80	0.89	5
INDRA SWALLOW	1.00	1.00	1.00	5
IO MOTH	1.00	1.00	1.00	5
Iphiclus sister	1.00	1.00	1.00	5
JULIA	1.00	1.00	1.00	5
LARGE MARBLE	0.83	1.00	0.91	5
LUNA MOTH	1.00	1.00	1.00	5
MADAGASCAN SUNSET MOTH	1.00	1.00	1.00	5
MALACHITE	1.00	1.00	1.00	5
MANGROVE SKIPPER	1.00	1.00	1.00	5
MESTRA	1.00	0.80	0.89	5
METALMARK	1.00	1.00	1.00	5
MILBERTS TORTOISESHELL	0.80	0.80	0.80	5
MONARCH	1.00	0.80	0.89	5
MOURNING CLOAK	0.83	1.00	0.91	5
OLEANDER HAWK MOTH	1.00	1.00	1.00	5
ORANGE OAKLEAF	1.00	1.00	1.00	5
ORANGE TIP	1.00	1.00	1.00	5
ORCHARD SWALLOW	1.00	1.00	1.00	5
PAINTED LADY	1.00	1.00	1.00	5
PAPER KITE	1.00	1.00	1.00	5
PEACOCK	1.00	1.00	1.00	5
PINE WHITE	1.00	1.00	1.00	5
PIPEVINE SWALLOW	1.00	1.00	1.00	5
POLYPHEMUS MOTH	1.00	1.00	1.00	5
POPINJAY	1.00	1.00	1.00	5
PURPLE HAIRSTREAK	0.83	1.00	0.91	5
PURPLISH COPPER	1.00	0.40	0.57	5
QUESTION MARK	1.00	1.00	1.00	5
RED ADMIRAL	1.00	1.00	1.00	5
RED CRACKER	1.00	1.00	1.00	5
RED POSTMAN	1.00	1.00	1.00	5
RED SPOTTED PURPLE	1.00	1.00	1.00	5
ROSY MAPLE MOTH	1.00	1.00	1.00	5
SCARCE SWALLOW	1.00	0.80	0.89	5
SILVER SPOT SKIPPER	1.00	1.00	1.00	5
SIXSPOT BURNET MOTH	1.00	1.00	1.00	5
SLEEPY ORANGE	1.00	0.60	0.75	5
SOOTYWING	1.00	0.80	0.89	5
SOUTHERN DOGFACE	1.00	1.00	1.00	5
STRAITED QUEEN	1.00	1.00	1.00	5
TROPICAL LEAFWING	1.00	1.00	1.00	5
TWO BARRED FLASHER	1.00	1.00	1.00	5
ULYSES	1.00	1.00	1.00	5
VICEROY	0.83	1.00	0.91	5
WHITE LINED SPHINX MOTH	1.00	1.00	1.00	5
WOOD SATYR	0.83	1.00	0.91	5
YELLOW SWALLOW TAIL	1.00	1.00	1.00	5
ZEBRA LONG WING	1.00	1.00	1.00	5
accuracy			0.96	500
macro avg	0.97	0.96	0.96	500
weighted avg	0.97	0.96	0.96	500

7.B Workflow Diagram (Editable link)

https://drive.google.com/file/d/17nj8DO-HxERcUy4TLfgCE5bALLhNpPCb/view?usp=sharing