Project Design Phase
Solution Architecture

Date	15 February 2025
Team ID	LTVIP2025TMID42449
Project Name	Butterfly Species Classification System
Maximum Marks	4 Marks

1 Solution Architecture

1.1 Purpose

Solution architecture bridges the gap between business problems and technology solutions by defining the structure, characteristics, and behavior of the software. Its goals are to:

- Identify the optimal tech solution for classifying butterfly species.
- Describe the system's components and data flow to stakeholders.
- Define features, development phases, and requirements.
- Provide specifications for system management and delivery.

1.2 Architecture Description

The Butterfly Species Classification System is a web-based application that uses transfer learning with a pre-trained convolutional neural network (CNN), specifically MobileNetV2, to classify butterfly images across 75 species. The architecture is designed for efficiency, accuracy, and user engagement, leveraging the following components:

- **Frontend**: Built with Streamlit, offering a user-friendly interface for uploading butterfly images, displaying species predictions, and providing educational facts about butterflies.
- **Backend**: Utilizes Python with TensorFlow/Keras for model inference, Flask for API management, and integration with the pre-trained CNN for classification tasks.
- **CNN Model**: MobileNetV2, pre-trained on ImageNet, is fine-tuned for butterfly species classification, with additional dense layers for 75-class output, optimized for lightweight deployment.

- **Database**: SQLite stores metadata about butterfly images and species information (e.g., scientific names, ecological facts) for quick retrieval in the application.
- **Infrastructure**: Deployed locally or on cloud platforms like Google Colab for model training and Streamlit Cloud for app hosting, ensuring scalability and accessibility.

1.3 Rationale for Transfer Learning

Transfer learning with MobileNetV2 is chosen for:

- **Efficiency**: Pre-trained weights reduce training time and computational resources, ideal for a dataset of 6499 images.
- **Accuracy**: Leverages ImageNet's feature extraction capabilities, fine-tuned for butterfly-specific patterns, achieving high classification accuracy (targeting 95%).
- **Lightweight Design**: MobileNetV2's compact architecture suits deployment in resource-constrained environments, such as web applications.

1.4 Data Flow

1. **User Input**: Users upload butterfly images via the Streamlit frontend.
2. **Preprocessing**: Backend resizes images to 224x224 pixels, normalizes pixel values, and applies augmentation (e.g., rotation, flipping) for robustness.
3. **Model Inference**: The fine-tuned MobileNetV2 processes images to predict species probabilities across 75 classes.
4. **Data Storage**: Image metadata and species details are stored in SQLite for retrieval of educational facts.
5. **Output Delivery**: The frontend displays the predicted species, confidence score, and educational facts to the user.

1.5 Features and Development Phases

- **Features**: Image-based species classification, real-time predictions, educational content delivery, and data preprocessing for robust model performance.
- **Phases**:
  - **Phase 1**: Data collection, preprocessing, and model training.
  - **Phase 2**: Model evaluation and testing.
  - **Phase 3**: Streamlit application development and deployment.

1.6 Solution Architecture Diagram

[Placeholder for architecture diagram: Illustrates data flow from user image upload through Streamlit frontend, Flask backend, MobileNetV2 inference, SQLite storage, and output display.]

Reference:
https://aws.amazon.com/blogs/industries/voice-applications-in-clinical-researchpowered-by-ai-on-aws-part-1-architecture-and-design-considerations/