Project Title: Enchanted Wings: Marvels of Butterfly Species - A CNN-Based Image Classification Model

---

Abstract: This project focuses on classifying butterfly species using a Convolutional Neural Network (CNN) trained on a Kaggle dataset. With the increasing role of machine learning in biodiversity, this project aims to automate the recognition of butterfly species based on image data. The deep learning model enhances accuracy by using data augmentation and normalization techniques. The solution has real-world applications in ecological research, conservation, and education.

Objective: To develop a CNN model that classifies images of butterflies into their respective species using supervised learning and image processing techniques.

Algorithm Explanation: The CNN algorithm used involves several convolutional layers to extract spatial features, pooling layers to reduce dimensionality, batch normalization for faster convergence, and dense layers to perform final classification. The model uses the softmax function for multi-class classification and is trained using the Adam optimizer and categorical cross-entropy loss.

Project Explanation:

1. The butterfly image dataset is downloaded from Kaggle using `kagglehub`.
2. Images are loaded and labeled using a CSV file.
3. The dataset is split into training and validation sets.
4. A CNN is built using TensorFlow/Keras with multiple layers.
5. The model is trained over 30 epochs.
6. Performance is evaluated using accuracy and loss curves.
7. Predictions are visualized to compare actual and predicted labels.

Project Flow:

1. Dataset Collection
2. Data Preprocessing
3. Image Augmentation
4. CNN Model Building
5. Training the Model
6. Model Evaluation
7. Predictions

Prior Knowledge Required:

- Basics of Python programming
- Understanding of neural networks and CNNs
- Familiarity with TensorFlow/Keras libraries
- Knowledge of image data processing
- Basic data handling with pandas and NumPy

Architecture:

```
Input Image (128x128x3)
       ↓
[Conv2D + ReLU] → [BatchNormalization] → [MaxPooling]
       ↓
[Conv2D + ReLU] → [BatchNormalization] → [MaxPooling]
       ↓
[Conv2D + ReLU] → [BatchNormalization] → [MaxPooling]
       ↓
[Flatten] → [Dense 128] → [Dense 512] → [Dense n_classes + Softmax]
```

Project Structure:

```
├── butterfly_classification_project
│   ├── dataset
│   │   ├── train
│   │   ├── test
│   │   └── Training_set.csv
│   ├── main.py (or .ipynb)
│   ├── model.h5
│   ├── requirements.txt
│   └── README.md
```

Dataset Collection:

- Source: Kaggle Dataset - `phucthaiv02/butterfly-image-classification`
- Format: Images (.jpg/.png) with a CSV file containing filename-label pairs.
- Categories: Multiple butterfly species

Data Visualization:

- Sample image grid from the dataset
- Bar chart showing the distribution of classes
- Accuracy vs Epochs line plot
- Loss vs Epochs line plot

Example Reference Images Used:

Model Building:

- Uses Keras Sequential API
- Three Conv2D + MaxPooling2D layers
- Dense layers for classification
- Optimizer: Adam
- Loss: categorical_crossentropy
- Metrics: Accuracy

Testing Model & Data Prediction:

- Predictions on validation set
- `np.argmax` to convert probabilities to class labels
- Display true vs predicted labels using `matplotlib`

Applications:

- Biodiversity conservation
- Educational tools for species identification
- Automated classification in ecological surveys
- Assisting biologists in insect taxonomy

Conclusion: This project successfully demonstrates the ability of CNNs to classify butterfly species based on image input. With good accuracy and visual validation, it sets a strong foundation for further enhancements using transfer learning or larger datasets.

End of Report