**Đề tài: Bees & Ants Classification - PyTorch**

**Mô tả nghiên cứu trong lĩnh vực phân tích dữ liệu**

**1. Giới thiệu về đề tài**

Đề tài *Bees & Ants Classification* sử dụng **PyTorch**, một thư viện học sâu mạnh mẽ, để xây dựng mô hình phân loại hình ảnh giữa hai loại côn trùng: **ong (bees) và kiến (ants)**. Nghiên cứu này thuộc lĩnh vực **phân tích dữ liệu hình ảnh (image analysis)** và **thị giác máy tính (computer vision)**, trong đó dữ liệu được xử lý, huấn luyện trên mô hình học sâu để tự động nhận diện và phân loại hình ảnh. Việc phân nhãn chính xác giữa các loài này có giá trị lớn trong nhiều lĩnh vực, bao gồm nghiên cứu về sinh thái học, kiểm định dịch bệnh, và bảo vệ môi trường.

**2. Mục tiêu nghiên cứu**

* Ứng dụng **học sâu (deep learning)** trong **phân loại hình ảnh (image classification)**.
* Khai thác sức mạnh của **mạng nơ-ron tích chập (CNN - Convolutional Neural Networks)** để phân biệt hai loại côn trùng.
* Sử dụng **PyTorch** để triển khai mô hình học sâu, tối ưu hiệu suất phân loại.
* Tìm hiểu cách tiền xử lý dữ liệu ảnh và tăng cường dữ liệu (data augmentation) để cải thiện độ chính xác.

**3. Dữ liệu sử dụng**

* **Nguồn dữ liệu**: Tập dữ liệu hình ảnh của **ong** và **kiến**, có thể được thu thập từ bộ dữ liệu ImageNet hoặc một tập dữ liệu tùy chỉnh.
* **Đặc điểm dữ liệu**:
  + Gồm hai lớp: *bees* (ong) và *ants* (kiến).
  + Mỗi hình ảnh có kích thước khác nhau, cần được tiền xử lý về cùng kích thước chuẩn.
  + Có thể chứa nhiễu hoặc các đặc điểm khó phân biệt, cần áp dụng **kỹ thuật tăng cường dữ liệu (data augmentation)**.

**4. Phương pháp nghiên cứu**

* **Tiền xử lý dữ liệu**:
  + Chuyển đổi hình ảnh về kích thước cố định.
  + Chuẩn hóa (normalization) để đưa các pixel về một phạm vi giá trị thích hợp.
  + Áp dụng kỹ thuật tăng cường dữ liệu như lật ảnh, xoay, thay đổi độ sáng để cải thiện khả năng tổng quát của mô hình.
* **Xây dựng mô hình học sâu**:
  + Sử dụng mô hình **CNN (Convolutional Neural Network)** hoặc fine-tune trên mô hình **ResNet, VGG hoặc AlexNet** để trích xuất đặc trưng từ ảnh.
  + Huấn luyện mô hình bằng **PyTorch**, tối ưu bằng **hàm mất mát Cross-Entropy và thuật toán tối ưu Adam hoặc SGD**.
* **Đánh giá hiệu suất**:
  + Sử dụng các chỉ số như **độ chính xác (accuracy), độ nhạy (recall), độ đặc hiệu (precision), và F1-score**.
  + Trực quan hóa kết quả bằng ma trận nhầm lẫn (confusion matrix).

**5. Ứng dụng của nghiên cứu**

* **Sinh học và nghiên cứu môi trường**: Hỗ trợ nhận diện côn trùng tự động trong các nghiên cứu sinh thái.
* **Nông nghiệp thông minh**: Phân loại ong và kiến để theo dõi số lượng loài có lợi hoặc gây hại cho mùa màng.
* **Phát triển hệ thống nhận diện hình ảnh**: Làm nền tảng cho các nghiên cứu sâu hơn về phân loại hình ảnh trong lĩnh vực AI.

**6. Kết luận**

Nghiên cứu này giúp áp dụng **học sâu** vào **phân loại hình ảnh**, sử dụng **PyTorch** để huấn luyện mô hình CNN. Việc phân loại ong và kiến không chỉ giúp cải thiện nhận diện sinh học mà còn mở ra nhiều ứng dụng trong thực tế.