

Thời gian làm bài: 120 phút

Dataset: [Plant Disease Dataset from Kaggle](#)

Chủ đề: Phân loại bệnh trên cây

Mục tiêu

Xây dựng một hệ thống nhận dạng hình ảnh bệnh trên lá cây bằng cách:

- Trích xuất đặc trưng hình ảnh dùng ELM (Extreme Learning Machine)
 - Phân loại bằng SVM (Support Vector Machine).
-

Câu 1 (2 điểm): Tiền xử lý dữ liệu

1. Lọc ra các ảnh **liên quan đến mọi loại cây** từ bộ dữ liệu.
 2. Thống kê số lượng ảnh trên từng lớp bệnh của cây
 3. Chia dữ liệu thành 3 tập: huấn luyện (70%), kiểm tra (15%), và xác thực (15%).
 4. Resize ảnh về kích thước phù hợp (ví dụ 224x224 hoặc 128x128), chuẩn hóa giá trị pixel về $[0, 1]$.
-

Câu 2 (3 điểm): Trích xuất đặc trưng

- Trình bày ý tưởng tổng quan về ELM.
 - Sử dụng ELM để chọn đặc trưng từ ảnh.
-

Câu 3 (3 điểm): Phân loại bệnh

1. SVM để huấn luyện mô hình phân loại dựa trên đặc trưng đã trích xuất.
2. Đánh giá độ chính xác mô hình trên tập kiểm tra.
3. Vẽ ma trận nhầm lẫn, nêu rõ các lớp dễ nhầm lẫn.
4. Nhận xét về độ chính xác và đề xuất cải tiến.

Câu 4 (2 điểm): So sánh mô hình

1. Chọn phương pháp trích xuất đặc trưng khác thay cho ELM
2. Đánh giá ưu nhược điểm của hai hướng tiếp cận (độ chính xác, tốc độ, khả năng triển khai).

Yêu cầu nộp bài:

- **Notebook** (.ipynb) hoặc **Python script** (.py) chứa mã nguồn và kết quả thực nghiệm.
- **Báo cáo PDF (1-2 trang)** mô tả quá trình xử lý, mô hình sử dụng, và đánh giá kết quả.
- Gửi kèm link Google Drive hoặc Colab.