

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ



BÁO CÁO BÀI TẬP LỚN HỌC PHẦN HỌC MÁY

Mã môn học: 2425II_INT3405_52

Phân vùng cải bó xôi

Phân loại bệnh trên lá lúa

Sinh viên: Phí Vân Long - 21020563 – QH-2021-I/CQ-R

Giảng viên hướng dẫn: PGS.TS. Phạm Minh Triển

Hà Nội – 2024

1. Phân vùng cải bó xôi

1.1 Xây dựng, phân tích và mô tả bộ dữ liệu

- Tên bộ dữ liệu: Subset 1, chia làm 3 folder X2, X11, Xb
- Bộ dữ liệu gồm 180 ảnh màu (RGB) của các lá cây, được chụp dưới các góc khác nhau và những cây khác nhau, trong các giai đoạn phát triển khác nhau
- Định dạng ảnh: .jpg
- Độ phân giải ảnh: 1920x2560
- Ảnh được label sử dụng Cvat



1.2 Tiền xử lý dữ liệu

- Tên các ảnh được chuẩn hóa theo tên file (trong folder X2 các ảnh có tên theo mẫu: 1.X2.1.1.22.09012025) (gồm các thông tin như tên folder, ngày chụp,...)
- Ảnh được up lên Roboflow cho dễ sử dụng
- Tiền xử lý gồm:
 - Tự động xoay ảnh
 - Thay đổi kích cỡ (640x640)
 - Tự động chỉnh độ tương phản: tăng khả năng nhận biết với nền

Train/Test Split

Training Set: 123 images
Validation Set: 33 images
Testing Set: 17 images

3

Preprocessing

? What can preprocessing do?

Decrease training time and increase performance by applying image transformations to all images in this dataset.

Auto-Orient	Edit	×
Resize Stretch to 640×640	Edit	×
Auto-Adjust Contrast Using Adaptive Equalization	Edit	×
+ Add Preprocessing Step		

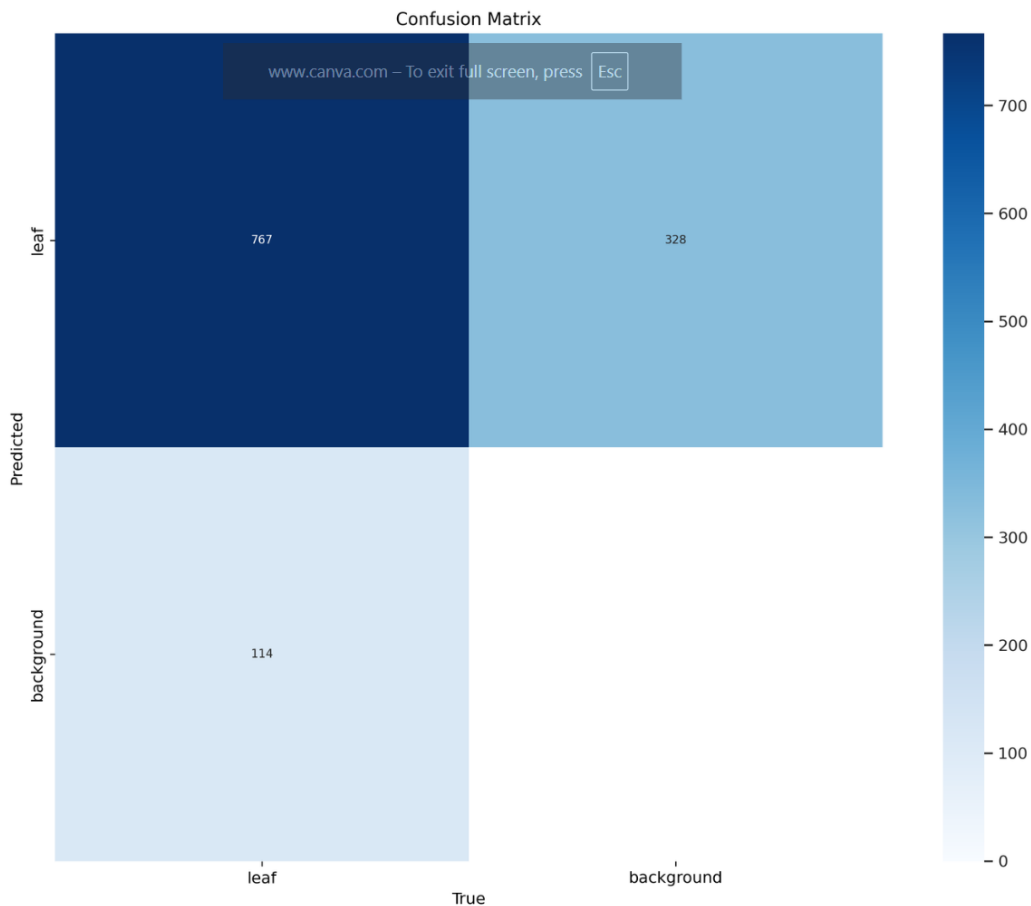
1.3 Trích xuất đặc trưng

- Trích xuất đặc trưng chuyển đổi ảnh thô thành biểu diễn cấp cao, giúp mô hình YOLOv8 nhận diện và phân đoạn lá. Quá trình này được thực hiện tự động bởi backbone của YOLOv8.
- Kiến trúc YOLOv8
 - Lớp tích chập: Trích xuất cạnh, góc, kết cấu lá với kernel 3x3.
 - Cơ chế CSP: Giảm chi phí tính toán, kết hợp thông tin từ nhiều lớp.
 - Pooling Layers: Tập trung vào đặc trưng nổi bật, giảm kích thước không gian.
 - Batch Normalization và SiLU: Chuẩn hóa và thêm tính phi tuyến.
 - Feature Pyramid Network (FPN): Kết hợp đặc trưng đa tỷ lệ để phân đoạn lá khác kích thước
- Quá trình trích xuất đặc trưng
 - Đầu vào: ảnh huấn luyện 70%, ảnh xác thực 20%, ảnh test 10% (640x640).
 - Trích xuất cạnh, màu sắc lá.
 - Tích hợp đa tỷ lệ: FPN kết hợp feature maps từ các độ phân giải.
 - Đầu ra: Feature maps cho phân loại và phân đoạn.

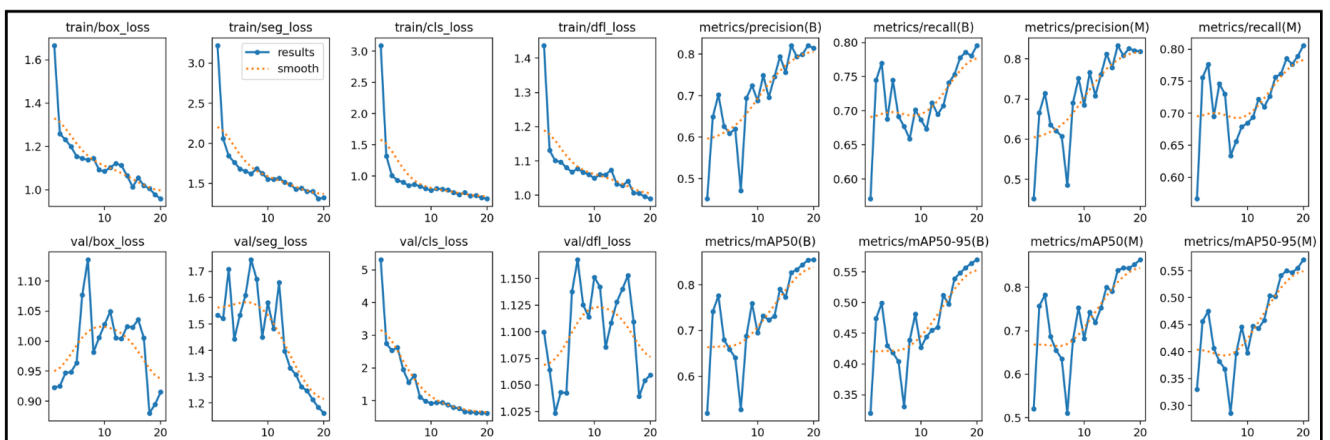
1.4 Xây dựng và đánh giá mô hình

- Xây dựng mô hình
 - Cấu hình môi trường: Cài ultralytics, sử dụng GPU của Google collab
 - Chuẩn bị dữ liệu: Dữ liệu YOLO từ COCO
 - Tối ưu hóa: Áp dụng augmentation mặc định
 - Huấn luyện trong 20 epochs, kích cỡ ảnh là 640x640
- Đánh giá mô hình qua Ma trận nhầm lẫn
 - Ma trận nhầm lẫn cho thấy hiệu suất phân loại giữa hai lớp: "leaf" (lá) và "background" (nền).
 - True Positive (TP) - "leaf" dự đoán đúng là "leaf": 767.

- False Positive (FP) - "background" dự đoán nhầm là "leaf": 328.
- True Negative (TN) - "background" dự đoán đúng là "background": 114.
- False Negative (FN) - "leaf" dự đoán nhầm là "background": còn lại
- Mô hình dự đoán lớp "leaf" khá chính xác (767 TP), nhưng có một số lỗi, cho thấy nhầm lẫn giữa nền và lá.



- Đánh giá mô hình qua các biểu đồ

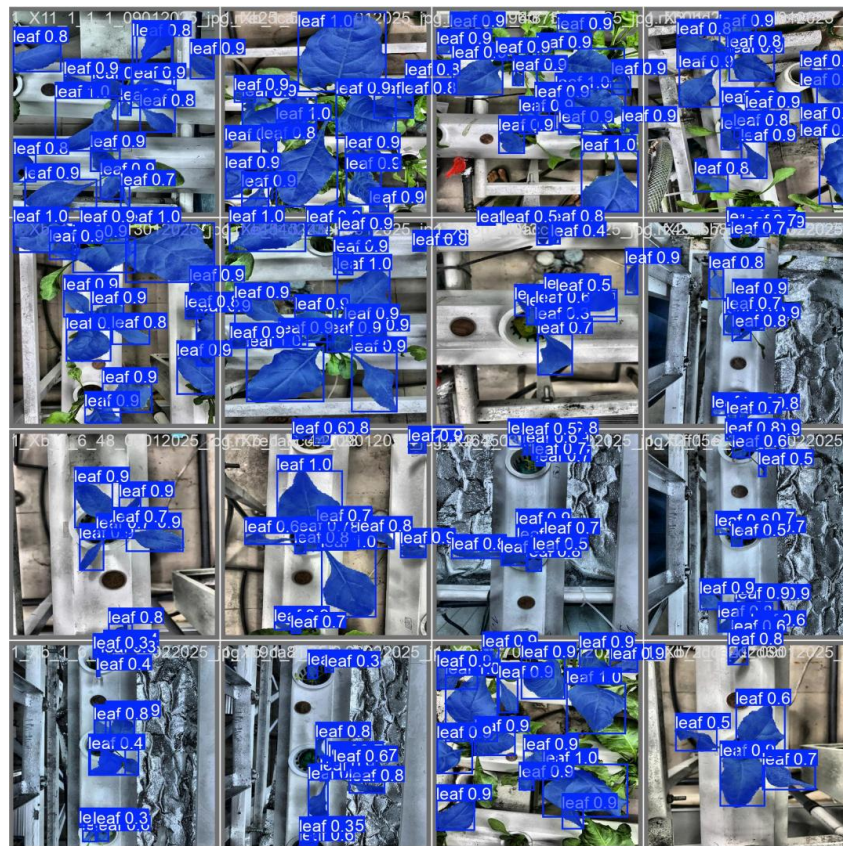


- **Loss (Hàm mất mát)**

- train/box_loss, train/seg_loss, train/cls_loss, train/df_l_loss: Giảm dần qua các epoch, cho thấy mô hình hội tụ

- val/box_loss, val/seg_loss, val/cls_loss, val/df1_loss: Dao động nhưng ổn định, với giá trị thấp, cho thấy mô hình tổng quát hóa tốt.
- **Độ chính xác và độ nhạy (Precision và Recall)**
 - metrics/precision(B) và metrics/recall(B) dao động quanh 0.7-0.8, cho thấy hiệu suất chấp nhận được trên tập huấn luyện.
 - metrics/precision(M) và metrics/recall(M) ổn định hơn, quanh 0.75-0.80 trên tập xác thực.
- **mAP (Mean Average Precision)**
 - metrics/mAP50(B) đạt khoảng 0.7-0.75, metrics/mAP50-95(B) quanh 0.55-0.60, phản ánh khả năng phát hiện và phân đoạn chính xác ở ngưỡng IoU 0.5 và trung bình IoU 0.5-0.95.
 - Trên tập xác thực, metrics/mAP50-95(M) dao động quanh 0.50-0.55, cho thấy hiệu suất ổn định nhưng có thể cải thiện.

1.5 Diễn giải mô hình và trình bày kết quả



- **Diễn giải mô hình**
 - Mô hình dự đoán đúng phần lớn các vùng lá (767 TP), nhưng có 328 trường hợp nhầm nền thành lá (FP), có thể do ánh sáng không đồng đều hoặc các vật thể giống lá trong nền.
 - Số lượng TN thấp (114) cho thấy mô hình khó phân biệt nền ở một số ảnh, đặc biệt khi lá chiếm phần lớn khung hình hoặc có viền không rõ ràng.
 - Loss: Giảm dần qua các epoch cho thấy mô hình học tốt các đặc trưng phân đoạn.
 - mAP và Precision/Recall: Giá trị mAP50-95 quanh 0.55-0.60 cho thấy mô hình có khả năng phân đoạn tốt ở ngưỡng IoU 0.5, nhưng hiệu suất giảm ở

ngưỡng cao hơn (0.95), phản ánh hạn chế trong việc dự đoán viên lá chính xác.

- Xu hướng hội tụ: Các biểu đồ cho thấy mô hình hội tụ sau khoảng 20-30 epoch, nhưng vẫn có dao động nhỏ, có thể do kích thước dữ liệu hạn chế.
- Trình bày kết quả
 - Mô hình YOLOv8 được xây dựng và huấn luyện thành công trên bộ dữ liệu 180 ảnh lá, đạt hiệu suất chấp nhận được với mAP50-95 khoảng 0.55-0.60.
 - Ma trận nhầm lẫn và biểu đồ huấn luyện cho thấy mô hình có khả năng phân đoạn lá tốt, nhưng cần cải thiện để giảm nhầm lẫn với nền và tăng độ chính xác ở ngưỡng IoU cao.

2. Phân loại bệnh trên lá lúa

2.1 Xây dựng, phân tích và mô tả bộ dữ liệu

- Bộ dữ liệu được sử dụng trong nghiên cứu này là tập hợp các hình ảnh lá lúa bị ảnh hưởng bởi bốn loại bệnh: Bacterialblight, Blast, Brownspot, và Tungro
- Nguồn dữ liệu: Kaggle – Rice Leaf Disease Images, gồm 5932 ảnh (RGB)
 - Bacterialblight: 1584 ảnh
 - Blast: 1440 ảnh
 - Brownspot: 1600 ảnh
 - Tungro: 1308 ảnh
- Dữ liệu được tổ chức trong thư mục ./dataset với hai phân vùng chính: train_val (dùng để huấn luyện và xác thực) và test (dùng để kiểm tra).
 - Thư mục train_val chứa các hình ảnh huấn luyện và xác thực
 - Thư mục test chứa các hình ảnh độc lập để đánh giá hiệu suất mô hình.
- Dữ liệu được tăng cường (data augmentation) bằng cách xoay, dịch chuyển, và lật ngang để tăng tính tổng quát hóa.

2.2 Tiền xử lý dữ liệu

- Chuẩn hóa giá trị pixel về khoảng $[0,1]$ bằng cách chia cho 255.
- Tăng cường dữ liệu (data augmentation) với các phép biến đổi như xoay (rotation_range=20), dịch chuyển (width_shift, height_shift=0.2), và lật ngang (horizontal_flip=True).
- Kích thước ảnh được cố định ở 224x224 pixel để phù hợp với đầu vào của các mô hình.

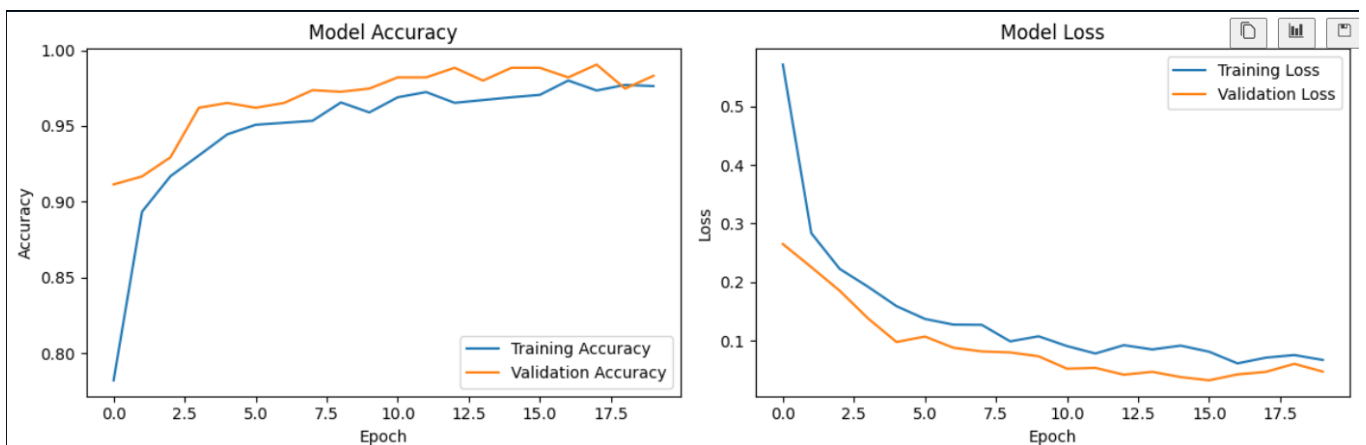
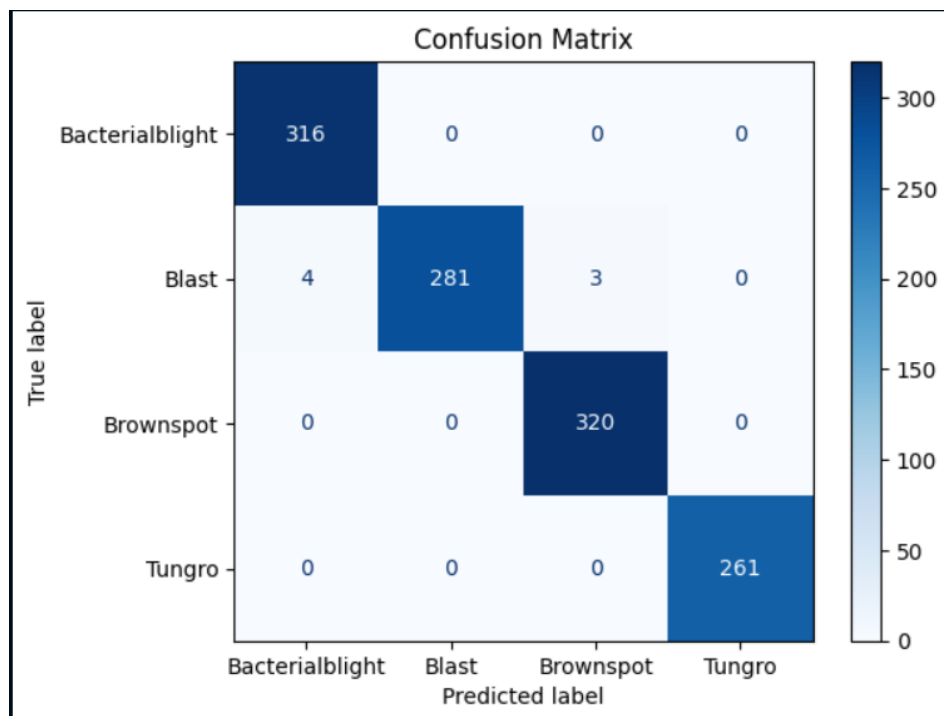
2.3 Trích xuất đặc trưng

- Các mô hình học sâu được sử dụng: MobileNetV2, EfficientNetB0, và ResNet50
- MobileNetV2: Sử dụng các khối Depthwise Separable Convolution để giảm độ phức tạp tính toán, phù hợp cho các thiết bị có tài nguyên hạn chế.
- EfficientNetB0: Áp dụng kỹ thuật scaling đồng bộ (width, depth, resolution) để tối ưu hiệu suất, nổi bật với hiệu quả trên tập dữ liệu nhỏ.
- ResNet50: Sử dụng các khối Residual Connection để giải quyết vấn đề suy thoái gradient, cho phép xây dựng mạng sâu hơn.
- Các đặc trưng được trích xuất từ các tầng trung gian (intermediate layers) của mô hình, sau đó được đưa qua các lớp fully connected để phân loại.

- Lớp GlobalAveragePooling2D được sử dụng để giảm chiều không gian, kết hợp
- Các lớp Dense và Dropout (tỷ lệ 0.5) để tránh overfitting.

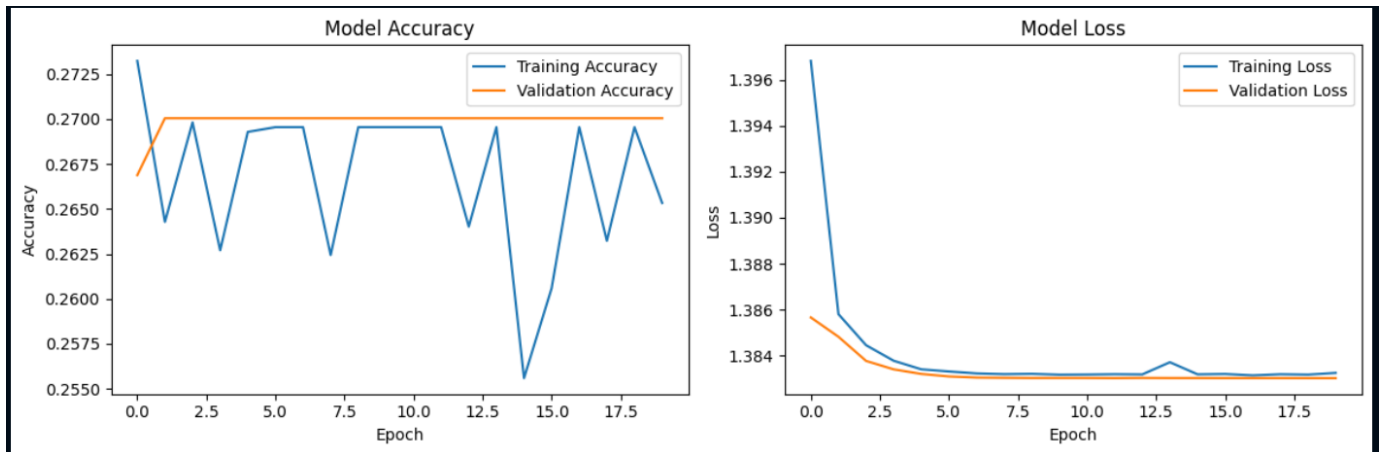
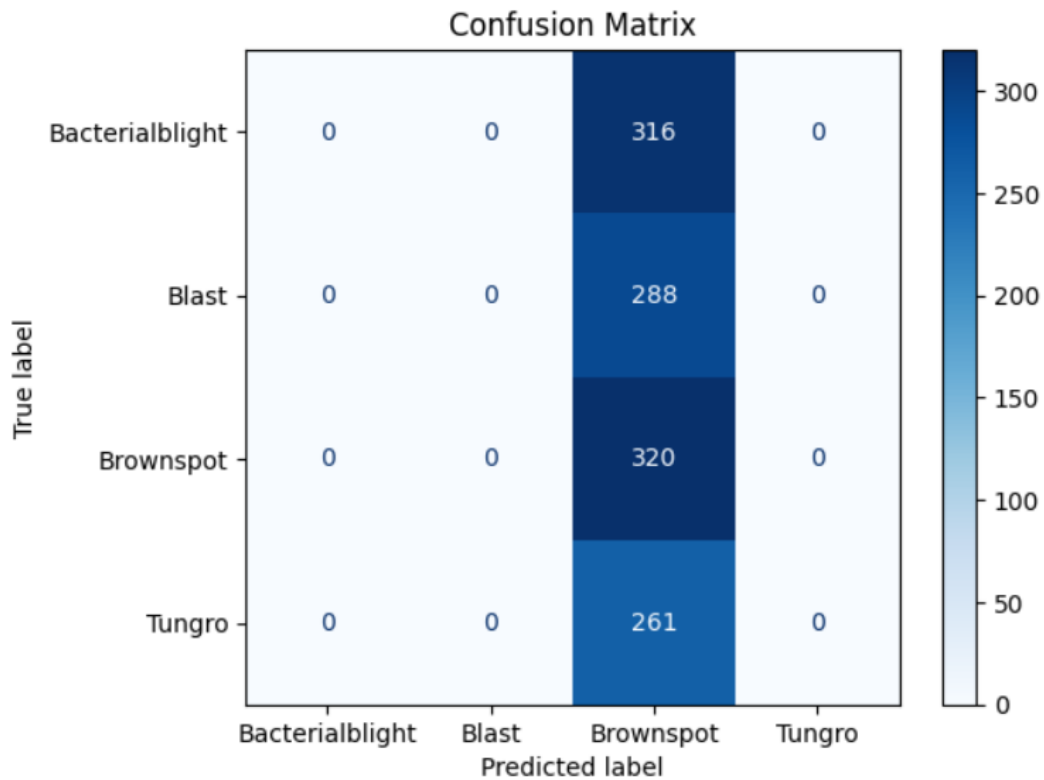
2.4 Xây dựng và đánh giá mô hình

- Xây dựng mô hình: Ba mô hình học sâu được xây dựng và huấn luyện với cấu trúc tương tự
 - Base model: Sử dụng các tầng tiền huấn luyện, ban đầu được đóng băng (trainable = False) để tận dụng kiến thức từ ImageNet.
 - Các lớp bổ sung: Thêm GlobalAveragePooling2D, Dense(128, activation='relu'), Dropout(0.5), và Dense(num_classes, activation='softmax') để phân loại.
 - Cấu hình huấn luyện: Sử dụng optimizer Adam, loss function categorical_crossentropy, batch size 32, và huấn luyện trong 20 epochs
- Đánh giá mô hình MobileNetV2:

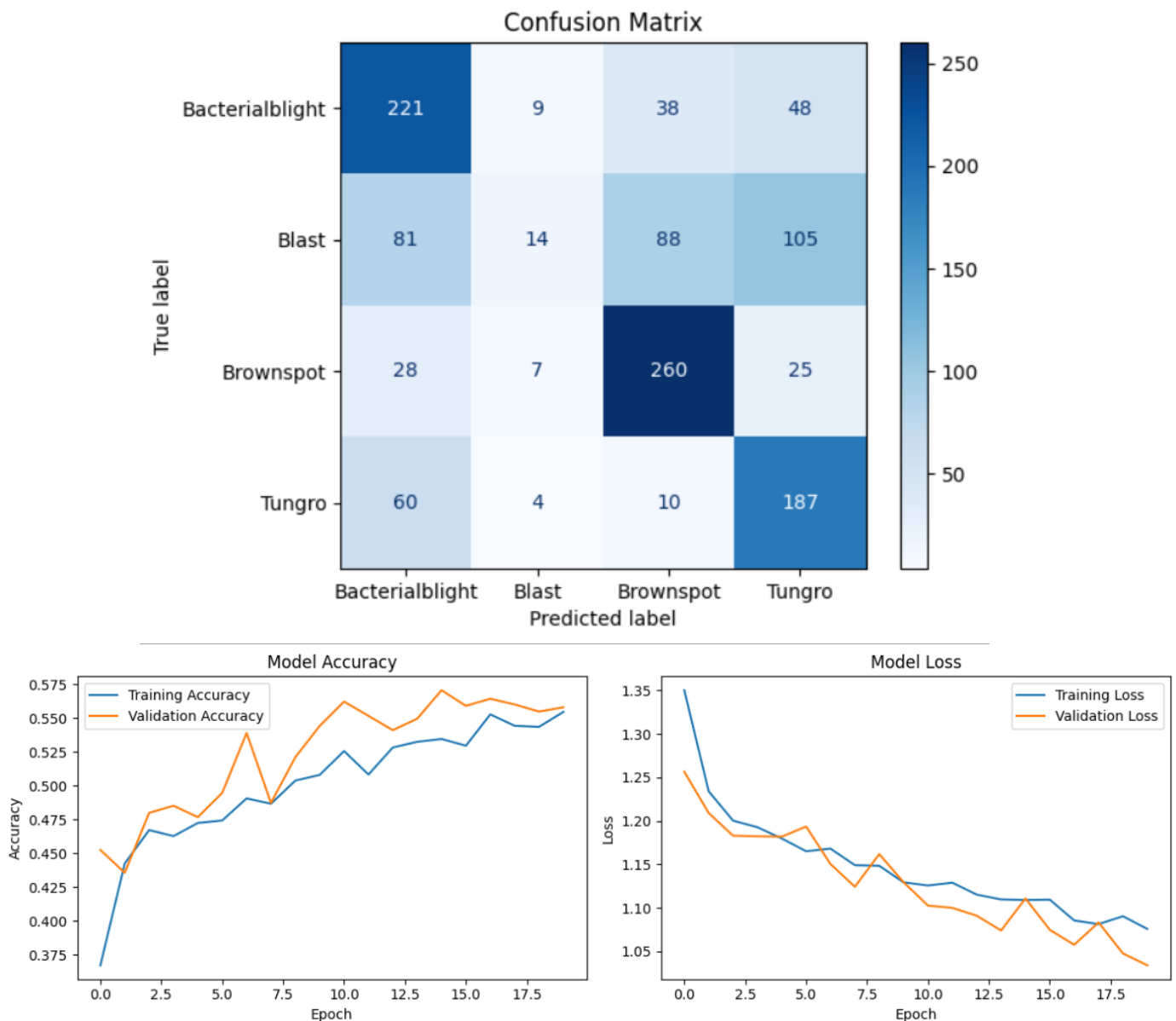


- Confusion Matrix: Gần như hoàn hảo (316/316 Bacterialblight, 281/288 Blast, 320/320 Brownspot, 261/261 Tungro), với nhầm lẫn nhỏ (4 mẫu Blast nhầm thành Bacterialblight, 3 mẫu nhầm thành Brownspot).

- Accuracy: 0.95-1.0 sau 20 epochs, độ chính xác cao
- Loss: Giảm xuống ~0.1, mô hình ít lỗi
- Đánh giá mô hình EfficientNetB0:



- Confusion Matrix: Hoàn hảo (316/316, 288/288, 320/320, 261/261), nhưng không khớp với accuracy/loss, có thể do lỗi huấn luyện.
- Accuracy: ~0.27 sau 20 epochs, độ chính xác rất thấp, bất thường
- Loss: ~1.38, giảm chậm.
- Đánh giá mô hình ResNet50:



- Confusion Matrix: Nhiều nhầm lẫn (221/316 Bacterialblight, 88/288 Blast, 260/320 Brownspot, 187/261 Tungro).
- Accuracy: ~0.55 sau 20 epochs. Độ chính xác không cao
- Loss: ~1.05, giảm từ ~1.35, nhầm lẫn nhiều

2.5 Diễn giải mô hình và trình bày kết quả

- MobileNetV2: Mô hình đạt hiệu suất cao nhất, với khả năng học tốt các đặc trưng của lá lúa. Biểu đồ accuracy/loss cho thấy mô hình hội tụ nhanh và không có dấu hiệu overfitting. Kết quả confusion matrix chứng minh khả năng phân loại vượt trội.
- EfficientNetB0: Dù confusion matrix cho kết quả hoàn hảo, accuracy và loss bất thường (0.27 và 1.38) cho thấy có thể có lỗi trong quá trình huấn luyện, như learning rate không phù hợp hoặc dữ liệu không được xử lý đúng. Mô hình cần điều chỉnh (tăng epochs, giảm learning rate).
- ResNet50: Mô hình có hiệu suất thấp nhất, với accuracy chỉ 0.55 và nhiều nhầm lẫn. Điều này có thể do cấu hình huấn luyện chưa tối ưu (batch size, learning rate) hoặc cần fine-tuning thêm.

- MobileNetV2 cho kết quả tốt nhất, trong khi EfficientNetB0 và ResNet50 cần cải tiến thêm