Enchanted Wings: Marvels of Butterfly Species

1. Introduction

1.1 Project Overview

Enchanted Wings is a deep learning–based project designed to classify butterfly species using computer vision and image augmentation. The goal is to assist researchers, students, and nature enthusiasts by automating butterfly image identification — enhancing accuracy and efficiency in biodiversity studies.

The system uses a custom dataset of butterfly images covering different species, such as Monarch, Swallowtail, and Morpho butterflies.

1.2 Objectives

- Automate butterfly species classification using CNNs.
- Improve generalization and accuracy with image augmentation.
- Provide an interactive, user-friendly web interface for uploading and predicting images.
- Demonstrate AI's role in supporting biodiversity research and conservation.

---

2. Project Initialization and Planning Phase

2.1 Define Problem Statement

Manual classification of butterfly species is time-consuming, requires expertise, and is prone to human error. The objective is to develop an accurate and automated system that classifies different butterfly species from images using a trained deep learning model.

2.2 Project Proposal (Proposed Solution)

The system uses a Convolutional Neural Network (CNN) trained on a labeled butterfly dataset. Data augmentation techniques (rotation, zoom, flip) improve model generalization. The trained model is integrated with a web application using Streamlit to allow easy image uploads and real-time predictions.

2.3 Initial Project Planning

- Defined workflow: Data collection, preprocessing, modeling, evaluation, deployment.
- Selected CNN architectures suitable for small-to-moderate datasets.
- Planned user interface development using Streamlit for deployment.

---

3. Data Collection and Preprocessing Phase

3.1 Data Collection Plan and Sources

- Dataset: Collected butterfly species images from sources like Wikimedia Commons, open datasets, and manually curated samples.
- Categories: Monarch, Swallowtail, Morpho.
- Total Samples: ~30–50 images per class (expandable).

3.2 Data Quality Report

- Data Shape: Balanced number of images across species.
- Missing/Corrupted Images: Removed.
- Image Size: Standardized to 128×128 pixels for CNN input.

3.3 Data Preprocessing

- Resized all images to 128×128 pixels.
- One-hot encoded class labels.
- Applied data augmentation (rotation, zoom, horizontal flip).
- Used an 80%–20% train-validation split for training.

---

4. Model Development Phase

4.1 Feature Selection Report

- Features: Pixel intensity values and spatial patterns learned through CNN layers.
- Target: Butterfly species.

4.2 Model Selection Report

Models Used:

- Custom CNN (built with Keras/TensorFlow)
- Activation Functions: ReLU for hidden layers, Softmax for output.

Metrics:

- Accuracy
- Confusion Matrix
- Loss plots

4.3 Initial Training & Evaluation

- Base model: Simple CNN architecture with two convolutional and pooling layers, fully connected dense layers, and dropout for regularization.
- Achieved ~80%–90% accuracy on validation data with the sample dataset.
- Performance expected to improve with more images and fine-tuning.

5. Model Optimization and Tuning Phase

5.1 Hyperparameter Tuning

- Learning Rate: Tuned with callbacks (ReduceLROnPlateau).
- Optimizer: Adam.
- Batch Size: Tested with 16 and 32.
- Epochs: 10–20 with early stopping.

5.2 Performance Comparison

Model	Accuracy
Custom CNN + Augmentation	85%

5.3 Final Model Selection

- Selected Model: Custom CNN with augmentation due to:
- Simplicity and adaptability for small datasets.
- Good balance between accuracy and training time.

- Suitable for deployment on web UI with limited resources.

6. Results

6.1 Output Screenshots

- Web interface built with Streamlit for image upload.
- Real-time species prediction displayed with confidence score.
- Accuracy and loss graphs plotted for training and validation data.

7. Advantages & Limitations

Advantages:

- Accurate butterfly classification with minimal human effort.
- User-friendly, interactive web interface.
- Supports biodiversity research and educational applications.

Limitations:

- Performance depends on the size and quality of the dataset.
- May need retraining for more species or different regional butterflies.

8. Conclusion

Enchanted Wings demonstrates how deep learning and computer vision can automate species identification, supporting biodiversity research and conservation. With a user-friendly deployment, it shows the practical impact of AI in ecological applications.

9. Future Scope

- Expand the dataset with more butterfly species and real-world variations.
- Optimize the model for mobile devices for field research use.
- Integrate with citizen science platforms for large-scale data collection.
- Explore advanced architectures like transfer learning for improved accuracy.

10. Appendix

10.1 Source Code

- app.py (Streamlit web app)
- train.py (Model training script)
- dataset/ (Butterfly images organized by species)
- model/ (Saved model and label map)

10.2 GitHub & Project Demo

https://github.com/Varun-panchakarla/enhancedWings_AIML