HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



ĐINH MẠNH HẢI

PHÂN LOẠI MỘT SỐ BỆNH TRÊN CÂY CHÈ DỰA TRÊN HỌC SÂU ÍT MẪU

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(*Theo định hướng nghiên cứu*)

HÀ NỘI - NĂM 2025

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



ĐINH MẠNH HẢI

PHÂN LOẠI MỘT SỐ BỆNH TRÊN CÂY CHÈ DỰA TRÊN HỌC SÂU ÍT MẪU

Chuyên ngành: Hệ thống thông tin
Mã số: 8.48.01.04

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT
(*Theo định hướng nghiên cứu*)

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS. TS. PHẠM VĂN CƯỜNG

HÀ NỘI - NĂM 2025

LỜI CAM ĐOAN

Tôi cam đoan đề án Thạc sĩ Hệ thống thông tin: "Phân loại một số bệnh trên cây chè dựa trên học sâu ít mẫu" được nghiên cứu và thực hiện bởi chính tôi. Những thông tin về nội dung, số liệu và các kết quả được đề cập trong đề án tốt nghiệp này là sự thật, trung thực. Tôi sẽ chịu trách nhiệm hoàn toàn nếu phát hiện bất kỳ hành vi sao chép trái phép hoặc vi phạm quy chế đào tạo.

Hà Nội, tháng 11 năm 2025

Tác giả đề án



Đinh Mạnh Hải

LỜI CẢM ƠN

Lời đầu tiên, tôi xin gửi lời cảm ơn chân thành đến PGS. TS. Phạm Văn Cường. Trong suốt quá trình thực hiện đề án tốt nghiệp lần này, thầy đã định hướng và dìu dắt tôi, giúp tôi có thể vượt qua khó khăn trong quá trình thực hiện và hoàn thiện đề tài nghiên cứu của mình.

Tiếp theo, tôi xin gửi lời cảm ơn sâu sắc đến các thầy cô Giảng viên của Khoa Công nghệ thông tin 1 và Khoa Sau đại học thuộc Học viện Công nghệ Bưu Chính Viễn thông. Suốt thời gian hơn 2 năm học tập Thạc sĩ tại trường, tôi đã được thầy cô truyền đạt rất nhiều kiến thức chuyên môn quý báu. Đây không chỉ là những kiến thức nền tảng để tôi thực hiện đề án tốt nghiệp mà còn là hành trang quý báu để tôi có thể bước vào tương lai.

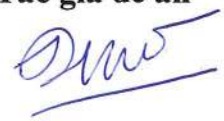
Ngoài ra, tôi cũng muốn gửi lời cảm ơn đến những người cộng sự đã sẵn sàng chia sẻ, tư vấn và hỗ trợ tôi về mặt kỹ thuật trong giai đoạn thực hiện đề án.

Dù đã đạt được một số kết quả nhất định nhưng tôi hiểu rằng đề án này không thể tránh khỏi những thiếu sót và hạn chế. Tôi hi vọng sẽ nhận được những sự thông cảm và các ý kiến đóng góp từ thầy cô để có thể cải thiện, nâng cao chất lượng nghiên cứu này trong tương lai.

Tôi xin chân thành cảm ơn!

Hà Nội, tháng 11 năm 2025

Tác giả đề án



Đinh Mạnh Hải

MỤC LỤC

LỜI CAM ĐOAN i
LỜI CẢM ƠN ii
MỤC LỤC iii
DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT v
DANH SÁCH BẢNG vi
DANH SÁCH HÌNH VẼ vii
MỞ ĐẦU 1
1. Tính cấp thiết của đề tài 1
2. Tổng quan vấn đề nghiên cứu 1
3. Mục tiêu của học ít mẫu FSL 3
4. Đối tượng và phạm vi nghiên cứu 4
5. Phương pháp nghiên cứu 4
CHƯƠNG 1 - TỔNG QUAN 7
1.1. Giới thiệu về bệnh hại trên lá chè 8
1.1.1. Những đặc điểm chung của bệnh hại 8
1.1.2. Tác hại của bệnh hại trên lá chè 9
1.1.3. Tầm quan trọng của việc kiểm soát bệnh hại trên lá chè 10
1.2. Giới thiệu về các loại bệnh và nghiên cứu 11
1.2.1. Giới thiệu về các loại bệnh 11
1.2.2. Mô tả chi tiết bài toán 13
1.2.3. Đề xuất hướng giải quyết 14
1.3. Kết luận chương 17
CHƯƠNG 2 - PHÂN LOẠI BỆNH TRÊN LÁ CHÈ DỰA TRÊN HỌC SÂU ÍT MẪU 18
2.1. Tiền xử lý dữ liệu 19
2.2. Học sâu ít mẫu phân loại bệnh trên lá chè 21
2.2.1. Few-shot learning là gì? 21
2.2.2. Thách thức của Few-shot learning 21

2.2.3. Mô hình kiến trúc 22
2.2.4. Backbone trích xuất Đặc trưng 23
2.2.5. Mô-đun học phép đo (Metric Learning Module) 24
2.2.6. Hàm mất mát 26
2.2.7. Khung tích hợp quan hệ theo Batch (RBI Framework) 27
2.3. Kết luận chương 32
CHƯƠNG 3: THỰC NGHIỆM VÀ ĐÁNH GIÁ 34
3.1. Thiết lập và cấu hình 35
3.2. Xây dựng, huấn luyện mô hình 36
3.2.1. Kiến trúc mô hình đề xuất 36
3.2.2. Quy trình huấn luyện và tối ưu hóa 40
3.3. Kết quả mô hình 42
3.4. Xây dựng ứng dụng application - servers 44
3.5. Kết chương 48
KẾT LUẬN 50
TÀI LIỆU THAM KHẢO 52

DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT

Viết tắt	Tiếng Anh	Tiếng Việt
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
RC	Residual connections	Kết nối tắt
FSL	Few-Shot Learning	Học sâu ít mẫu
TL	Transfer Learning	Học chuyển giao
PN	Prototypical Networks	Mạng nguyên mẫu
RBI	Relationship Batch Integration	Khung Tích hợp Quan hệ
VD	Vertical Duplication	Sao chép dọc
HD	Horizontal Duplication	Sao chép ngang
RRA	Residual Relationship Attention	Cơ chế chú ý quan hệ
RPE	Relationship Position Encoding	Mã hoá vị trí quan hệ
ISM	Image Similarity Matrix	Ma trận tương đồng hình ảnh
PSNR	Peak Signal-to-Noise Ratio	
MAML	Model-Agnostic Meta-Learning	

DANH SÁCH BẢNG

Bảng 1: Thống kê tham số được sử dụng.....35
Bảng 2: Độ chính xác phân loại (%) trên tập kiểm thử bệnh lá chè 43
Bảng 3: Độ chính xác phân loại (%) giữa các kiến trúc 44

DANH SÁCH HÌNH VẼ

Hình 1: Phân bổ số lượng ảnh theo lớp 13
Hình 2: Prototypical networks trong few-shot 25
Hình 3: Kiến trúc Relation network 26
Hình 4: Sơ đồ kiến trúc 36
Hình 5: Cơ chế tập trung và backbone Resnet50 37
Hình 6: Cơ chế tập trung và backbone Resnet101 37
Hình 7: Cơ chế tập trung vào backbon ConvNeXt-Base 38
Hình 8: Giá trị hàm mất mát, độ chính xác trong quá trình huấn luyện với mạng xương sống Resnet101 42
Hình 9: Giá trị hàm mất mát, độ chính xác trong quá trình huấn luyện với mạng xương sống Resnet50 43
Hình 10: Giá trị hàm mất mát, độ chính xác trong quá trình huấn luyện với mạng xương sống ConvNeXt-Base 43
Hình 11: Sơ đồ kiến trúc application - servers 45
Hình 12: Giao diện web 47
Hình 13: Giao diện web phản hồi kết quả 47

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Cây chè (Camellia sinensis) là một trong những cây công nghiệp mũi nhọn tại Khu vực Châu Á và một vài khu vực khác trên thế giới. Tuy nhiên, hiện nay nền công nghiệp trồng chè đang bị đe doạ bởi sâu bệnh, khiến năng suất và chất lượng chè bị suy giảm đáng kể. Đã có nhiều phương pháp phát hiện sâu bệnh được triển khai, tuy nhiên vẫn tồn đọng nhiều hạn chế như: Tốn thời gian, chi phí, độ chính xác không cao và khó để ứng dụng trên quy mô lớn, ... Việc tìm ra giải pháp để phát hiện sớm và chẩn đoán đúng tình trạng sâu bệnh trên lá chè trở nên quan trọng và cấp thiết hơn bao giờ hết.

Trong bối cảnh nông nghiệp thông minh (Precision Agriculture), việc tự động hóa quy trình giám sát sức khỏe cây trồng, ứng dụng thị giác máy tính để nhận diện bệnh hại cây trồng đã trở thành một xu hướng và yêu cầu cấp thiết [6, 7, 15, 17, 18, 19, 20]. Các nghiên cứu mới nhất cũng chỉ ra vai trò quan trọng của trí tuệ nhân tạo trong việc phát hiện bệnh hại nông nghiệp, tối ưu hóa quy trình canh tác.

Để trình bày nội dung một cách hệ thống và khoa học, đề án sẽ được cấu trúc thành các chương chính như sau:

- **Chương 1: Tổng quan về bài toán bệnh hại trên lá chè.** Chương này sẽ trình bày những khái niệm tổng quan nhất về bệnh hại trên cây, vấn đề xung quanh và các nghiên cứu liên quan cũng sẽ được tổng hợp trong chương này.
- **Chương 2: Phân loại bệnh trên lá chè dựa trên học sâu ít mẫu.** Chương này sẽ trình bày chi tiết hơn về các phương pháp để giải quyết bài toán phân loại bệnh trên lá chè.
- **Chương 3: Thực nghiệm và đánh giá.** Chương này sẽ trình bày về kết quả thực nghiệm và đánh giá hiệu quả của mô hình học sâu ít mẫu đã đề xuất.

2. Tổng quan vấn đề nghiên cứu

Trong thập kỷ qua, Học sâu (Deep Learning) đã tạo ra cuộc cách mạng trong lĩnh vực nhận dạng hình ảnh. Sự ra đời của các kiến trúc Mạng nơ-ron tích chập

(CNN) đã đẩy hiệu suất phân loại ảnh lên mức vượt trội so với các phương pháp truyền thống.

Các kiến trúc nền tảng như AlexNet đã mở đầu cho kỷ nguyên CNN hiện đại.

Tiếp đó, sự xuất hiện của ResNet với các kết nối RCđã giải quyết vấn đề đạo hàm về 0, cho phép huấn luyện các mạng sâu.

Kiến trúc DenseNet tăng cường khả năng tái sử dụng đặc trưng, hay EfficientNet tối ưu hóa sự cân bằng giữa độ sâu và độ phân giải của mạng, đã trở thành xương sống (backbone) cho nhiều bài toán thị giác máy tính.

Tuy nhiên, sự thành công của các mô hình học sâu CNN truyền thống phụ thuộc rất lớn vào dữ liệu lớn (Big Data). Để đạt được độ chính xác cao, các mô hình này yêu cầu hàng nghìn, thậm chí hàng triệu hình ảnh được gán nhãn thủ công cho mỗi lớp đối tượng. Đây là một rào cản khổng lồ đưa vào ứng dụng thực tế với cây chè trong nông nghiệp.

Trong môi trường tự nhiên, việc thu thập dữ liệu bệnh hại gặp 3 trở ngại sau:

- Sự xuất hiện của các loại bệnh mới hoặc hiếm gặp khiến số lượng mẫu thu được rất ít.
- Chi phí gán nhãn dữ liệu bởi các chuyên gia nông nghiệp là rất cao và tốn thời gian.
- Sự biến đổi của môi trường, ánh sáng và giai đoạn phát triển của bệnh làm cho dữ liệu trở nên phức tạp.

Khi số lượng mẫu huấn luyện hạn chế, các mô hình học sâu truyền thống thường rơi vào tình trạng quá khớp (overfitting), từ đó khiến khả năng tổng quát hoá kém hiệu quả trên dữ liệu thực tế. Điều này yêu cầu phải tìm kiếm các phương pháp học tập mới có khả năng hoạt động hiệu quả với lượng dữ liệu ít [6, 7, 15, 17, 18, 19, 20].

Nhiều kỹ thuật FSL mới đã được đề xuất với đối tượng cây trồng. Ví dụ, việc sử dụng miền tần số (frequency domain) để trích xuất đặc trưng bệnh hại [23], hay kết hợp học bán giám sát [19] để tận dụng dữ liệu chưa gán nhãn. Các mô hình kết hợp (ensemble) sử dụng trọng số chuỗi cũng cho thấy hiệu quả tốt khi nhận dạng

bệnh cây trồng. Bên cạnh đó, các kỹ thuật như nén và kích hoạt đa tỷ lệ (multi-scale feature fusion) [23] kết hợp với cơ chế chú ý RBI cũng được áp dụng để nắm bắt các đặc điểm qua hệ bệnh giữa các mẫu.

3. Mục tiêu của học ít mẫu FSL

Để giải quyết vấn đề khan hiếm dữ liệu, FSL đã nổi lên như một hướng nghiên cứu đầy tiềm năng. FSL hướng tới việc mô phỏng khả năng tư duy của con người: có thể nhận diện một đối tượng mới chỉ sau khi quan sát một vài ví dụ mẫu.

Các nghiên cứu nền tảng về FSL đã đề xuất nhiều hướng tiếp cận khác nhau để "học cách học" (meta-learning):

- **Phương pháp dựa trên độ đo (Metric-based):** Các mạng như Matching Networks [3] sử dụng cơ chế chú ý và bộ nhớ ngoài để so sánh sự tương đồng giữa các mẫu. PN học một không gian metric nơi các lớp được biểu diễn bởi một điểm nguyên mẫu - là trung bình của các vector đặc trưng hỗ trợ. Relation Networks [1] tiến xa hơn bằng cách học một mô-đun quan hệ phi tuyến để đánh giá độ tương đồng thay vì sử dụng các khoảng cách cố định như Euclidean hay Cosine.
- **Phương pháp dựa trên tối ưu hóa (Optimization-based):** Tiêu biểu là MAML [4, 5] cho phép tìm kiếm một bộ tham số khởi tạo tốt để mô hình có thể thích nghi nhanh chóng với tác vụ mới chỉ sau vài bước cập nhật gradient. Các cải tiến sau đó về quy trình huấn luyện MAML cũng đã giúp tăng tính ổn định và hiệu quả của phương pháp này.

Các nghiên cứu sâu hơn về cơ chế phân loại trong FSL và các kỹ thuật học động không quên kiến thức cũ [8] đã củng cố nền tảng lý thuyết chắc chắn cho việc áp dụng FSL vào các bài toán với tập dữ liệu phức tạp như bệnh lá chè.

Việc áp dụng học tương phản có giám sát (supervised contrastive learning) cũng đã được chứng minh là giúp cải thiện khả năng phân biệt giữa các loại bệnh có triệu chứng tương tự nhau trong điều kiện ít mẫu. Các khung làm việc Meta-learning cũng đang được tinh chỉnh cụ thể cho miền nông nghiệp [22].

4. Đối tượng và phạm vi nghiên cứu

Phạm vi nghiên cứu của đề tài trước hết được xác định ở khía cạnh đối tượng, trong đó tập trung vào các hình ảnh lá chè thuộc mười lớp dữ liệu gồm lá khỏe và chín loại bệnh phổ biến. Việc phân tích được giới hạn ở dấu hiệu bệnh thể hiện trên phiến lá, không mở rộng sang các bộ phận khác như thân, rễ hay búp chè. Đề tài cũng không hướng tới việc bao quát toàn bộ các loại bệnh hại tồn tại trong thực tế mà chỉ xem xét nhóm bệnh xuất hiện trong tập dữ liệu được thu thập và chuẩn hóa phục vụ cho mục tiêu nghiên cứu.

Ở phương diện kỹ thuật và mô hình, phạm vi nghiên cứu được giới hạn trong các phương pháp học sâu ít mẫu, đặc biệt là những tiếp cận dựa trên độ đo như Prototypical Networks và cơ chế tăng cường quan hệ RBI [4]. Các mô hình CNN truyền thống yêu cầu lượng dữ liệu lớn chỉ được sử dụng với vai trò mạng trích xuất đặc trưng. Bên cạnh đó, đề tài không triển khai các hướng tiếp cận khác như các mô hình dựa trên Transformer, học tăng cường hay các phương pháp sinh dữ liệu như GAN.

Ở khía cạnh ứng dụng, đề tài chỉ triển khai một hệ thống mẫu cho phép người dùng tải ảnh và nhận kết quả phân loại. Hệ thống này nhằm minh chứng tính khả thi của mô hình có thể áp dụng trong nông nghiệp quy mô lớn hay một hệ thống IoT tích hợp nhiều thiết bị cảm biến ngoài thực tế sản xuất.

Cuối cùng, phạm vi đánh giá mô hình được giới hạn trong hai kịch bản phổ biến của bài toán học ít mẫu là 5-way 1-shot và 5-way 5-shot. Các tiêu chí đánh giá được xây dựng tập trung vào độ chính xác phân loại trong khuôn khổ dữ liệu ít mẫu, không mở rộng sang các tiêu chí đánh giá mang tính công nghiệp như khả năng chịu tải, tối ưu mô hình trên thiết bị biên hay độ ổn định của mô hình trong môi trường triển khai dài hạn.

5. Phương pháp nghiên cứu

Dù hiện nay đã có nhiều nghiên cứu về FSL cho cây trồng nói chung, nhưng các nghiên cứu chuyên sâu cho cây chè vẫn còn hạn chế và tồn tại những thách thức riêng biệt.

- **Đặc thù của bệnh hại chè:** Các bệnh trên lá chè thường có hình thái phức tạp, dễ nhầm lẫn và chịu tác động bởi yếu tố môi trường. Các mô hình TL thông thường có thể không đạt hiệu quả tối ưu nếu không có sự tinh chỉnh phù hợp với đặc trưng hình thái của lá chè.
- **Các nghiên cứu hiện có và hạn chế:** Một số nghiên cứu tiên phong đã áp dụng phương pháp học ít mẫu cho bệnh chè và đánh giá các phương pháp FSL khác nhau trên miền dữ liệu này. Gần đây, mô hình MAF-MixNet sử dụng cơ chế chú ý hỗn hợp đã được đề xuất cho bệnh chè. Tuy nhiên, vẫn còn thiếu các nghiên cứu so sánh toàn diện về hiệu quả của các kiến trúc mạng quan hệ (Relation Network) (1) so với các phương pháp dựa trên nguyên mẫu (Prototypical Network) (2) hay tối ưu hóa (MAML) cụ thể trên tập dữ liệu bệnh chè Việt Nam hoặc các bộ dữ liệu chè chuẩn hóa.
- **Yêu cầu về độ chính xác và tốc độ:** Để triển khai thực tế, mô hình không chỉ cần chính xác trên dữ liệu ít mẫu mà còn phải đảm bảo tốc độ phán đoán và khả năng mở rộng. Việc kết hợp các backbone hiện đại như EfficientNet hay ResNet với các đầu phân loại FSL như Relation Network (1) là một hướng đi đầy hứa hẹn ngữ cảnh bệnh hại chè.

Có thể thấy rằng việc nghiên cứu đề tài "Phân loại một số bệnh trên cây chè dựa trên học sâu ít mẫu" là vô cùng cấp thiết và có ý nghĩa khoa học cũng như thực tiễn sâu sắc. Đề tài không chỉ giải quyết bài toán nút thắt về dữ liệu đang kìm hãm ứng dụng AI trong nông nghiệp, mà còn góp phần hoàn thiện các phương pháp luận về Few-Shot Learning trong một miền ứng dụng cụ thể và phức tạp.

Việc thực hiện đề tài này sẽ:

1. Thu thập dữ liệu từ 2 tập dữ liệu bệnh lá chè được công khai, “Identifying Disease in Tea leaves” trên Kaggle và “Tea Leaf Dataset” trên Mendeley.

2. Kế thừa các thành tựu của các kiến trúc CNN tiên tiến và các kỹ thuật tối ưu hóa.

3. Áp dụng và cải tiến các thuật toán FSL hiện đại như Prototypical Network kết với cơ chế tập trung RBI [4].

Đây là bước đi cần thiết để đưa công nghệ cao vào canh tác chè, giúp nâng cao năng suất và chất lượng lá chè, từ đó tạo tiền đề phát triển bền vững nền nông nghiệp trong tương lai.

CHƯƠNG 1 - TỔNG QUAN

Trong kỷ nguyên chuyển đổi số mạnh mẽ hiện nay, sự ra đời của nông nghiệp thông minh (Smart Agriculture) và nông nghiệp chính xác (Precision Agriculture) đang định hình lại hoàn toàn phương thức canh tác truyền thống. Tại đây, việc giám sát sức khỏe cây trồng không còn đơn thuần dựa vào kinh nghiệm cảm quan của người nông dân mà đã trở thành một quy trình kỹ thuật số dựa trên dữ liệu, nơi công nghệ thị giác máy tính đóng vai trò tiên quyết. Khả năng tự động hóa quy trình nhận diện và chẩn đoán sớm các loại bệnh hại với tốc độ xử lý thời gian thực và độ chính xác vượt trội so với mắt thường chính là chìa khóa để bảo vệ năng suất, giảm thiểu việc lạm dụng thuốc bảo vệ thực vật và hướng tới một nền nông nghiệp bền vững. Tuy nhiên, việc triển khai các mô hình trí tuệ nhân tạo vào thực tế, đặc biệt là đối với cây chè, đang vấp phải một rào cản kỹ thuật to lớn: sự khan hiếm và đắt đỏ của dữ liệu gán nhãn chất lượng cao.

Thực tế cho thấy, các phương pháp học sâu truyền thống (Deep Learning) dù mạnh mẽ nhưng lại phụ thuộc quá lớn vào dữ liệu lớn (Big Data). Khi đối mặt với các loại bệnh hiếm gặp, mới phát sinh hoặc có hình thái phức tạp trên lá chè, các mô hình này thường thất bại trong việc tổng quát hóa do không đủ mẫu huấn luyện. Chính vì vậy, việc nghiên cứu các cơ chế học tập mới có khả năng mô phỏng tư duy nhận thức của con người – học nhanh và chính xác chỉ từ một vài ví dụ mẫu (Few-shot Learning) – không chỉ là một thách thức khoa học thú vị mà còn là yêu cầu cấp bách để giải quyết bài toán "nút thắt cổ chai" về dữ liệu trong sản xuất.

Với vai trò là chương mở đầu, đặt nền móng kiến thức cho toàn bộ đề tài, Chương 1 sẽ cung cấp một cái nhìn toàn cảnh và hệ thống về vấn đề nghiên cứu. Nội dung chương sẽ bắt đầu bằng việc đi sâu phân tích các đặc trưng sinh học và tác hại của các nhóm bệnh phổ biến trên cây chè, làm rõ sự phức tạp trong hình thái vết bệnh để thấy được độ khó của bài toán nhận diện. Tiếp đó, chúng ta sẽ khảo sát các cơ sở lý thuyết nền tảng của mạng nơ-ron tích chập (CNN) và các kỹ thuật Học sâu ít mẫu hiện đại, đồng thời tổng hợp các nghiên cứu liên quan để xác định rõ những khoảng

trống công nghệ mà đề tài cần giải quyết. Những kiến thức này sẽ là tiền đề vững chắc để phát triển các giải pháp chuyên sâu được trình bày ở các chương sau.

1.1. Giới thiệu về bệnh hại trên lá chè

1.1.1. Những đặc điểm chung của bệnh hại

Trong canh tác nông nghiệp, bệnh hại trên lá chè [20] không chỉ đơn thuần là các tổn thương vật lý bề mặt mà là kết quả của một quá trình biến đổi sinh lý và hóa sinh phức tạp, diễn ra khi cây chè phản ứng lại sự xâm nhiễm của các tác nhân gây hại như nấm, vi khuẩn, tảo hoặc côn trùng. Đặc điểm chung bao trùm lên hầu hết các loại bệnh này là khả năng phá hủy cấu trúc diệp lục và tế bào biểu bì. Về mặt thị giác, các bệnh hại này thể hiện tính chất biến thiên cực kỳ cao theo không gian và thời gian. Một vết bệnh đơn lẻ có thể thay đổi hình thái liên tục từ giai đoạn ủ bệnh (thường là các chấm nhỏ khó quan sát) sang giai đoạn phát triển mạnh (lan rộng, thay đổi màu sắc từ vàng sang nâu hoặc đen) và cuối cùng là giai đoạn hoại tử (khô, rách hoặc hình thành bào tử nấm). Chính sự biến đổi liên tục này tạo ra thách thức lớn về sự "đa dạng nội lớp" (intra-class variation), nghĩa là cùng một loại bệnh nhưng lại có hình dạng, kích thước và màu sắc hoàn toàn khác nhau tùy theo điều kiện môi trường (ánh sáng, độ ẩm...) và giai đoạn phát triển của cây.

Đặc điểm chung quan trọng khác làm gia tăng độ khó cho bài toán nhận diện là sự "tương đồng giữa các lớp" (inter-class similarity). Các tác nhân gây bệnh khác nhau thường tạo ra các triệu chứng phản ứng tương tự trên phiến lá, ví dụ như hiện tượng cháy mép, đốm vòng hay biến vàng cục bộ. Các vết bệnh này thường không tồn tại độc lập mà xuất hiện xen kẽ, chồng chéo lên nhau trên cùng một bề mặt lá, hòa lẫn với các yếu tố nhiễu nền như bụi bẩn, sương mù hoặc các vết rách cơ học. Do đó, để phân loại chính xác, hệ thống không chỉ dựa vào các đặc trưng cơ bản như màu sắc hay hình dáng, mà buộc phải học được các đặc trưng trừu tượng hơn về kết cấu bề mặt, cấu trúc viền vết bệnh của các tổn thương.

Trên cơ sở các đặc điểm chung đó, tập dữ liệu nghiên cứu được xây dựng nhằm bao quát các đại diện tiêu biểu nhất cho các hình thái bệnh lý này, bên cạnh nhóm lá khỏe mạnh dùng làm đối chứng. Cụ thể, nhóm bệnh gây hoại tử và biến đổi màu sắc

diện rộng bao gồm bệnh thán thư, bệnh cháy nâu và bệnh chấm xám, thường đặc trưng bởi các vết cháy loang lổ hoặc các vòng đồng tâm sẫm màu. Nhóm bệnh gây tổn thương dạng điểm khu trú bao gồm bệnh đốm mắt cua, đốm trắng và đốm đỏ, với các kích thước nhỏ và ranh giới vết bệnh rõ rệt hơn. Đặc biệt, sự đa dạng của tập dữ liệu còn được thể hiện qua các loại bệnh có kết cấu bề mặt độc đáo như bệnh ghẻ sẹo đỏ - tạo vết sần sùi, bệnh đốm tảo tạo lớp nhung mượt do tảo ký sinh và các tổn thương giả bệnh do côn trùng gây ra.

1.1.2. Tác hại của bệnh hại trên lá chè

Bệnh hại trên lá chè gây ra nhiều tác động tiêu cực kéo dài lên cả năng suất, chất lượng và sức sinh trưởng của cây, cũng như ảnh hưởng trực tiếp đến hiệu quả kinh tế của người trồng. Trước hết, các loại bệnh như thán thư, cháy nâu hay chấm xám làm tổn thương mô lá và phá hủy cấu trúc diệp lục, dẫn đến giảm hiệu suất quang hợp - quá trình sống còn đối với sinh trưởng của cây chè. Khi diện tích lá bị phá hủy tăng dần, khả năng sinh trưởng bị suy giảm rõ rệt, khiến cây còi cọc, chậm phát triển và khả năng chống chịu với điều kiện khắc nghiệt của môi trường kém.

Bệnh hại trên lá chè không chỉ đơn thuần là những tổn thương sinh học cục bộ mà thực sự là một rào cản nghiêm trọng, gây ra chuỗi tác động tiêu cực đa chiều kéo dài từ sức khỏe sinh lý của cây trồng, chất lượng thương phẩm cho đến hiệu quả kinh tế vĩ mô của người nông dân và sự bền vững của môi trường sinh thái.

Thứ nhất, xét trên phương diện sinh học và nông học, sự xâm nhiễm của các tác nhân gây bệnh phổ biến như thán thư, cháy nâu hay chấm xám đã trực tiếp tấn công và phá hủy hệ thống mô lá, nơi đóng vai trò là "nhà máy năng lượng" của cây chè. Quá trình hoại tử này làm phá vỡ cấu trúc diệp lục, dẫn đến sự sụt giảm nghiêm trọng hiệu suất quang hợp – cơ chế sống còn để cây tổng hợp chất hữu cơ nuôi dưỡng cơ thể. Khi diện tích lá xanh bị thu hẹp dần bởi các vết bệnh loang lổ, khả năng sinh trưởng của cây bị kìm hãm rõ rệt; cây trở nên còi cọc, búp ra thưa thớt, chậm phát triển và suy giảm đáng kể sức đề kháng tự nhiên trước các điều kiện khắc nghiệt của thời tiết như hạn hán hay sương muối.

Thứ hai, bệnh hại là nguyên nhân trực tiếp làm xói mòn giá trị thương mại và phẩm cấp của sản phẩm chè. Nguyên nhân sâu xa nằm ở những biến đổi rối loạn trong quá trình sinh lý – hóa sinh bên trong lá nhiễm bệnh, làm đảo lộn tỷ lệ cân bằng của các thành phần dược tính quan trọng như polyphenol, axit amin và các sắc tố tự nhiên. Hậu quả là chè nguyên liệu sau khi chế biến bị mất đi hương vị tinh túy đặc trưng, thay vào đó là vị đắng chát kém chất lượng, màu nước pha trở nên xỉn đục thiếu thẩm mỹ và thậm chí xuất hiện các mùi lạ khó chịu. Sự xuống cấp về mặt cảm quan này ảnh hưởng trực tiếp đến doanh thu tài chính của người trồng do bị ép giá, đồng thời làm giảm uy tín và vị thế cạnh tranh của thương hiệu chè vùng miền trên thị trường quốc tế.

Thứ ba, một hệ lụy nghiêm trọng mang tính xã hội và môi trường là nguy cơ mất an toàn vệ sinh thực phẩm do áp lực phòng trừ sâu bệnh. Trong điều kiện khí hậu nóng ẩm đặc trưng, tốc độ lây lan của nấm bệnh diễn ra cực kỳ nhanh chóng, dễ dàng bùng phát thành dịch trên diện rộng nếu không được phát hiện và khoanh vùng kịp thời. Tâm lý lo lắng, sốt ruột trước nguy cơ mất trắng vụ thu hoạch thường thúc đẩy người nông dân lạm dụng thuốc bảo vệ thực vật hoặc sử dụng quá liều lượng quy định nhằm dập dịch nhanh chóng.

Cuối cùng, tác hại của bệnh hại còn mang tính chất tích lũy và kéo dài, đe dọa sự ổn định của cấu trúc vườn trồng trong dài hạn. Trong trường hợp bệnh tái diễn liên tục hoặc kéo dài không dứt, cấu trúc tán lá và sức sống tổng thể của vườn chè sẽ bị suy kiệt nghiêm trọng, buộc người canh tác phải thực hiện các biện pháp can thiệp mạnh tay như đốn đau để trẻ hóa hoặc thậm chí phải phá bỏ hoàn toàn để trồng mới. Những gián đoạn này làm đứt gãy chu kỳ sản xuất trong nhiều năm, gây thiệt hại năng suất lâu dài và ảnh hưởng tiêu cực đến chiến lược phát triển bền vững của cả chuỗi cung ứng ngành chè.

1.1.3. Tầm quan trọng của việc kiểm soát bệnh hại trên lá chè

Trước hết, sâu bệnh hại trên lá chè sẽ làm giảm năng suất, sản lượng lá chè thu hoạch. Các loại bệnh phổ biến như thán thư, phồng lá hay chấm xám thường tấn công vào búp non và lá bánh tẻ - những phần mang lại giá trị kinh tế chính. Khi diện tích

lá bị tổn thương, khả năng quang hợp của cây giảm sút, dẫn đến sinh trưởng kém, còi cọc và rụng lá hàng loạt. Trong những đợt dịch bùng phát mạnh, người nông dân có thể bị mất trắng cả vụ thu hoạch, mất mát cả công sức và chi phí giống cây, phân bón, ...

Bên cạnh sự thiệt hại về số lượng, bệnh hại còn làm giảm chất lượng của chè thương phẩm. Lá chè bị nhiễm bệnh thường trải qua những biến đổi hóa sinh bất lợi, thường có vị đắng chát không đạt chuẩn, màu nước pha xỉn màu và mất đi hương thơm đặc trưng. Sự suy giảm chất lượng này khiến giá trị thương mại của sản phẩm giảm sút mạnh, chè bị ép giá hoặc bị từ chối bởi các thị trường khó tính, làm giảm uy tín của thương hiệu chè địa phương và quốc gia trên trường quốc tế.

Trong trường hợp sâu bệnh bùng phát, người dân trồng chè có thể lo lắng quá độ, dẫn đến việc lạm dụng lượng lớn thuốc bảo vệ thực vật hoá học để kiểm soát vấn đề. Việc phun thuốc bảo vệ thực vật hoá học quá liều sẽ dẫn đến nguy cơ tồn dư chất hoá học độc hại trong chè thành phẩm, ảnh hưởng trực tiếp đến sức khỏe của người tiêu dùng và gây ô nhiễm nguồn nước, đất đai và tiêu diệt các loài thiên địch có lợi, làm mất cân bằng hệ sinh thái nông nghiệp.

Cuối cùng, bệnh hại trên lá chè còn gây ra những tổn thất lâu dài về mặt cấu trúc vườn trồng. Một số loại bệnh nguy hiểm không chỉ gây hại trên lá mà còn lây lan xuống thân và cành, làm suy yếu sức đề kháng tổng thể của cây chè lâu năm. Việc này buộc người nông dân phải tiến hành đốn đau hoặc thậm chí phá bỏ để trồng mới, làm gián đoạn chu kỳ kinh doanh trong nhiều năm. Tóm lại, có thể thấy bệnh hại trên lá chè không chỉ ảnh hưởng đến năng suất, chất lượng cây chè mà còn ảnh hưởng đến vấn đề kinh tế, môi trường và đời sống người dân nếu tình trạng sâu bệnh không được phát hiện và kiểm soát kịp thời.

1.2. Giới thiệu về các loại bệnh và nghiên cứu

1.2.1. Giới thiệu về các loại bệnh

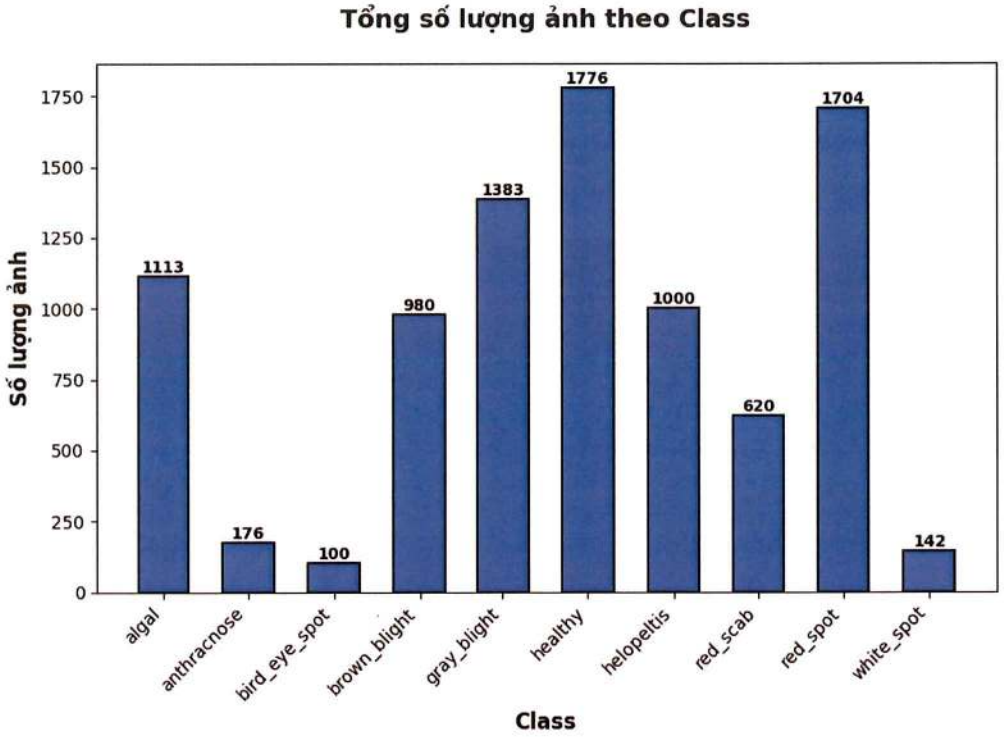
Các bệnh hại trên lá chè còn là kết quả của các quá trình xâm nhiễm vi sinh vật đa dạng, làm biến đổi cấu trúc tế bào, suy giảm chức năng quang hợp và ảnh hưởng trực tiếp đến phẩm cấp chè thương phẩm.

Nhóm bệnh phổ biến nhất thường có tác nhân nấm gây hoại tử mô lá diện rộng. Điển hình là bệnh Thán thư (Anthracnose) và các loại bệnh cháy lá như Cháy nâu (Brown blight) và Chấm xám (Gray blight). Đặc điểm chung của những bệnh này là sự xuất hiện của các vết tổn thương lớn, thường bắt đầu từ mép hoặc chóp lá và lan dần vào trong. Nếu như thán thư tạo ra các vết bệnh màu nâu sẫm, khô xác với ranh giới rõ ràng, thì bệnh chấm xám và cháy nâu (thường do nấm Pestalotiopsis spp. gây ra) lại đặc trưng bởi các vòng đồng tâm màu xám trắng hoặc nâu đen, trên đó thường xuất hiện các hạt nhỏ li ti là cơ quan sinh sản của nấm. Những tổn thương này làm giảm đáng kể diện tích quang hợp, khiến cây sinh trưởng kém và lá chè thành phẩm bị vụn nát, mất hương vị.

Bên cạnh các bệnh gây hoại tử lớn, lá chè còn chịu sự tấn công của các loại bệnh tạo đốm biệt lập với hình thái tinh vi hơn, đòi hỏi sự quan sát chi tiết để phân biệt. Bệnh Đốm mắt cua (Bird eye spot) là một ví dụ điển hình, biểu hiện qua các đốm nhỏ hình tròn với tâm màu xám trắng và viền nâu đỏ, tựa như mắt chim. Tương tự, các bệnh như Đốm trắng (White spot) và Đốm đỏ (Red spot) tạo ra các tổn thương dạng chấm nhỏ rải rác trên phiến lá, tuy kích thước không lớn nhưng mật độ xuất hiện dày đặc có thể gây vàng lá và rụng sớm. Ngoài ra, bệnh Ghẻ sẹo đỏ (Red scab) lại mang đến một thách thức khác về mặt thị giác máy tính khi tạo ra các vết sần sùi, nổi gờ trên bề mặt lá, đòi hỏi mô hình không chỉ nhận diện màu sắc mà còn phải phân tích được đặc trưng kết cấu (texture).

Không chỉ giới hạn ở các bệnh do nấm, danh mục các bệnh hại cần nghiên cứu còn bao gồm các tác nhân sinh học đặc thù khác như tảo và côn trùng. Bệnh Đốm tảo (Algal leaf spot), hay còn gọi là bệnh chè đỏ, do loài tảo ký sinh Cephaleuros virescens gây ra, tạo nên các vết bệnh phủ lớp nhung mượt màu cam đỏ hoặc xanh xám, thường xuất hiện ở các lá già và cành, làm suy yếu sức đề kháng của cây. Song song đó, sự hiện diện của các tổn thương do Bọ xít muỗi (Helopeltis) là một yếu tố gây nhiễu đáng kể. Mặc dù đây là vết châm hút của côn trùng, nhưng các vết thâm đen khô lại trên búp non và lá non có hình thái rất giống với các vết bệnh hoại tử do nấm, dễ dẫn đến chẩn đoán sai nếu không được phân tích kỹ lưỡng.

Nhằm phục vụ nghiên cứu, ctập dữ liệu được thu thập từ các tập dữ liệu công khai.



Hình 1: Phân bổ số lượng ảnh theo lớp

1.2.2. Mô tả chi tiết bài toán

Trong phạm vi nghiên cứu bài toàn phát hiện bệnh trên lá chè được định nghĩa là một bài toán phân loại hình ảnh đa lớp dưới điều kiện ràng buộc nghiêm ngặt về dữ liệu, hay còn được gọi là bài toán Học ít mẫu (Few-shot Learning). Khác với các bài toán nhận diện truyền thống vốn yêu cầu hàng nghìn mẫu dữ liệu để tối ưu hóa tham số, mục tiêu cốt lõi tại đây là xây dựng một hệ thống có khả năng "học cách học" (meta-learning), cho phép mô hình tổng quát hóa tri thức để nhận diện chính xác các loại bệnh mới chỉ từ một vài ví dụ mẫu đầu vào (Support set). Điều này mô phỏng lại quá trình nhận thức của con người, khi một chuyên gia nông nghiệp chỉ cần quan sát vài mẫu bệnh lạ là có thể ghi nhớ và nhận biết chúng trong các lần gặp sau.

Không gian dữ liệu của bài toán được cấu thành từ tập hợp các hình ảnh lá chè thuộc 10 lớp trạng thái khác nhau, bao gồm trạng thái khỏe mạnh và 9 nhóm bệnh lý phức tạp. Thách thức lớn nhất của bài toán nằm ở sự chồng chéo giữa các đặc trưng thị giác (inter-class similarity) và sự biến thiên lớn trong cùng một loại bệnh (intra-class variation). Cụ thể, hệ thống phải phân biệt được nhóm bệnh gây hoại tử có hình

thái rất giống nhau như bệnh thán thư, cháy nâu và chấm xám, nơi các vết bệnh đều biểu hiện dưới dạng các mảng nâu sẫm lan rộng và khô xác. Song song đó, mô hình cũng phải xử lý các chi tiết tinh vi ở nhóm bệnh dạng đốm nhỏ như đốm mắt cua, đốm trắng và đốm đỏ, đòi hỏi khả năng trích xuất đặc trưng cục bộ cực kỳ chính xác. Đặc biệt, bài toán trở nên phức tạp hơn với sự xuất hiện của các tổn thương có kết cấu bề mặt độc đáo như vết sần sùi của bệnh ghẻ sẹo đỏ, lớp nhung mượt của bệnh đốm tảo, hay các vết châm kim do côn trùng bọ xít muỗi gây ra - những tác nhân rất dễ gây nhiễu cho các thuật toán thị giác máy tính thông thường.

Về mặt hình thức hóa, bài toán được thiết lập dưới dạng cấu trúc N-way K-shot, trong đó mô hình được cung cấp N lớp bệnh cần phân loại, nhưng mỗi lớp chỉ có K ví dụ mẫu (Support set) với K là số nguyên rất nhỏ, thường là 1, 5 hoặc 10. Nhiệm vụ của mô hình là ánh xạ hình ảnh truy vấn (Query image) vào một không gian đặc trưng nhiều chiều (embedding space), sao cho khoảng cách giữa các véc-tơ đặc trưng của cùng một loại bệnh được thu nhỏ tối đa, trong khi khoảng cách giữa các loại bệnh khác nhau được nới rộng triệt để. Đầu ra của bài toán là nhãn phân loại duy nhất có độ tương đồng cao nhất với ảnh đầu vào. Sự thành công của bài toán được đánh giá qua độ chính xác trên tập kiểm thử, cùng với đó khả năng thích nghi nhanh chóng: khi xuất hiện một biến thể bệnh mới trên đồi chè, hệ thống chỉ cần cập nhật một vài ảnh mẫu là có thể vận hành mà không cần tốn kém tài nguyên để huấn luyện lại.

1.2.3. Đề xuất hướng giải quyết

Để giải quyết những khó khăn về khan hiếm dữ liệu mẫu và tính phức tạp trong đặc trưng hình thái của hệ sinh thái bệnh hại chè, đề tài đề xuất hướng tiếp cận dựa trên kỹ thuật học sâu ít mẫu, cụ thể là khai thác nhóm các mô hình học dựa trên độ đo (Metric-based Learning). Thay vì đi theo lối mòn của các mô hình học giám sát truyền thống - vốn cố gắng tối ưu hóa hàm mất mát để phân lớp trực tiếp trên một tập dữ liệu cố định, giải pháp này chuyển dịch trọng tâm sang việc huấn luyện mô hình "học cách so sánh". Chiến lược cốt lõi là xây dựng một không gian véc-tơ tối ưu, nơi sự tương đồng về mặt bệnh học được lượng hóa bằng khoảng cách hình học, giúp mô

hình duy trì hiệu năng cao bất kể sự thay đổi của điều kiện môi trường hay sự xuất hiện của các biến thể bệnh mới.

Quy trình giải quyết bài toán được hiện thực hóa thông qua một kiến trúc ba giai đoạn chặt chẽ. Đầu tiên là việc xây dựng bộ trích xuất đặc dựa trên kiến trúc mô hình Attentive Batch Model (3) (4). Nhiệm vụ của thành phần này là chuyển đổi hình ảnh lá chè thô thành các véc-tơ đặc trưng giàu thông tin ngữ nghĩa. Yêu cầu đặt ra cho bộ trích xuất là khả năng nắm bắt đồng thời cả các đặc trưng tương đồng giữa tập support và query để có thể đưa ra các đặc trưng cục bộ chuẩn xác hơn khi phán đoán.

Tiếp theo là việc ứng dụng thuật toán Học dựa trên độ đo, trong đó đề tài tập trung nghiên cứu kiến trúc PN. Đây là mô hình được đánh giá cao về sự đơn giản nhưng hiệu quả trong việc tạo ra không gian biểu diễn phân lớp.

$$C_k = \frac{1}{|S_k|} \sum_{(x_i, y_i)} f_\theta(x_i) \tag{1}$$

Trong đó:

- C_k: Véc-tơ nguyên mẫu (Prototype) đại diện cho lớp bệnh k.
- S_k: Tập hỗ trợ (Support set) chứa các cặp dữ liệu mẫu x_i, y_i thuộc lớp k.
- $|S_k|$: Lực lượng của tập S_k (hay số lượng ảnh mẫu có trong tập hỗ trợ của lớp k).
- (x_i, y_i): Một cặp dữ liệu trong tập hỗ trợ, bao gồm ảnh đầu vào x_i và nhãn tương ứng y_i.
- f_θ: Hàm nhúng (embedding function) hay mạng nơ-ron trích xuất đặc trưng với tập tham số θ.
- $f_\theta(x_i)$: Véc-tơ đặc trưng (feature vector) của ảnh x_i sau khi đi qua mạng nơ-ron.

Về mặt toán học, giả sử tập hỗ trợ S_k chứa các hình ảnh mẫu của lớp bệnh k, mô hình sẽ tính toán một "nguyên mẫu" C_k (prototype) đại diện cho lớp đó. Nguyên mẫu này thực chất là trung bình (mean vector) của các véc-tơ đặc trưng sau khi đi qua mạng nhúng f_θ:

$$p(y = k|x) = \frac{\exp\left(-d\left(f_\theta(x), (C_k)\right)\right)}{\sum_{k'} \exp\left(-d(f_\theta(x), (C_{k'})\right)} \quad (2)$$

Trong đó:

- $p(y = k|x)$: Xác suất có điều kiện để ảnh truy vấn x có nhãn y là lớp k
- x: Ảnh truy vấn (Query image) cần được phân loại.
- f_θ: Véc-tơ đặc trưng của ảnh truy vấn x sau khi đi qua mạng nơ-ron.
- C_k: Véc-tơ nguyên mẫu của lớp k (đã tính ở công thức 1).
- $d(.,.)$: Hàm khoảng cách (Distance metric). Trong bài báo cáo này, văn bản ghi rõ đây là khoảng cách Euclidean.
- $d\left(f_\theta(x), (C_k)\right)$: Khoảng cách giữa đặc trưng của ảnh truy vấn và nguyên mẫu lớp k.
- exp (): Hàm mũ cơ số tự nhiên, dùng để chuyển đổi giá trị khoảng cách âm thành giá trị dương (dùng cho hàm Softmax).
- k': Biến chạy đại diện cho tất cả các lớp bệnh có trong tập dữ liệu (dùng ở mẫu số).
- $\sum_{k'}$: Tổng xác suất trên tất cả các lớp k' có thể (dùng để chuẩn hóa sao cho tổng xác suất bằng 1).

Với d là hàm khoảng cách Euclidean. Cơ chế này đảm bảo rằng trong không gian đặc trưng, điểm dữ liệu của ảnh truy vấn sẽ được kéo về phía nguyên mẫu của lớp bệnh đúng nhất và đẩy ra xa khỏi các nguyên mẫu khác.

Cuối cùng, để đảm bảo khả năng tổng quát hóa, quy trình huấn luyện sẽ được tổ chức theo chiến lược huấn luyện theo tập (Episodic Training). Dữ liệu sẽ được chia nhỏ thành hàng nghìn "tập" (episodes) mô phỏng lại đúng kịch bản kiểm thử N-way K-shot. Trong mỗi tập, mô hình buộc phải thực hiện nhiệm vụ phân loại dựa trên một tập hỗ trợ ngẫu nhiên rất nhỏ. Việc lặp lại quá trình này giúp mạng nơ-ron không "học thuộc lòng" các đặc điểm cố định của một bức ảnh cụ thể, mà học được quy luật chung để so sánh và nhận diện sự khác biệt. Đây chính là chìa khóa giúp hệ thống đạt được sự cân bằng giữa độ chính xác và khả năng thích nghi, giải quyết thành công bài toán nhận diện bệnh chè trong thực tế sản xuất thiếu hụt dữ liệu.

1.3. Kết luận chương

Chương 1 đã cung cấp một bức tranh tổng quan về bối cảnh thực tiễn cũng như các khía cạnh then chốt của bài toán nhận diện bệnh hại trên cây chè. Chúng ta đã bắt đầu bằng việc khám phá hệ sinh thái dịch bệnh phức tạp trên lá chè, từ những tổn thương hoại tử diện rộng của bệnh thán thư, chấm xám đến những đốm bệnh tinh vi như mắt cua hay đốm tảo, qua đó cho thấy tầm quan trọng của việc phân loại chính xác tình trạng sâu bệnh. Tiếp theo, các giới hạn của phương pháp học sâu truyền thống trong điều kiện dữ liệu hạn chế và tiềm năng của kỹ thuật Học sâu ít mẫu (Few-shot Learning) đã được phân tích, cung cấp bộ công cụ lý thuyết cần thiết như các mô hình dựa trên độ đo (Metric-based) để giải quyết bài toán "khan hiếm dữ liệu".

Trên cơ sở hiểu biết về đặc điểm sinh học của bệnh hại và các mô hình học máy tiên tiến, bài toán phân loại bệnh chè trong ngữ cảnh N-way K-shot đã được định nghĩa và mô tả chi tiết. Bài toán này làm nổi bật các thách thức kỹ thuật về sự tương đồng giữa các lớp bệnh (inter-class similarity) và biến thiên nội lớp (intra-class variation), mà còn phản ánh nhu cầu thực tiễn cấp bách trong việc tự động hóa quy trình giám sát nông nghiệp thông minh, giảm thiểu sự phụ thuộc vào chuyên gia và dữ liệu gán nhãn thủ công. Việc điểm qua các nghiên cứu liên quan cũng đã cho thấy xu hướng dịch chuyển tất yếu sang các phương pháp Meta-learning trong lĩnh vực thị giác máy tính nông nghiệp.

Với những kiến thức nền tảng về bệnh học cây chè, cơ sở lý thuyết Few-shot Learning và bài toán cụ thể đã được trang bị, chúng ta sẽ đi sâu hơn vào các phương pháp cụ thể nhằm xây dựng hệ thống nhận diện tối ưu. Chương tiếp theo sẽ đi sâu vào việc đề xuất mô hình chi tiết, phân tích kiến trúc mạng trích xuất đặc trưng và cơ chế so sánh độ đo, nhằm giải quyết việc phân loại bệnh chè với độ chính xác cao ngay cả khi chỉ có rất ít mẫu dữ liệu đầu vào.

CHƯƠNG 2 - PHÂN LOẠI BỆNH TRÊN LÁ CHÈ DỰA TRÊN HỌC SÂU ÍT MẪU

Nếu như Chương 1 đã hoàn thành nhiệm vụ khái quát hóa bối cảnh nông nghiệp thông minh và tính cấp thiết của việc kiểm soát bệnh hại trên cây chè, thì Chương 2 sẽ đóng vai trò là hạt nhân lý thuyết, tập trung giải quyết triệt để thách thức kỹ thuật lớn nhất mà đề tài đặt ra: Bài toán phân loại hình ảnh chính xác trong điều kiện dữ liệu cực kỳ hạn chế (Few-shot Learning). Đây là bước chuyển dịch quan trọng từ việc nhận diện vấn đề sang việc xây dựng phương pháp luận giải quyết vấn đề, nơi các mô hình học máy không chỉ đơn thuần là công cụ nhận dạng mà phải sở hữu khả năng tư duy và thích nghi linh hoạt tương tự như nhận thức của con người.

Trong thực tế triển khai các hệ thống AI tại các vùng chuyên canh chè, chúng ta thường xuyên đối mặt với nghịch lý về dữ liệu. Trong khi các mô hình mạng nơ-ron tích chập (CNN) truyền thống đòi hỏi hàng nghìn, thậm chí hàng triệu mẫu dữ liệu được gán nhãn thủ công để đạt độ chính xác tối ưu và tránh hiện tượng quá khớp (overfitting), thì điều kiện thực địa lại hiếm khi đáp ứng được yêu cầu khắt khe này. Chi phí thu thập dữ liệu cao, sự xuất hiện bất thường của các biến thể bệnh mới, và tính chất thời vụ của sâu bệnh khiến cho việc xây dựng các bộ dữ liệu lớn (Big Data) trở nên bất khả thi về mặt thời gian và kinh phí. Chính vì vậy, nhu cầu cấp thiết đặt ra là phải nghiên cứu và ứng dụng các cơ chế học tập tiên tiến như Meta-learning, cho phép mô hình có khả năng tổng quát hóa tri thức mạnh mẽ chỉ từ một vài ví dụ mẫu (Support Set) đơn lẻ.

Với mục tiêu đó, chương này sẽ được cấu trúc một cách chặt chẽ để đi sâu vào bản chất của bài toán học sâu ít mẫu. Nội dung chương sẽ bắt đầu bằng việc hình thức hóa bài toán dưới dạng cấu trúc N-way K-shot chuẩn mực, tạo cơ sở toán học cho các bước xử lý tiếp theo. Trọng tâm nghiên cứu sẽ được đặt vào việc phân tích cơ sở lý thuyết của nhóm thuật toán học dựa trên độ đo (Metric-based Learning), một hướng tiếp cận đầy tiềm năng giúp mô hình học được không gian đặc trưng tối ưu để so sánh sự tương đồng giữa các loại bệnh thay vì học thuộc lòng các đặc điểm cố định.

Trên nền tảng lý thuyết đó, chương sẽ trình bày chi tiết kiến trúc mô hình đề xuất, từ mạng xương sống (backbone) trích xuất đặc trưng đến các mô-đun tính toán khoảng cách và phân loại. Những phân tích chuyên sâu trong chương này sẽ đóng vai trò là nền móng vững chắc, định hướng cho toàn bộ quá trình hiện thực hóa hệ thống và đánh giá thực nghiệm sẽ được trình bày tại Chương 3 của đề án.

2.1. Tiền xử lý dữ liệu

Trong lĩnh vực thị giác máy tính, chất lượng và định dạng của dữ liệu đầu vào là yếu tố then chốt, quyết định trực tiếp đến hiệu suất hoạt động của các mô hình học sâu (deep learning). Đặc biệt trong nhiệm vụ phân loại bệnh trên lá chè, dữ liệu ảnh thô được thu thập từ môi trường thực tế thường mang nhiều thách thức như: Kích thước ảnh không đồng nhất (khác biệt về độ phân giải), chứa các yếu tố nhiễu nền (background noise) không liên quan, hay bị tác động bởi điều kiện ánh sáng đa dạng (ánh sáng mạnh, yếu, bóng râm...). Chính vì những vấn đề này, một quy trình tiền xử lý dữ liệu (data pre-processing) đã được xây dựng. Mục đích của quy trình này là chuẩn hóa và tinh chỉnh thông tin, đảm bảo dữ liệu đạt được định dạng tối ưu và loại bỏ nhiễu trước khi được cung cấp cho mô hình mạng nơron để tiến hành huấn luyện. Thực hiện tiền xử lý dữ liệu bao gồm 3 bước:

1. **Thay đổi kích thước (Resizing):** Hình Bước đầu tiên và cơ bản nhất trong quy trình tiền xử lý là chuẩn hóa không gian hình học của dữ liệu thông qua kỹ thuật thay đổi kích thước ảnh (Resizing). Mọi hình ảnh đầu vào thu thập từ thực địa, vốn có độ phân giải và tỷ lệ khung hình không đồng nhất, sẽ được quy hoạch về một kích thước chuẩn cố định là 224×224 pixel. Việc lựa chọn con số cụ thể này không phải là ngẫu nhiên mà là kết quả của sự cân nhắc kỹ lưỡng nhằm đảm bảo tính tương thích tuyệt đối với kiến trúc đầu vào của các mạng nơ-ron xương sống (Backbone) phổ biến và hiện đại nhất hiện nay như ResNet hay ConvNeXt. Hơn nữa, kích thước 224×224 đại diện cho một điểm cân bằng tối ưu về mặt kỹ thuật: nó giúp giảm thiểu khối lượng tính toán để mô hình vận hành nhẹ nhàng hơn, nhưng đồng thời vẫn đủ lớn để bảo toàn các chi tiết thị giác quan trọng của các vết bệnh, ngăn chặn việc mất mát thông tin hữu ích trong quá trình lấy mẫu lại

2. **Chuẩn hóa (Normalization):** Sau khi đồng bộ kích thước, kỹ thuật chuẩn hóa (Normalization) được áp dụng để điều chỉnh phân phối giá trị của các điểm ảnh (pixel), đưa chúng về một phạm vi số học tiêu chuẩn giúp mô hình xử lý hiệu quả hơn. Quá trình này thực hiện việc biến đổi dữ liệu dựa trên các tham số thống kê tham chiếu nổi tiếng từ tập dữ liệu khổng lồ ImageNet, cụ thể là sử dụng véc-tơ giá trị trung bình (mean) là $[0.485, 0.456, 0.406]$ và độ lệch chuẩn (standard deviation - std) là $[0.229, 0.224, 0.225]$ cho ba kênh màu. Mục tiêu tối thượng của bước chuẩn hóa này không chỉ dừng lại ở việc đưa dữ liệu về cùng một miền giá trị, mà quan trọng hơn là kiến tạo một bề mặt lỗi (loss surface) trơn tru hơn, từ đó thúc đẩy đáng kể tốc độ hội tụ của các thuật toán tối ưu hóa, điển hình như Gradient Descent, giúp quá trình huấn luyện diễn ra nhanh chóng và đạt độ ổn định cao.

3. **Tăng cường dữ liệu (Data Augmentation):** Để giải quyết triệt để thách thức cốt lõi về sự khan hiếm dữ liệu đầu vào – một đặc thù khó khăn của bài toán học ít mẫu, các kỹ thuật tăng cường dữ liệu (Data Augmentation) được triển khai mạnh mẽ ngay trong pha huấn luyện. Chiến lược này nhằm mục đích làm phong phú không gian mẫu và ngăn chặn hiện tượng quá khớp bằng cách tạo ra các biến thể dữ liệu nhân tạo thông qua các phép biến đổi ngẫu nhiên:

- *Lật ngẫu nhiên (Random Horizontal Flip):* Phép biến đổi này được sử dụng để mô phỏng tính đối xứng trục tự nhiên của cấu trúc lá chè, giúp mô hình học được các đặc trưng bất biến về hướng không gian.
- *Biến đổi màu sắc (Color Jitter):* Kỹ thuật này thực hiện việc thay đổi ngẫu nhiên các thông số thị giác như độ sáng, độ tương phản và độ bão hòa của ảnh. Ý nghĩa của việc này là giúp mô hình học được các đặc trưng hình học cốt lõi của vết bệnh mà không bị nhiễu hay thay đổi bởi các điều kiện ánh sáng môi trường khác nhau trên thực địa, qua đó tăng cường đáng kể tính chính xác và khả năng tổng quát hóa của hệ thống.

2.2. Học sâu ít mẫu phân loại bệnh trên lá chè

2.2.1. Few-shot learning là gì?

Học sâu ít mẫu FSL là một nhánh nghiên cứu của học sâu, tập trung vào việc xây dựng các mô hình có khả năng học các đặc trưng, tổng quát hóa tri thức từ một lượng dữ liệu rất hạn chế. Khác với các mô hình học sâu truyền thống yêu cầu hàng nghìn mẫu dữ liệu cho mỗi lớp (Supervised Learning), FSL hướng tới việc mô phỏng khả năng nhận thức của con người: có thể nhận diện một đối tượng mới chỉ sau khi quan sát một hoặc vài ví dụ mẫu.

Trong ngữ cảnh bài toán phân loại, FSL thường được định nghĩa dưới dạng bài toán $N\ way\ -\ K\ shot$, trong đó mô hình được cung cấp một tập hỗ trợ (Support Set) chứa N lớp bệnh, mỗi lớp chỉ có K hình ảnh (với K thường rất nhỏ, ví dụ $K = 1$, 5). Nhiệm vụ của mô hình là phân loại các hình ảnh trong tập truy vấn (Query Set) vào một trong N lớp đó dựa trên sự tương đồng với tập hỗ trợ.

2.2.2. Thách thức của Few-shot learning

Việc huấn luyện mô hình với dữ liệu cực ít đặt ra những thách thức kỹ thuật lớn:

1. **Hiện tượng quá khớp (Overfitting):** Khi số lượng tham số của mạng nơ-ron lớn hơn rất nhiều so với số lượng mẫu huấn luyện, mô hình có xu hướng "học thuộc lòng" các đặc điểm nhiễu của dữ liệu thay vì học các đặc trưng bản chất, dẫn đến khả năng dự đoán kém trên dữ liệu mới.

2. **Độ biến thiên nội lớp lớn (High Intra-class Variance):** Các bệnh trên lá chè có hình thái thay đổi mạnh tùy theo giai đoạn phát triển và điều kiện môi trường. Với chỉ vài mẫu đại diện (K-shot), rất khó để mô hình bao quát hết các biến thể của cùng một loại bệnh.

3. **Độ tương đồng giữa các lớp (Inter-class Similarity):** Một số loại bệnh khác nhau (ví dụ: bệnh cháy nâu và bệnh thán thư) có thể biểu hiện các triệu chứng hình ảnh rất giống nhau, gây khó khăn cho việc phân tách ranh giới giữa các lớp trong không gian đặc trưng.

2.2.3. Mô hình kiến trúc

Để giải quyết triệt để và hiệu quả bài toán học sâu ít mẫu (FSL) đầy thách thức, nghiên cứu này đã lựa chọn hướng tiếp cận dựa trên nền tảng học chuyển giao (Meta-learning), cụ thể là khai thác chiến lược học dựa trên độ đo (Metric-based Learning). Kiến trúc tổng thể của hệ thống được thiết kế dưới dạng một quy trình xử lý tuần tự và chặt chẽ, bao gồm ba thành phần chức năng chính tương tác mật thiết với nhau để tối ưu hóa khả năng phân loại:

1. **Mạng xương sống (Backbone Network):** Thành phần đầu tiên và nền tảng của kiến trúc là mạng nơ-ron tích chập (CNN), trong đó nghiên cứu sử dụng các kiến trúc tiên tiến như ResNet hoặc ConvNeXt nhằm mục đích trích xuất các đặc trưng thị giác một cách mạnh mẽ và ổn định nhất. Nhiệm vụ cốt lõi của hai loại backbone này là thực hiện quá trình chuyển đổi dữ liệu hình ảnh đầu vào: biến đổi từ không gian pixel thô sơ ban đầu thành các véc-tơ đặc trưng trừu tượng có tính phân biệt cao (discriminative features) nằm trong không gian nhúng (embedding space) nhiều chiều. Việc tạo ra một không gian nhúng chất lượng tại bước này là tiền đề quyết định hiệu suất của toàn bộ hệ thống.

2. **Cơ chế tập trung RBI (Relationship Batch Integration):** Điểm sáng tạo và khác biệt của mô hình nằm ở việc tích hợp một cơ chế tập trung chuyên biệt gọi là RBI. Cơ chế này được thiết kế để tăng cường độ chính xác tối đa trong việc nhận diện sự tương đồng giữa các mẫu bệnh. Cụ thể, đây là một cơ chế tăng cường tương quan giúp mô hình không chỉ phân tích từng ảnh riêng lẻ mà còn nâng cao khả năng nhận diện các đặc điểm giống và khác nhau tinh vi giữa các mẫu bệnh trong cùng một tập xử lý. Cơ chế này đặc biệt phát huy tác dụng trong bối cảnh dữ liệu cực kỳ ít ỏi và các đặc trưng thị giác giữa các loại bệnh có sự chồng lấn lớn, giúp mô hình phân tách rõ ràng hơn giữa các lớp bệnh dễ gây nhầm lẫn.

3. **Mô-đun học phép đo (Metric Module):** Thành phần cuối cùng hoàn thiện quy trình suy luận là mô-đun học phép đo. Đây có thể là một hàm toán học định nghĩa trước hoặc một mạng nơ-ron con, chịu trách nhiệm thực hiện việc tính toán độ tương đồng hoặc khoảng cách hình học giữa véc-tơ đặc trưng của ảnh truy vấn (Query) và

các véc-tơ đặc trưng đại diện của tập hỗ trợ (Support). Kết quả của phép đo lường này chính là cơ sở định lượng để hệ thống đưa ra quyết định phân loại cuối cùng một cách chính xác nhất.

Để vận hành kiến trúc này, quy trình huấn luyện được tổ chức một cách khoa học theo cơ chế tập (Episodic Training). Thay vì huấn luyện theo lô truyền thống, dữ liệu được chia nhỏ thành hàng loạt các tác vụ giả lập gọi là các "tập" (episodes), trong đó mỗi episode được thiết kế để mô phỏng chính xác cấu trúc N-way K-shot mà mô hình sẽ phải đối mặt trong giai đoạn kiểm thử. Chiến lược huấn luyện này giúp mô hình hình thành tư duy "học cách so sánh" linh hoạt giữa các đối tượng thay vì lối mòn "học cách nhận diện" cố định các lớp đã biết.

2.2.4. Backbone trích xuất Đặc trưng

Mạng xương sống đóng vai trò nền tảng trong việc tạo ra các biểu diễn đặc trưng chất lượng cao. Trong đồ án này, chúng tôi thực nghiệm với các kiến trúc CNN tiên tiến nhất hiện nay, được khởi tạo bằng trọng số đã huấn luyện trước (pre-trained) trên tập dữ liệu ImageNet:

- **ResNet (Residual Networks):** Sử dụng các kết nối RC để giải quyết vấn đề biến mất đạo hàm, cho phép huấn luyện các mạng rất sâu. Các phiên bản ResNet-50 và ResNet-101 được sử dụng để trích xuất các đặc trưng phân cấp từ thô (cạnh, góc) đến tinh (cấu trúc vết bệnh).
- **ConvNeXt:** Một kiến trúc CNN hiện đại được thiết kế lại dựa trên cảm hứng từ Vision Transformers (ViT). ConvNeXt sử dụng các kernel kích thước lớn, thay thế ReLU bằng GELU và chuẩn hóa LayerNorm, giúp mô hình nắm bắt ngữ cảnh toàn cục tốt hơn và đạt hiệu năng vượt trội trong các bài toán nhận diện hình ảnh.

Đầu ra của các mạng backbone này thường được đưa qua lớp gộp trung bình toàn cục (Global Average Pooling) để thu được véc-tơ đặc trưng 1 chiều trước khi đi vào mô-đun phân loại.

2.2.5. Mô-đun học phép đo (Metric Learning Module)

Trong hướng tiếp cận Metric-based, mục tiêu là học một không gian đặc trưng sao cho các ảnh cùng loại bệnh sẽ nằm gần nhau, và các ảnh khác loại sẽ nằm xa nhau. Đề án xem xét hai phương pháp tiêu biểu:

2.2.5.1. Phép đo cố định

PN dựa trên ý tưởng rằng mỗi lớp bệnh có thể được đại diện bởi một điểm duy nhất trong không gian đặc trưng, gọi là Nguyên mẫu (Prototype).

Nguyên mẫu C_k của lớp bệnh k được tính là trung bình cộng của các véc-tơ đặc trưng thuộc tập hỗ trợ S_k:

$$C_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\theta(x_i) \tag{3}$$

Trong đó:

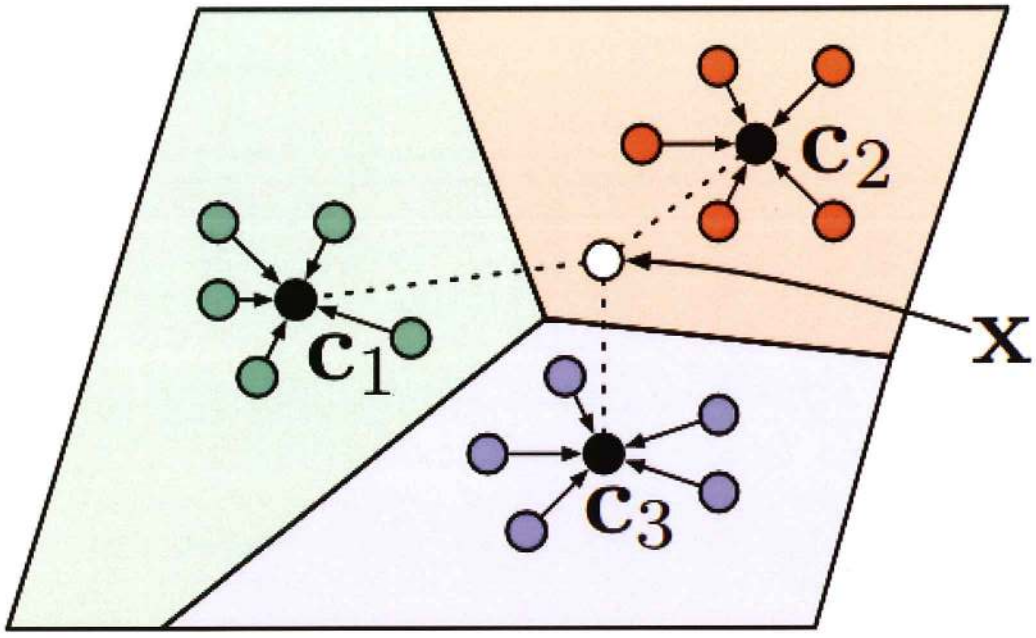
- C_k: Véc-tơ nguyên mẫu (Prototype) đại diện cho lớp bệnh k.
- S_k: Tập hỗ trợ (Support set) chứa các cặp dữ liệu mẫu x_i, y_i thuộc lớp k.
- $|S_k|$: Lực lượng của tập S_k (hay số lượng ảnh mẫu có trong tập hỗ trợ của lớp k).
- (x_i, y_i): Một cặp dữ liệu trong tập hỗ trợ, bao gồm ảnh đầu vào x_i và nhãn tương ứng y_i.
- f_θ: Hàm nhúng (embedding function) hay mạng nơ-ron trích xuất đặc trưng với tập tham số θ.
- $f_\theta(x_i)$: Véc-tơ đặc trưng (feature vector) của ảnh x_i sau khi đi qua mạng nơ-ron.

Việc phân loại cho một ảnh truy vấn x_q được tính bằng cách thực hiện đo khoảng cách từ $f_\theta(x_q)$ đến tất cả các nguyên mẫu C_k. Phép đo được sử dụng là khoảng cách Euclidean. Lớp dự đoán sẽ tương ứng với nguyên mẫu có khoảng cách gần nhất.

$$(d(z, c) = \|z - c\|^2 \tag{4}$$

Trong đó:

- $d(z, c)$: Hàm khoảng cách (distance function) giữa hai véc-tơ z và c. Giá trị này càng nhỏ thì hai véc-tơ càng giống nhau.
- z: Biến đại diện cho véc-tơ đặc trưng của ảnh truy vấn (Query).
- c: Biến đại diện cho véc-tơ nguyên mẫu (Prototype)
- $\|z - c\|$: Ký hiệu của Chuẩn L2 (L2-norm) hay còn gọi là độ dài Euclidean của hiệu véc-tơ



Hình 2: Prototypical networks trong few-shot

2.2.5.2. Phép đo có thể học (Relation Network)

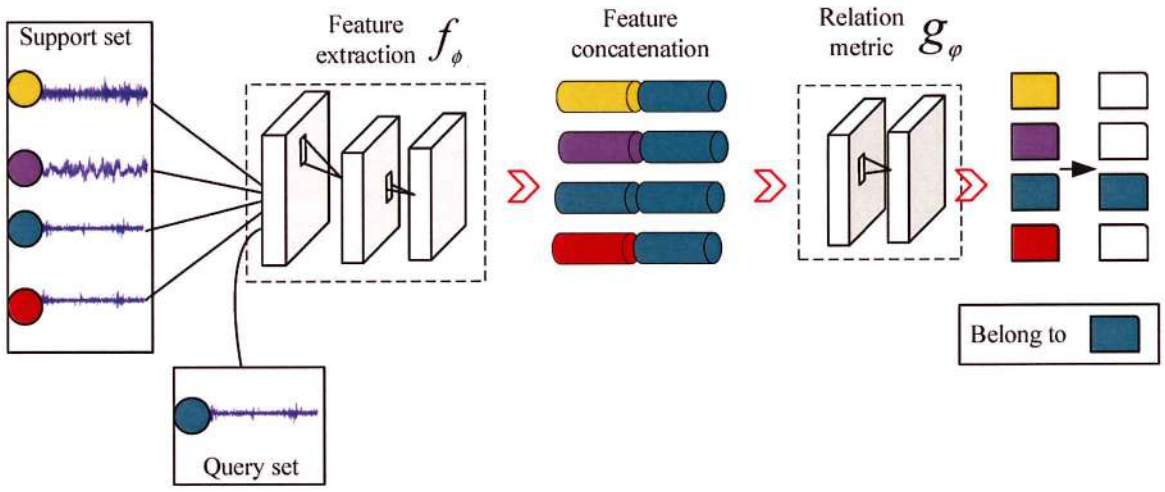
Khác với PN sử dụng khoảng cách Euclidean cố định, Relation Network (Mạng quan hệ) học một hàm đo độ tương đồng phi tuyến thông qua một mạng nơ-ron con (Relation Module C% Véc-tơ đặc trưng của ảnh truy vấn B :T và véc-tơ nguyên mẫu của lớp (hoặc từng mẫu trong tập hỗ trợ) được ghép nối (concatenate) lại với nhau:

N L C% l? K J?@B :P ;ÆB kT oÄp :w;

Trong đó:

- r_{ij} Kết quả đầu ra là một giá trị vô hướng (scalar) nằm trong khoảng >rAsGiá trị này biểu thị độ tương đồng giữa hai ảnh đầu vào Tvà T
- C: Đây là một mạng nơ-ron con (thường là các lớp Fully Connected hoặc Convolutional) với tập tham số .
- ? K J?:A=:;Phép ghép nối véc-tơ
- B Mạng nơ-ron trích xuất đặc trưng (backbone) với tham số

Đầu ra Nlà một giá trị scalar trong khoảng >rAs1biểu thị "điểm quan hệ" hay xác suất tương đồng giữa hai ảnh. Mạng C:được huấn luyện đồng thời với backbone để tự tìm ra cách so sánh tối ưu nhất cho đặc trưng bệnh chè.



Hình 3: Kiến trúc Relation network

2.2.6. Hàm mất mát

Để mô hình có thể "học" được cách trích xuất đặc trưng và so sánh độ tương đồng, chúng ta cần một thước đo định lượng sai số giữa kết quả dự đoán và nhãn thực tế. Trong khuôn khổ của bài toán phân loại N-way K-shot sử dụng mạng Nguyên mẫu (Prototypical Networks), hàm mục tiêu được thiết kế để tối đa hóa xác suất đúng của các mẫu trong tập truy vấn (Query Set).

Trước khi tính toán mất mát, xác suất để một ảnh truy vấn Tthuộc về lớp đúng Gđược xác định thông qua hàm Softmax dựa trên khoảng cách Euclidean âm.

Việc sử dụng dấu âm là rất quan trọng, bởi vì hàm Softmax ưu tiên các giá trị lớn, trong khi mục tiêu của chúng ta là tối thiểu hóa khoảng cách.

Dựa trên phân phối xác suất trên, hàm mất mát Cross-Entropy (hay cụ thể hơn trong trường hợp này là Negative Log-Likelihood) được áp dụng cho từng tập huấn luyện (episode). Mục tiêu của việc huấn luyện là tìm kiếm bộ tham số θ của mạng backbone và các trọng số của mô-đun RBI sao cho giá trị hàm mất mát L là nhỏ nhất.

Việc tối ưu hóa hàm mất mát (15) thực chất là thực hiện đồng thời hai tác động lực học trong không gian đặc trưng.

Lực kéo (Attraction): Tối thiểu hóa khoảng cách giữa ảnh truy vấn x_q và nguyên mẫu đúng C_{yq}. Điều này làm cho tử số trong công thức Softmax lớn hơn, giúp các điểm dữ liệu cùng một lớp bệnh co cụm lại gần nhau (Intra-class compactness).

Lực đẩy (Repulsion): Tối đa hóa khoảng cách giữa ảnh truy vấn x_q và các nguyên mẫu sai. Điều này nằm ở mẫu số của công thức Softmax, giúp đẩy các cụm dữ liệu của các bệnh khác nhau ra xa nhau (Inter-class separability).

Quá trình này được thực hiện thông qua thuật toán lan truyền ngược (Backpropagation). Gradient của hàm mất mát sẽ được tính toán và truyền ngược qua hàm khoảng cách, qua bộ tính nguyên mẫu, và cuối cùng cập nhật trọng số cho mạng Backbone (ResNet/ConvNeXt) và các lớp Attention trong mô-đun RBI. Kết quả là sau nhiều vòng lặp (epochs), mô hình sẽ học được cách nhúng các hình ảnh lá chè vào một không gian metric tối ưu, nơi các loại bệnh được phân tách rõ ràng.

2.2.7. Khung tích hợp quan hệ theo Batch (RBI Framework)

Đây là trái tim của kiến trúc đề xuất. Trong các mô hình Few-shot truyền thống, các ảnh trong cùng một tập (batch/episode) thường được xử lý độc lập cho đến khâu tính loss. Ngược lại, khung RBI chủ động khai thác mối tương quan ngữ cảnh giữa các ảnh ngay trong quá trình trích xuất đặc trưng thông qua hai mô-đun con chuyên biệt:

Mô-đun 1: Mã hóa vị trí quan hệ RPE

Trong kiến trúc của các mạng nơ-ron tích chập truyền thống, các mẫu dữ liệu trong cùng một lô (batch) thường được xử lý một cách độc lập, khiến cho thông tin về mối tương quan giữa các mẫu bị bỏ qua. Để khắc phục hạn chế này, mô-đun Mã hóa vị trí Quan hệ (RPE) đóng vai trò là thành phần khởi tạo ngữ cảnh, chịu trách nhiệm xây dựng một bản đồ quan hệ toàn cục giữa các mẫu hình ảnh trong cùng một tập huấn luyện. Đầu ra của mô-đun này là ma trận tương đồng hình ảnh (Image Similarity Matrix - ISM), ký hiệu là $S \in R^{B \times B}$ (với B là kích thước batch), nơi mỗi phần tử S_{ij} đại diện cho mức độ tương đồng sâu sắc về mặt đặc trưng giữa ảnh thứ i và ảnh thứ j.

Điểm đột phá trong thiết kế của RPE nằm ở việc lựa chọn hàm đo độ đo. Thay vì sử dụng các khoảng cách hình học thông thường như Euclidean hay Cosine Similarity – vốn chỉ đo lường khoảng cách tuyến tính trong không gian véc-tơ, nghiên cứu này ứng dụng chỉ số tỷ số tín hiệu trên nhiễu Đỉnh (Peak Signal-to-Noise Ratio - PSNR). Trong xử lý ảnh truyền thống, PSNR thường được dùng để đánh giá chất lượng ảnh sau nén. Tuy nhiên, trong ngữ cảnh của học sâu ít mẫu cho bệnh chè, việc áp dụng PSNR mang lại một lợi thế đặc biệt: khả năng khuếch đại sự khác biệt đối với các đặc trưng có độ tương đồng cao. Do bản chất của hàm Logarit trong công thức PSNR, mô hình trở nên nhạy bén hơn với các sai biệt nhỏ giữa các đặc trưng bệnh học có hình thái gần giống nhau (ví dụ: phân biệt giữa bệnh thán thư giai đoạn đầu và bệnh cháy nâu), giúp tách biệt các lớp bệnh "khó" hiệu quả hơn so với các phép đo tuyến tính.

Công thức toán học để tính toán phần tử S_{ij} (độ tương đồng quan hệ giữa véc-tơ đặc trưng z_i và z_j) được thể hiện như sau:

$$S_{ij} = 20 \cdot \log_{10}\left(\frac{\max(|z_i|, |z_j| + \epsilon}{\sqrt{MSE(z_i, z_j) + \epsilon}}\right) \quad (6)$$

Trong đó:

- z_i, z_j Î $R^{d_{model}}$: véc-tơ đặc trưng của ảnh i và j .

- $MSE(z_i, z_j)$: sai số bình phương trung bình giữa hai véc-tơ.
- $max(|z_i|, |z_j|)$: giá trị biên độ lớn nhất, đại diện cho cường độ tín hiệu.
- ϵ : hằng số nhỏ để đảm bảo tính ổn định số học.

Kết quả S_{ij} thu được không chỉ là một con số vô hướng mà nó đóng vai trò như một mã hóa vị trí tương đối. Trong các bước xử lý tiếp theo của mô-đun RRA, ma trận S này sẽ được cộng trực tiếp vào các véc-tơ đặc trưng (Key và Value), cho phép mô hình "nhìn thấy" vị trí của một mẫu bệnh trong mối tương quan với toàn bộ các mẫu bệnh khác đang có mặt trong batch, từ đó học được cách phân loại dựa trên ngữ cảnh thay vì chỉ dựa trên đặc điểm riêng lẻ.

Mô-đun 2: Chú ý quan hệ thặng dư (Residual Relationship Attention - RRA)

Sau khi ma trận tương đồng hình ảnh (ISM) được thiết lập bởi mô-đun RPE, hệ thống tiếp tục đưa dữ liệu vào trái tim của khung RBI: mô-đun chú ý quan hệ thặng dư (RRA). Nếu RPE đóng vai trò là "bản đồ" chỉ ra vị trí tương đối giữa các ảnh, thì RRA chính là "bộ não" xử lý bản đồ đó để tinh chỉnh đặc trưng. Mục tiêu tối thượng của RRA là khắc phục điểm yếu cốt tử của các mạng CNN truyền thống: việc xử lý từng ảnh một cách cô lập. Thay vào đó, RRA thực hiện cơ chế lan truyền thông tin ngữ cảnh, cho phép đặc trưng của một lá chè bị bệnh "mượn" thông tin từ các lá chè khác trong cùng một tập huấn luyện (batch) để làm rõ chính mình.

Quá trình hoạt động của RRA được chia thành ba giai đoạn xử lý chuyên sâu: (1) Biến đổi không gian và tích hợp vị trí, (2) Tính toán trọng số chú ý tương quan, và (3) Tổng hợp đặc trưng thích nghi qua cơ chế cổng.

Đầu tiên, véc-tơ đặc trưng đầu vào x (kích thước $B \times d_{model}$) được chiếu qua ba lớp tuyến tính riêng biệt để tạo ra các ma trận Truy vấn Q, Khóa K và Giá trị K. Tuy nhiên, để mô hình hóa mối quan hệ "mọi - mọi" (all-to-all) giữa các ảnh trong batch, RRA không giữ nguyên kích thước này mà áp dụng kỹ thuật nhân bản tensor đặc biệt:

- Sao chép Dọc (Vertical Duplication - VD): Được áp dụng cho ma trận Khóa K. Tensor được mở rộng theo chiều dọc.

- Sao chép Ngang (Horizontal Duplication - HD): Được áp dụng cho ma trận Truy vấn Q và Giá trị V. Tensor được mở rộng theo chiều ngang.

Tại đây, một cơ chế đổi mới sáng tạo được áp dụng tích hợp vị trí quan hệ. Thay vì chỉ sử dụng đặc trưng hình ảnh đơn thuần, RRA cộng trực tiếp thông tin từ ma trận tương đồng S (ISM) vào trong K và K. Điều này biến đổi bản chất của dữ liệu: giá trị của Key và Value giờ đây không chỉ chứa "nội dung" của bức ảnh mà còn chứa cả "tọa độ quan hệ" của nó so với các ảnh khác:

$$K'_j = K_j + S_{ij} \quad (7)$$

$$V'_j = V_j + S_{ij} \quad (8)$$

Trong đó:

- K_j, V_j: Véc-tơ Key (Khóa) và Value (Giá trị) gốc được trích xuất từ đặc trưng của mẫu thứ j.
- S_{ij}: Véc-tơ biểu diễn mối quan hệ vị trí hoặc không gian (spatial relation) giữa vị trí i (đang xét) và vị trí j.
- K'_j, V'_j: Véc-tơ Key và Value đã được hiệu chỉnh

S_{ij} đóng vai trò như một dạng Positional Embedding mềm, giúp mô hình hiểu được rằng: "Mẫu bệnh A này rất giống mẫu B, nhưng lại xa lạ với mẫu C", từ đó định hướng sự tập trung của cơ chế chú ý.

Sau khi đã có các véc-tơ K' và V' mang thông tin quan hệ, RRA tiến hành tính toán mức độ quan trọng (trọng số chú ý) của từng mẫu trong batch đối với mẫu đang xét. Trọng số này quyết định xem mô hình nên "tin tưởng" và "học hỏi" bao nhiêu từ các mẫu lân cận. Công thức tính toán dựa trên tích vô hướng (Dot-product) đã được chuẩn hóa:

$$A_{ij} = softmax\left(\frac{Q_i \cdot (K'_j)^T}{\sqrt{d_k}}\right) \quad (9)$$

Trong đó:

- A_{ij}: Trọng số chú ý (Attention weight). Giá trị này càng lớn nghĩa là đặc trưng tại j càng quan trọng đối với i.

- Q_i: Véc-tơ Query (Truy vấn) tại vị trí i.
- K_j': Véc-tơ Key đã điều chỉnh.
- $(K_j')^T$: Phép chuyển vị (Transpose) của véc-tơ Key.
- $\sqrt{d_k}$: Hệ số tỷ lệ (scaling factor), với d_k là kích thước chiều của véc-tơ Key.

Kết quả đầu ra Z^{attn} là tổng hợp có trọng số của toàn bộ thông tin trong batch:

$$Z^{attn} = \sum_j A_{ij} \cdot V_j' \tag{10}$$

Trong đó:

- Z^{attn}: Véc-tơ đặc trưng đầu ra sau khi áp dụng cơ chế chú ý.
- A_{ij}: Trọng số chú ý.
- V_j': Véc-tơ Value đã điều chỉnh.

Cuối cùng, điểm khác biệt lớn nhất của RRA so với các cơ chế chú ý thông thường (như trong Transformer) nằm ở bước cơ chế cổng (Gating Mechanism). Không phải lúc nào thông tin từ các ảnh khác cũng hữu ích; đôi khi chúng có thể gây nhiễu nếu tập dữ liệu chứa nhiều mẫu ngoại lai. Do đó, RRA sử dụng một tham số cổng β để tự động điều tiết luồng thông tin. Mô hình tính toán một véc-tơ quyết định β dựa trên sự so sánh giữa đặc trưng mới Z^{attn} và đặc trưng gốc (Z):

$$\beta = \sigma(W_b[Z^{attn} || W_S(Z) || Z^{attn} - W_S(Z)]) \tag{11}$$

Trong đó:

- β: Véc-tơ tham số cổng, có giá trị nằm trong khoảng $(0, 1)$.
- σ: Hàm kích hoạt Sigmoid.
- W_b: Ma trận trọng số (learnable weights) của mạng nơ-ron tuyến tính.
- Z^{attn}: Đặc trưng đầu ra từ lớp chú ý.
- $W_S(Z)$: Đặc trưng gốc Z đã được biến đổi tuyến tính qua trọng số W_S để khớp kích thước.

Cuối cùng, đặc trưng đầu ra Z^{out} được tổng hợp thông qua phép cộng có trọng số động:

$$Z^{out} = BatchNorm1d\big((1 - \beta) \odot Z^{attn} + \beta \odot W_S(Z)\big) \tag{12}$$

Trong đó:

- Z^{out}: Véc-tơ đặc trưng đầu ra cuối cùng.
- $BatchNorm1d(.)$: Lớp chuẩn hóa theo batch.
- Z^{attn}: Đặc trưng đầu ra từ lớp chú ý.

Cổng β hoạt động như một "van điều áp" thông minh. Nếu $\beta \approx 1$, mô hình quyết định giữ nguyên đặc trưng gốc (bảo toàn thông tin thị giác). Nếu $\beta \approx 0$, mô hình ưu tiên sử dụng đặc trưng quan hệ (tận dụng ngữ cảnh). Sự linh hoạt này giúp mô hình Attentive Batch Model vừa có khả năng phân biệt chi tiết sắc bén của CNN, vừa có khả năng tổng quát hóa cao của Meta-learning, giải quyết triệt để bài toán biến thiên nội lớp lớn trong dữ liệu bệnh chè.

2.3. Kết luận chương

Chương 2 đã hoàn tất xuất sắc việc xây dựng nền tảng lý luận và kiến trúc kỹ thuật, qua đó xác lập một phương pháp tiếp cận toàn diện và chiến lược nhằm giải quyết thách thức cốt lõi của đề tài: Phân loại bệnh trên lá chè trong bối cảnh dữ liệu ít mẫu (Few-shot Learning - FSL).

Trọng tâm của giải pháp kỹ thuật nằm ở việc ứng dụng kiến trúc học dựa trên Độ đo (Metric-based Learning), tiêu biểu là mạng Nguyên mẫu (Prototypical Networks). Theo đó, triết lý cốt lõi là tạo ra một không gian nhúng đặc trưng (embedding space) có tính phân biệt cao, nơi mà các mẫu của cùng một lớp bệnh, dù là mẫu huấn luyện hay mẫu mới, đều được ánh xạ về gần nhau. Việc khởi tạo mạng bằng trọng số đã được pre-trained là một biện pháp không thể thiếu, giúp mô hình nhanh chóng tiếp thu các đặc trưng thị giác cấp thấp và cấp trung, từ đó tối ưu hóa quá trình học tập meta trên dữ liệu chè giới hạn. Song song đó, các kỹ thuật tiền xử lý ảnh và tăng cường dữ liệu (Data Augmentation) nghiêm ngặt được áp dụng để tăng cường tính mạnh mẽ và khả năng tổng quát hóa của mô hình trước độ biến thiên tự nhiên của ảnh chụp lá chè.

Điểm đột phá của phương pháp này chính là cơ chế nguyên mẫu (Prototype). Trong quá trình huấn luyện meta, mỗi lớp bệnh được đại diện bằng một trung tâm

được tính toán dựa trên trung bình các véc-tơ đặc trưng của tập hỗ trợ. Khi một ảnh truy vấn mới query được đưa vào, thay vì thông qua một lớp phân loại Softmax truyền thống, mô hình thực hiện phân loại bằng cách tính toán khoảng cách giữa véc-tơ đặc trưng của ảnh truy vấn đó với tất cả các Prototype đã được học. Lớp có khoảng cách gần nhất sẽ là dự đoán cuối cùng. Phương pháp này trực tiếp huấn luyện mô hình để "học cách so sánh" và "học cách nhóm" thay vì "học thuộc lòng ranh giới phân loại" cứng nhắc, từ đó khắc phục hiệu quả vấn đề quá khớp thường gặp khi dữ liệu huấn luyện quá ít. Với nền tảng lý thuyết vững chắc và thiết kế kiến trúc chi tiết đã được trình bày. Ngoài ra cơ chế tập trung theo mối hệ theo Batch giúp mô hình đã có khả năng tìm ra những lớp có độ tương đồng giống nhau cao trong một intra-class. Chương 2 đã trang bị đầy đủ công cụ cần thiết để chuyển sang Chương 3. Chúng ta đã sẵn sàng thực nghiệm hóa mô hình, định lượng hóa hiệu suất và khẳng định tính ưu việt của phương pháp học sâu ít mẫu trong việc giải quyết bài toán chẩn đoán bệnh trên lá chè.

CHƯƠNG 3: THỰC NGHIỆM VÀ ĐÁNH GIÁ

Nếu như Chương 2 đóng vai trò thiết lập nền tảng lý luận vững chắc thông qua việc đề xuất kiến trúc mô hình "Attentive Batch Model" kết hợp giữa mạng nơ-ron tích chập và cơ chế học dựa trên độ đo, thì Chương 3 chính là giai đoạn hiện thực hóa và kiểm chứng tính đúng đắn của các giả thuyết khoa học đó thông qua các bằng chứng thực nghiệm định lượng. Mục tiêu cốt lõi của chương này không chỉ dừng lại ở việc vận hành mô hình, mà là chứng minh khả năng vượt trội của giải pháp đề xuất trong việc giải quyết bài toán "nút thắt" về dữ liệu của ngành nông nghiệp: làm thế nào để nhận diện chính xác các bệnh hại trên lá chè khi chỉ có một lượng mẫu tham chiếu cực kỳ hạn chế.

Để đạt được mục tiêu đó, nội dung chương được cấu trúc một cách hệ thống, bắt đầu bằng quy trình xây dựng và chuẩn hóa tập dữ liệu chuyên biệt cho cây chè. Tại đây, các kịch bản kiểm thử tiêu chuẩn $N - way\ K - shot$ (5-way 1-shot, 5-way 5-shot) sẽ được thiết lập nghiêm ngặt nhằm mô phỏng chính xác những thách thức thực tế về sự khan hiếm dữ liệu và độ tương đồng cao giữa các lớp bệnh. Tiếp theo, sự minh bạch trong nghiên cứu khoa học sẽ được đảm bảo thông qua việc trình bày chi tiết các thông số cấu hình môi trường huấn luyện, các chiến lược tối ưu hóa như Episodic Training, cũng như hệ thống các độ đo hiệu năng được sử dụng để đánh giá.

Trọng tâm phân tích của chương sẽ đi sâu vào việc so sánh hiệu quả giữa kiến trúc đề xuất (tích hợp khung RBI) với các mạng xương sống (Backbone) học sâu truyền thống như ResNet và ConvNeXt. Những phân tích này nhằm làm rõ vai trò của cơ chế chú ý quan hệ trong việc cải thiện độ chính xác phân loại và khả năng thích nghi nhanh chóng của hệ thống. Cuối cùng, để khẳng định tính khả thi và giá trị ứng dụng thực tiễn, chương sẽ giới thiệu quá trình xây dựng một hệ thống demo Client-Server hoàn chỉnh, chuyển hóa các kết quả nghiên cứu từ phòng thí nghiệm thành một công cụ hỗ trợ chẩn đoán trực quan, sẵn sàng phục vụ cho bài toán giám sát sức khỏe cây trồng trong nền nông nghiệp thông minh.

3.1. Thiết lập và cấu hình

Các thực nghiệm được triển khai trên môi trường Ubuntu 22.04, CUDA 12.8, Python 3.10 cùng các thư viện phổ biến như PyTorch và TorchVision. Toàn bộ thí nghiệm được tối ưu cho kiến trúc Few-shot Learning theo cơ chế huấn luyện theo tập (Episodic Training), do đó các tham số được sử dụng nhằm đảm bảo tính ổn định, khả năng tổng quát hóa và tốc độ hội tụ của mô hình.

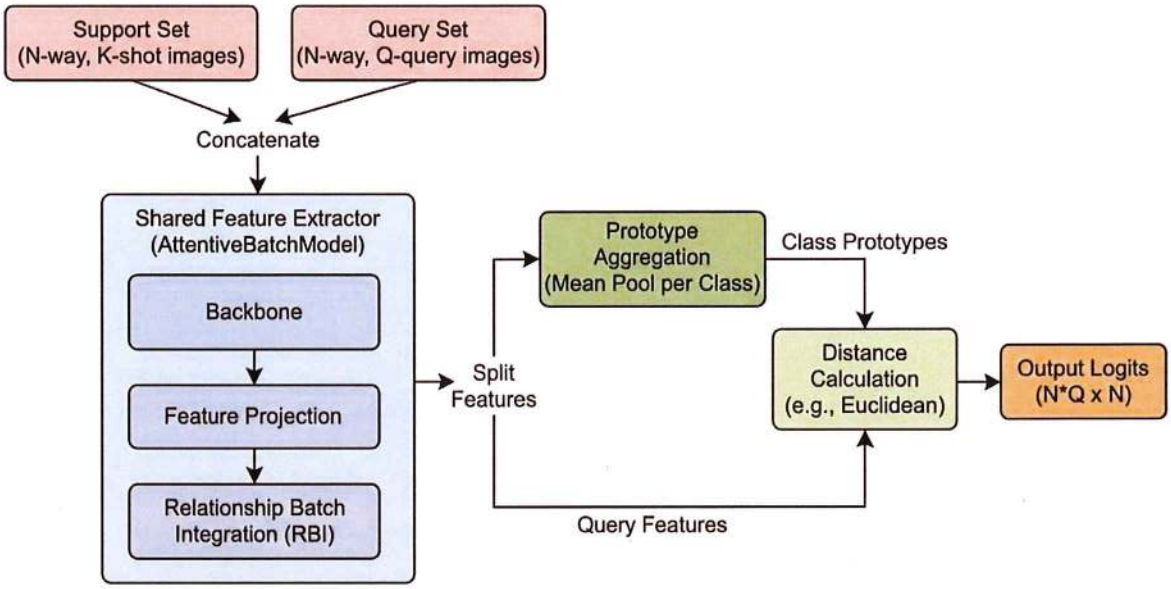
Bảng 1: Thống kê tham số được sử dụng

Tên tham số	Giá trị	Giải thích
backbone	resnet101	Mạng xương sống trích xuất đặc trưng (sâu hơn ResNet50).
d_model	2048	Số chiều của không gian nhúng đặc trưng.
dropout	0.3	Tỷ lệ dropout để giảm quá khớp (khá cao).
ways	1/5	Số lớp bệnh trong mỗi tập (1/5-way).
shots	5	Số ảnh mẫu cho mỗi lớp (5-shot).
queries	10	Số ảnh truy vấn cho mỗi lớp trong một tập.
episodes	100	Số lượng tập huấn luyện trong mỗi epoch.
epochs	60	Tổng số vòng lặp huấn luyện.
lr (Learning Rate)	0.0003	Tốc độ học khởi điểm.
weight_decay	1e-4	Hệ số suy giảm trọng số (L2 Regularization).
grad_clip	10.0	Ngưỡng cắt gradient để ổn định huấn luyện.
warmup_epochs	10	Số epoch khởi động để ổn định backbone.
scheduler_type	plateau	Giảm LR khi độ chính xác bão hòa.
lr_patience	5	Số epoch chờ trước khi giảm LR nếu không cải thiện.
lr_factor	0.5	Hệ số giảm LR.

lr_min	1e-6	Tốc độ học tối thiểu.
early_stopping	20	Dừng sớm nếu không cải thiện sau 20 epochs.

3.2. Xây dựng, huấn luyện mô hình

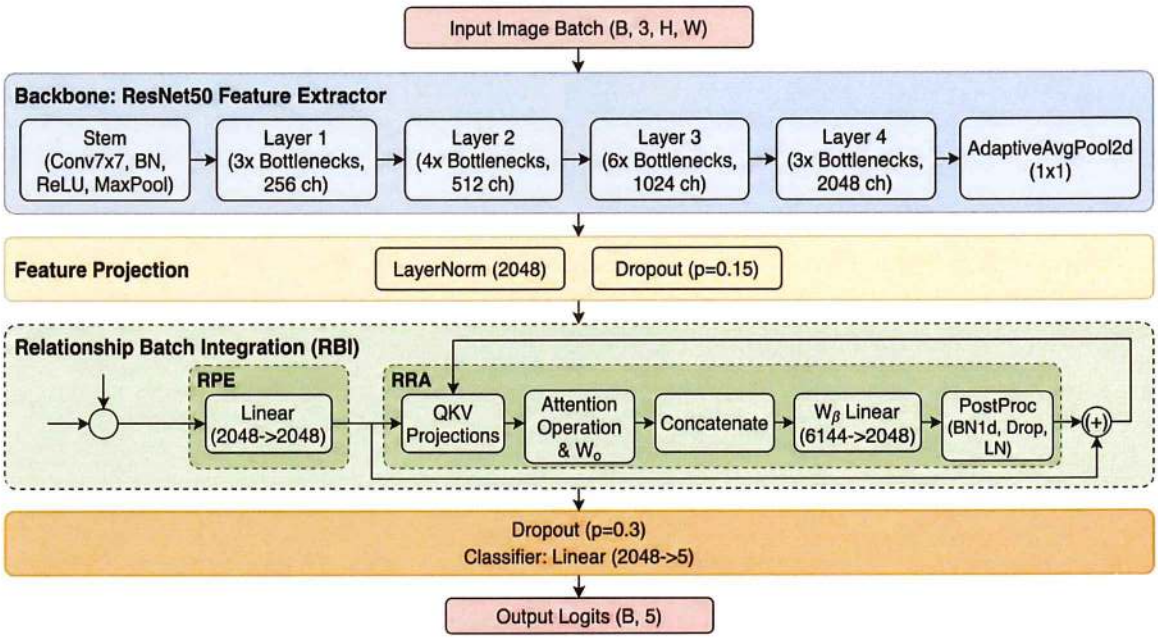
Để giải quyết bài toán phân loại bệnh trên lá chè với thách thức dữ liệu hạn chế, đề tài xây dựng một hệ thống FSL dựa trên kiến trúc Attentive Batch Model. Mô hình này là sự kết hợp chặt chẽ giữa khả năng trích xuất đặc trưng mạnh mẽ của các mạng CNN hiện đại và cơ chế học dựa trên độ đo (Metric-based Learning) được tăng cường bởi RBI.



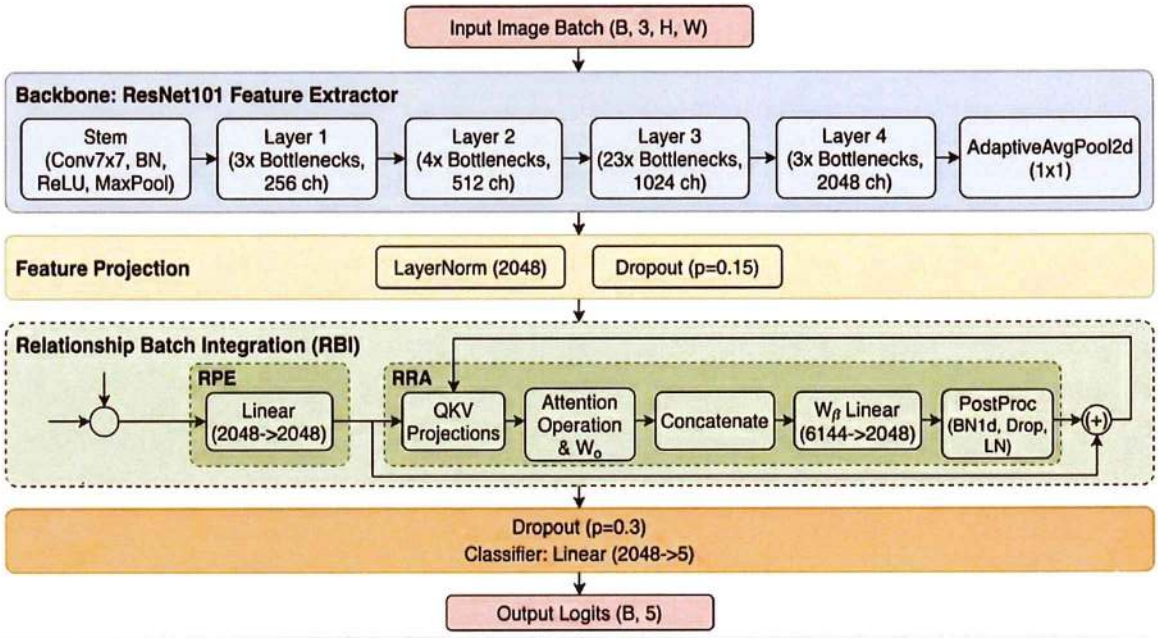
Hình 4: Sơ đồ kiến trúc

3.2.1. Kiến trúc mô hình đề xuất

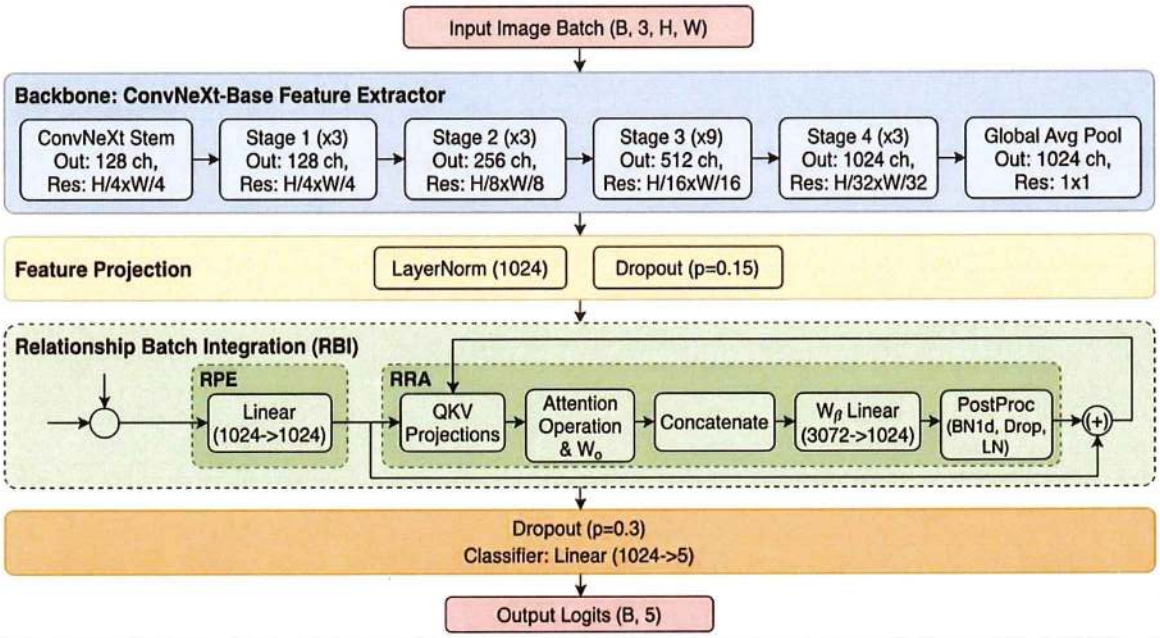
Kiến trúc tổng thể của hệ thống được thiết kế theo luồng xử lý tuần tự gồm ba khối chức năng chính: (1) Mạng xương sống trích xuất đặc trưng, (2) Khung tăng cường quan hệ RBI, và (3) Bộ phân loại dựa trên nguyên mẫu.



Hình 5: Cơ chế tập trung và backbone Resnet50



Hình 6: Cơ chế tập trung và backbone Resnet101



Hình 7: Cơ chế tập trung vào backbon ConvNeXt-Base

a. Mạng xương sống và Chiếu đặc trưng (Backbone Network & Feature Projection)

Thành phần đầu tiên tiếp nhận dữ liệu hình ảnh là mạng xương sống (Backbone). Thay vì khởi tạo ngẫu nhiên, hệ thống sử dụng các kiến trúc CNN tiên tiến như **ResNet-50**, **ResNet-101**, hoặc **ConvNeXt-Base**. Điểm mấu chốt là các mạng này đều được nạp bộ trọng số đã huấn luyện trước (pre-trained) trên tập dữ liệu ImageNet quy mô lớn. Chiến lược TL này giúp mô hình "kế thừa" các bộ lọc thị giác phong phú-từ việc nhận diện cạnh, góc, màu sắc ở các lớp nông đến các họa tiết phức tạp như vết đốm, vết hoại tử ở các lớp sâu-qua đó giảm thiểu đáng kể thời gian hội tụ và nguy cơ quá khớp khi huấn luyện trên tập dữ liệu bệnh chè nhỏ.

Đầu ra của mạng backbone là các bản đồ đặc trưng (feature maps) có kích thước lớn. Để chuyển đổi chúng thành dạng véc-tơ gọn nhẹ phù hợp cho việc tính toán khoảng cách, hệ thống áp dụng lớp gộp trung bình thích nghi (Adaptive Average Pooling) để thu về kích thước. Tiếp theo, véc-tơ này được đưa qua khối chiếu đặc trưng (Feature Projection). Khối này bao gồm một lớp tuyến tính (Linear) để ánh xạ đặc trưng vào không gian nhúng có số chiều cố định (được thiết lập là 2048 chiều

trong thực nghiệm), kết hợp với lớp chuẩn hóa (LayerNorm) và Dropout để ổn định phân phối dữ liệu và tăng cường khả năng tổng quát hóa.

b. Bộ phân loại Mạng Nguyên mẫu (Prototypical Classifier)

Cơ chế phân loại dựa trên nguyên lý của PN. Trong không gian đặc trưng metric đã được tối ưu hóa bởi RBI, mỗi lớp bệnh được đại diện bởi một điểm duy nhất gọi là "Nguyên mẫu" (Prototype) C_k.

Trong mỗi tập huấn luyện (episode) N-way K-shot, nguyên mẫu của lớp k được tính là trung bình cộng của các véc-tơ nhúng thuộc tập hỗ trợ S_k:

$$C_k = \frac{1}{|S_k|} \sum_{(x_i, y_i)} f_\theta(x_i) \tag{13}$$

Trong đó:

- C_k: Véc-tơ nguyên mẫu (Prototype) đại diện cho lớp bệnh k.
- S_k: Tập hỗ trợ (Support set) chứa các cặp dữ liệu mẫu x_i, y_i thuộc lớp k.
- $|S_k|$: Lực lượng của tập S_k (hay số lượng ảnh mẫu có trong tập hỗ trợ của lớp k).
- (x_i, y_i): Một cặp dữ liệu trong tập hỗ trợ, bao gồm ảnh đầu vào x_i và nhãn tương ứng y_i.
- f_θ: Hàm nhúng (embedding function) hay mạng nơ-ron trích xuất đặc trưng với tập tham số θ.

Đối với một mẫu truy vấn (Query sample) x_q, xác suất để nó thuộc về lớp bệnh k được tính toán dựa trên khoảng cách đến nguyên mẫu C_k. Nghiên cứu sử dụng Khoảng cách Euclidean âm làm hàm đo độ tương đồng:

$$p(y = k|x_q) = \frac{\exp\left(-\left\|Z_q^{out} - C_k\right\|^2\right)}{\sum_{k'} \exp\left(-\left\|Z_q^{out} - C_{k'}\right\|^2\right)} \tag{14}$$

Trong đó:

- $p(y = k|x_q)$: Xác suất dự đoán ảnh truy vấn x_q thuộc về lớp bệnh k.
- Z_q^{out}: Véc-tơ đặc trưng đầu ra của ảnh truy vấn x_q.
- C_k: Nguyên mẫu của lớp k.

3.2.2. Quy trình huấn luyện và tối ưu hóa

a. Chiến lược Huấn luyện theo Tập (Episodic Training Strategy)

Khác biệt hoàn toàn với phương pháp huấn luyện theo lô (minibatch) trong các mô hình học sâu giám sát truyền thống—nơi mục tiêu là tối ưu hóa trọng số để nhận diện cố định một tập hợp các lớp đã biết, phương pháp Meta-learning tiếp cận bài toán theo hướng "học cách thích nghi". Để đạt được điều này, quy trình huấn luyện được tổ chức lại theo cơ chế Episodic Training (Huấn luyện theo Tập). Mục đích cốt lõi của chiến lược này là mô phỏng chính xác kịch bản kiểm thử Few-shot ngay trong giai đoạn huấn luyện, đảm bảo sự đồng nhất về phân phối tác vụ giữa hai pha (train-test consistency).

Thay vì đưa toàn bộ dữ liệu vào mạng một cách ngẫu nhiên, dữ liệu được chia nhỏ và đóng gói thành hàng nghìn "tập" (episodes). Mỗi episode được xem như một tác vụ phân loại giả lập (pseudo-task) độc lập. Cấu trúc của một episode được thiết kế ngẫu nhiên và tuân thủ nghiêm ngặt nguyên tắc $N - way\ K - shot$ như sau:

Mỗi episode được cấu trúc ngẫu nhiên như sau:

- Lựa chọn không gian nhãn (Task Sampling): Hệ thống sẽ chọn ngẫu nhiên N lớp bệnh (Ways) từ tập dữ liệu huấn luyện khả dụng. Việc thay đổi ngẫu nhiên các lớp N này qua từng episode giúp mô hình không bị "học vẹt" đặc điểm của một loại bệnh cụ thể, mà buộc phải học cách phân biệt sự khác nhau giữa các lớp bất kỳ.
- Xây dựng tập Hỗ trợ (Support Set Construction): Với mỗi lớp trong N lớp đã chọn, hệ thống lấy ngẫu nhiên K hình ảnh để tạo thành Tập hỗ trợ S.
  - Tổng số ảnh trong tập này là $N \times K$
  - Trong kiến trúc Prototypical Networks, tập này đóng vai trò là dữ liệu tham chiếu để mạng nơ-ron tính toán ra các nguyên mẫu (Prototypes) C_k — các điểm đại diện trung tâm cho từng lớp bệnh trong không gian đặc trưng.

Xây dựng tập truy vấn (Query Set Construction):

Từ các lớp đã chọn, hệ thống tiếp tục lấy ngẫu nhiên Q hình ảnh khác (không trùng với tập hỗ trợ) để tạo thành Tập truy vấn Q.

- Tập này đóng vai trò như dữ liệu kiểm thử trong tác vụ giả lập đó.

- Nhiệm vụ của mô hình là phân loại chính xác các ảnh trong tập truy vấn dựa trên các nguyên mẫu được tạo ra từ tập hỗ trợ.

Cơ chế học tập: Trong mỗi episode, mô hình buộc phải giải quyết bài toán: "Làm thế nào để phân loại đúng N lớp bệnh này chỉ dựa trên $N \times K$ ví dụ mẫu vừa nhìn thấy?". Sai số (Loss) được tính toán dựa trên kết quả dự đoán của tập truy vấn và được dùng để cập nhật trọng số mạng thông qua lan truyền ngược.

Thông qua việc lặp lại quá trình này hàng nghìn lần với các tổ hợp lớp bệnh và mẫu ảnh khác nhau, mô hình dần hình thành được một không gian metric tối ưu. Tại đó, mô hình học được tri thức tổng quát về cách so sánh độ tương đồng và trích xuất đặc trưng, cho phép nó dễ dàng thích nghi với các loại bệnh mới chưa từng gặp (unseen classes) trong giai đoạn kiểm thử thực tế.

b. Hàm mất mát và kỹ thuật tối ưu

Hàm mất mát Cross-Entropy được tính toán trên tập truy vấn của mỗi episode để hướng dẫn quá trình cập nhật trọng số:

$$L = -\sum_{(x_q, y_q) \in Q} \log p(y = y_q | x_q) \quad (15)$$

Trong đó:

- L: Giá trị hàm mất mát.
- Q: Tập dữ liệu truy vấn được sử dụng trong episode hiện tại.
- x_q: Ảnh truy vấn.
- y_q: Nhãn thực tế.
- (x_q, y_q) :Một cặp mẫu dữ liệu trong tập Q , bao gồm ảnh đầu vào x_q và nhãn thực tế y_q.

Để đảm bảo quá trình hội tụ ổn định cho kiến trúc mạng phức tạp (Backbone + RBI), các kỹ thuật tối ưu hóa sau được áp dụng đồng bộ:

- **Optimizer:** Sử dụng thuật toán **AdamW** với tốc độ học (learning rate) khởi tạo là và hệ số suy giảm trọng số (weight decay) để kiểm soát độ phức tạp mô hình.
- **LR Scheduler:** Áp dụng chiến lược **Warmup** (khởi động nóng) trong 10 epochs đầu tiên để ổn định trọng số của mạng backbone đã pre-trained. Sau giai đoạn

này, tốc độ học được giảm dần theo hàm Cosine (**CosineAnnealingLR**) hoặc giảm khi độ chính xác bão hòa (**ReduceLROnPlateau**).

• **Gradient Clipping:** Ngưỡng cắt gradient được thiết lập ở mức 10.0. Kỹ thuật này cực kỳ quan trọng khi huấn luyện các mạng có cơ chế chú ý (như RRA) để ngăn chặn hiện tượng bùng nổ gradient, đảm bảo quá trình cập nhật trọng số diễn ra trơn tru.

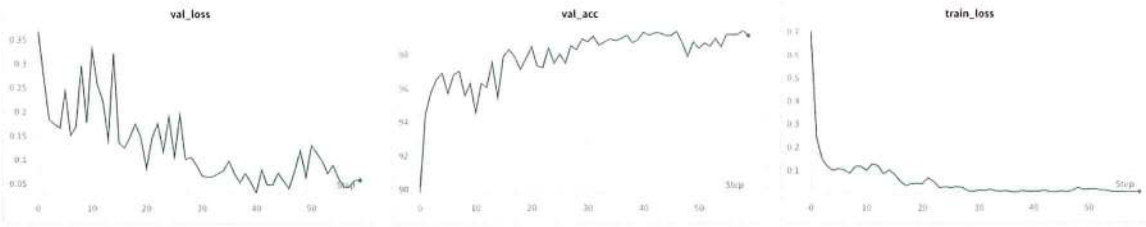
• **Data Augmentation:** Do dữ liệu bệnh chè rất ít, việc tăng cường dữ liệu là bắt buộc. Trong pha huấn luyện, các phép biến đổi như RandomHorizontalFlip và ColorJitter (thay đổi độ sáng, tương phản, bão hòa) được áp dụng ngẫu nhiên để mô phỏng các điều kiện ánh sáng và góc chụp khác nhau trên thực địa.

3.3. Kết quả mô hình

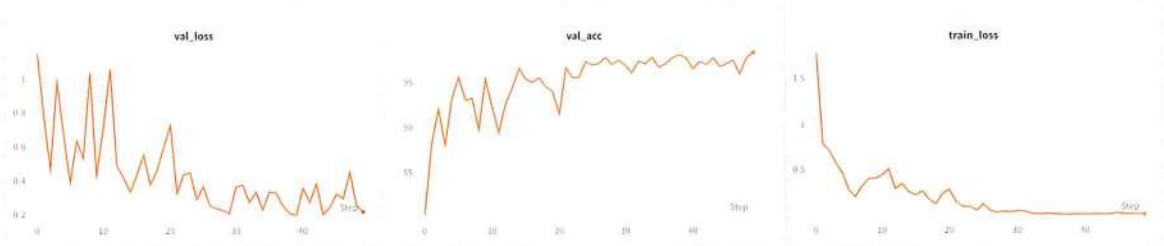
Phần này trình bày các kết quả định lượng thu được từ quá trình thực nghiệm mô hình đề xuất trên tập dữ liệu bệnh lá chè. Để đánh giá toàn diện hiệu năng của kiến trúc mạng, chúng tôi tiến hành so sánh trên ba mạng xương sống khác nhau: ResNet-50, ResNet-101 và ConvNeXt-Base. Các thí nghiệm được thiết lập theo hai kịch bản Few-shot tiêu chuẩn: 5-way 1-shot và 5-way 5-shot. Độ chính xác được báo cáo là giá trị trung bình trên 200 tập kiểm thử ngẫu nhiên với độ tin cậy lớn hơn 96%.



Hình 8: Giá trị hàm mất mát, độ chính xác trong quá trình huấn luyện với mạng xương sống Resnet101



Hình 9: Giá trị hàm mất mát, độ chính xác trong quá trình huấn luyện với mạng xương sống Resnet50



Hình 10: Giá trị hàm mất mát, độ chính xác trong quá trình huấn luyện với mạng xương sống ConvNeXt-Base

Kết quả thực nghiệm cho thấy sự tác động rõ rệt của độ sâu và kiến trúc mạng backbone đối với khả năng trích xuất đặc trưng trong bài toán ít mẫu. Bảng 3.2 dưới đây tóm tắt độ chính xác của mô hình Attentive Batch Model (tích hợp RBI) với các backbone khác nhau.

Bảng 2: Độ chính xác phân loại (%) trên tập kiểm thử bệnh lá chè

STT	Mạng xương sống (Backbone)	5-way 1-shot	5-way 5-shot
1	ResNet-50	96.73%	97.43%
2	ResNet-101	97.38%	97.93%
3	ConvNeXt-Base	**97.42%**	**98.27%**

Mô hình sử dụng ConvNeXt-Base cho kết quả vượt trội nhất ở cả hai kịch bản 1-shot và 5-shot. Điều này có thể được lý giải bởi kiến trúc hiện đại của ConvNeXt, vốn kết hợp các ưu điểm của Vision Transformers (như kernel kích thước lớn, các lớp chuẩn hóa LayerNorm) với tính hiệu quả của CNN. Khả năng trích xuất các đặc trưng ngữ nghĩa mạnh mẽ của ConvNeXt giúp mô hình tạo ra các véc-tơ nhúng chất lượng cao hơn, làm tăng khoảng cách giữa các lớp bệnh khác nhau trong không gian metric. Trong khi đó, ResNet-101 cho kết quả tốt hơn ResNet-50 nhờ độ sâu lớn hơn,

cho phép học được các đặc trưng trừu tượng phức tạp hơn, tuy nhiên mức độ cải thiện không đột phá bằng việc chuyển sang kiến trúc ConvNeXt.

Khi tăng số lượng mẫu hỗ trợ từ 1-shot lên 5-shot, độ chính xác của tất cả các mô hình đều tăng đáng kể (trung bình tăng khoảng 1%).

- Ở kịch bản 1-shot, mô hình phải xác định nguyên mẫu (prototype) chỉ dựa trên một hình ảnh duy nhất, do đó rất nhạy cảm với nhiễu hoặc các biến thể bất thường (outliers) của vết bệnh.
- Ở kịch bản 5-shot, nguyên mẫu được tính trung bình từ 5 véc-tơ đặc trưng, giúp loại bỏ nhiễu và tạo ra một đại diện ổn định hơn cho lớp bệnh. Điều này khẳng định rằng mặc dù mô hình Few-shot có thể hoạt động với 1 ảnh, nhưng việc cung cấp thêm một vài mẫu nhỏ (5 ảnh) sẽ mang lại độ chính xác cao.

Bảng 3: Độ chính xác phân loại (%) giữa các kiến trúc

STT	Mạng xương sống (Backbone)	5-way 1-shot	5-way 5-shot
1	ResNet-50	94.33%	95.28%
2	ResNet-101	95.41%	96.81%
3	**Kiến trúc đề xuất (Backbone + RBI)**	**97.42%**	**98.27%**

Mặc dù bảng trên so sánh các backbone, nhưng mức hiệu năng cao đạt được 96% là minh chứng gián tiếp cho hiệu quả của RBI. Nếu chỉ sử dụng PN thuần túy, các đặc trưng thường bị rời rạc. Việc ConvNeXt-Base kết hợp với RBI đạt kết quả cao nhất cho thấy cơ chế chú ý quan hệ (RRA) đã thành công trong việc tinh chỉnh đặc trưng dựa trên ngữ cảnh, giúp mô hình phân biệt tốt các loại bệnh có hình thái tương đồng như Chấm xám và Thán thư ngay cả khi dữ liệu huấn luyện rất ít.

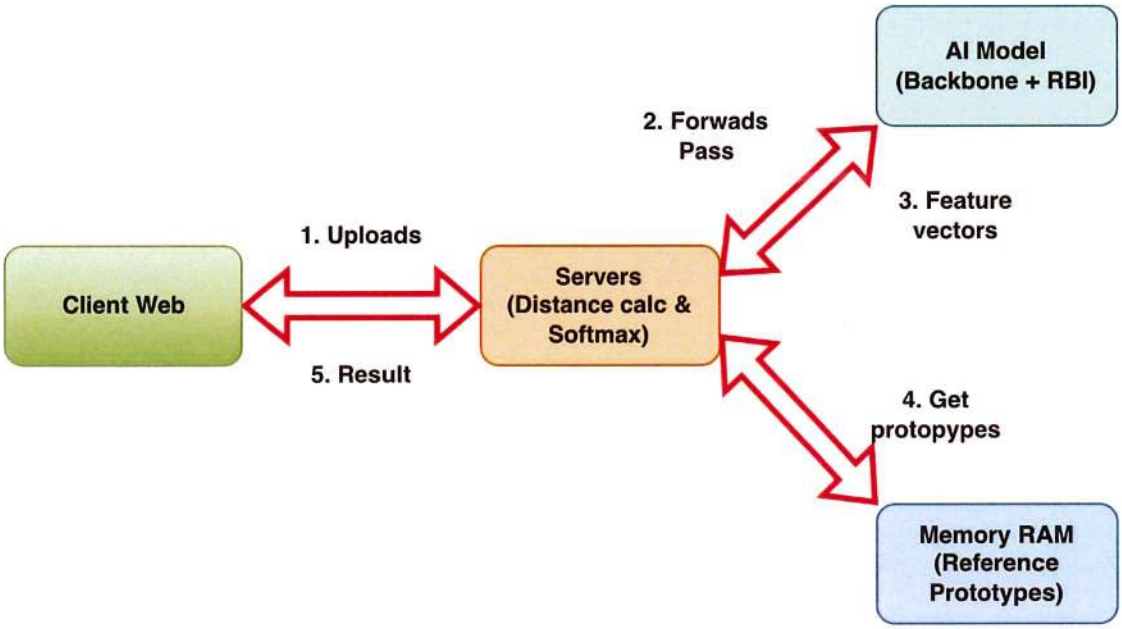
3.4. Xây dựng ứng dụng application - servers

Để hiện thực hóa các kết quả nghiên cứu lý thuyết và chuyển đổi mô hình học sâu ít mẫu FSL thành một giải pháp công nghệ hữu hình, đề tài đã tiến hành xây dựng một hệ thống phần mềm hỗ trợ chẩn đoán bệnh trên lá chè. Mục tiêu của phân hệ này không chỉ dừng lại ở việc kiểm thử tính đúng đắn của thuật toán, mà còn hướng tới

việc cung cấp một công cụ trực quan, thân thiện, cho phép người dùng cuối (nông dân, chuyên gia nông nghiệp) tương tác trực tiếp với trí tuệ nhân tạo, minh chứng cho tính khả thi của việc ứng dụng công nghệ cao vào thực tiễn sản xuất.

Hệ thống được thiết kế dựa trên mô hình Khách - Chủ (Client-Server) theo kiến trúc đơn khối (Monolithic Architecture) tinh gọn. Việc lựa chọn kiến trúc này nhằm tối ưu hóa tốc độ triển khai, giảm thiểu độ trễ giao tiếp giữa các thành phần và đơn giản hóa quá trình bảo trì trong giai đoạn thử nghiệm.

Tổng quan kiến trúc (Hình 8) bao gồm hai phân hệ chính tương tác chặt chẽ với nhau thông qua giao thức HTTP/RESTful API



Hình 11: Sơ đồ kiến trúc application - servers

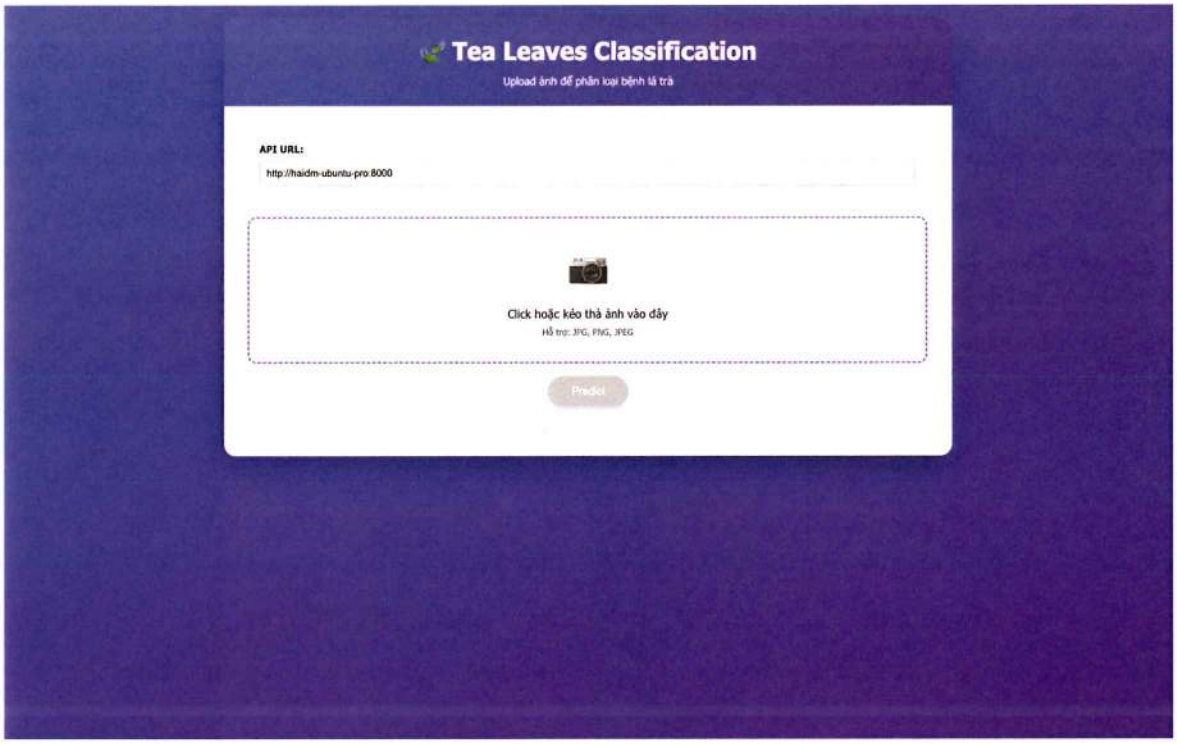
Về mặt kỹ thuật, hệ thống được thiết kế dựa trên mô hình Khách - Chủ (Client-Server) kinh điển, vận hành theo kiến trúc đơn khối (Monolithic Architecture) tinh gọn. Quyết định lựa chọn kiến trúc đơn khối thay vì các kiến trúc phân tán phức tạp (như Microservices) trong giai đoạn này là sự cân nhắc chiến lược nhằm tối ưu hóa tốc độ triển khai, giảm thiểu tối đa độ trễ giao tiếp giữa các thành phần nội bộ và đơn giản hóa quá trình bảo trì, sửa lỗi trong giai đoạn thử nghiệm. Tổng quan kiến trúc hệ thống được chia thành hai phân hệ chính tương tác chặt chẽ với nhau thông qua giao thức chuẩn HTTP/RESTful API, đảm bảo sự tách biệt rõ ràng giữa giao diện hiển thị và năng lực xử lý tính toán.

Phân hệ Máy chủ (Backend Server - Inference Engine):

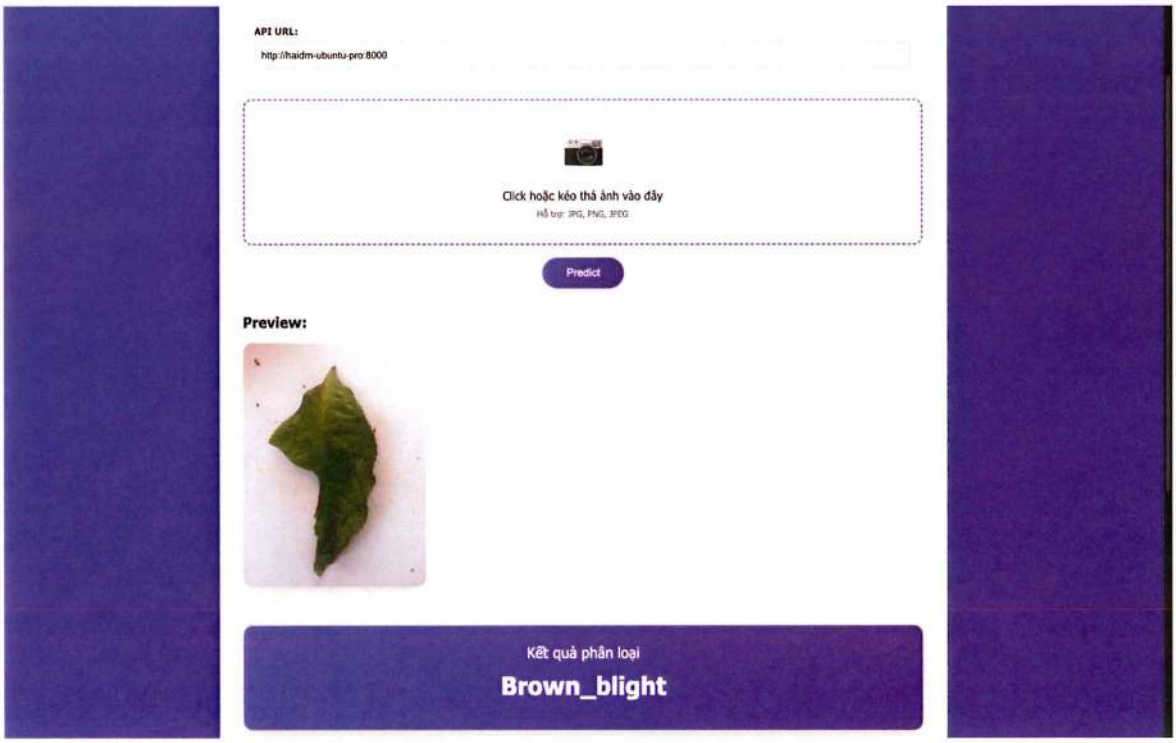
- Đây là trung tâm xử lý của toàn bộ hệ thống, được xây dựng trên nền tảng ngôn ngữ Python - ngôn ngữ tiêu chuẩn trong lĩnh vực khoa học dữ liệu.
- Máy chủ chịu trách nhiệm quản lý vòng đời của mô hình AI (Model Lifecycle Management). Ngay khi khởi động, hệ thống sẽ tải bộ trọng số đã được huấn luyện tối ưu (của Backbone ResNet/ConvNeXt và các mô-đun RBI) vào bộ nhớ (RAM/VRAM).
- Điểm đột phá về mặt kỹ thuật trong thiết kế backend của đề tài nằm ở việc cài đặt cơ chế "Bộ nhớ Nguyên mẫu" (Prototype Memory). Đây là khu vực lưu trữ chuyên biệt dành cho các véc-tơ đặc trưng đại diện C_k của từng loại bệnh, vốn được tính toán trước từ tập dữ liệu hỗ trợ (Support Set). Cơ chế này giải quyết triệt để bài toán hiệu năng trong mô hình Few-shot Learning: thay vì phải tính toán lại nguyên mẫu mỗi khi có yêu cầu mới – một tác vụ tốn kém tài nguyên, hệ thống chỉ cần truy xuất từ bộ nhớ đệm. Điều này giúp loại bỏ các bước tính toán dư thừa, giảm đáng kể thời gian phản hồi (latency) và mang lại trải nghiệm mượt mà cho người dùng ngay cả khi lượng truy cập tăng cao.

Phân hệ Người dùng (Client Web):

- Giao diện người dùng được phát triển dựa trên các công nghệ web tiêu chuẩn (JavaScript), đảm bảo khả năng tương thích đa nền tảng (Cross-platform), hoạt động tốt trên cả máy tính cá nhân và thiết bị di động.
- Client đóng vai trò là cổng thu thập dữ liệu, cho phép người dùng tải lên hình ảnh lá chè từ thực địa và hiển thị kết quả chẩn đoán một cách trực quan.



Hình 12: Giao diện web



Hình 13: Giao diện web phản hồi kết quả

3.5. Kết chương

Chương 3 đã hoàn thành vai trò then chốt trong cấu trúc tổng thể của đồ án tốt nghiệp, đó là hiện thực hóa các mô hình lý thuyết trừu tượng thành các kết quả định lượng cụ thể và sản phẩm công nghệ hữu hình. Thông qua một quy trình thực nghiệm được thiết kế nghiêm ngặt, từ khâu chuẩn hóa dữ liệu theo cấu trúc N-way K-shot đến việc tinh chỉnh các siêu tham số cho chiến lược huấn luyện theo tập (Episodic Training), chương này đã cung cấp những bằng chứng thực nghiệm đanh thép khẳng định tính đúng đắn của giả thuyết nghiên cứu.

Điểm nhấn quan trọng nhất mà thực nghiệm đã làm sáng tỏ chính là vai trò cốt lõi của Meta-learning (Học cách học) trong việc giải quyết bài toán dữ liệu khan hiếm. Kết quả thực nghiệm cho thấy, khác với các mô hình học sâu truyền thống (vốn cố gắng tối ưu hóa để nhận diện cố định một tập hợp lớp bệnh cụ thể), mô hình đề xuất đã thành công trong việc học được "tri thức tổng quát về cách so sánh". Bằng cách tổ chức quá trình huấn luyện thành hàng ngàn tập nhỏ (episodes) mô phỏng kịch bản kiểm thử, hệ thống đã được rèn luyện để nhanh chóng thích nghi với các tác vụ phân loại mới.

Cụ thể, sự kết hợp giữa mạng xương sống hiện đại ConvNeXt-Base và khung Tích hợp Quan hệ (RBI) trong mô hình Attentive Batch Model đã minh chứng cho sức mạnh của tư duy Meta-learning:

- Chuyển dịch từ "học nhận diện" sang "học so sánh": Thay vì ghi nhớ đặc điểm cứng nhắc của bệnh thán thư hay bệnh chấm xám, mô hình đã học được cách xây dựng một không gian metric linh hoạt. Tại đó, khoảng cách giữa các mẫu cùng loại được thu nhỏ tối đa và khoảng cách giữa các mẫu khác loại được nới rộng triệt để.

- Hiệu quả trên dữ liệu cực ít FSL: Với kết quả đạt ngưỡng trên 98% ở kịch bản 5-shot và xấp xỉ 97.5% ở kịch bản 1-shot, hệ thống chứng minh rằng nó không cần hàng nghìn ảnh để "hiểu" một loại bệnh mới. Chỉ cần 1 đến 5 ảnh mẫu (Support set), cơ chế Meta-learning đã cho phép mô hình nội suy và đưa ra phán đoán chính

xác cho các ảnh truy vấn (Query set), giải quyết triệt để vấn đề biến thiên nội lớp và tương đồng giữa các lớp.

Đi sâu vào kiến trúc, các cơ chế bổ trợ như mã hóa Vị trí Quan hệ (RPE) và chú ý quan hệ thặng dư (RRA) chính là những công cụ hiện thực hóa tư tưởng Meta-learning này. Chúng giúp mô hình khai thác thông tin ngữ cảnh toàn cục của tập dữ liệu (batch), cho phép "mượn" thông tin từ các mẫu lân cận để làm rõ các đặc trưng bệnh học mờ nhạt. Đây là điều mà các phương pháp học giám sát truyền thống không thể thực hiện được.

Bên cạnh những đóng góp sâu sắc về mặt thuật toán, chương này còn đánh dấu bước tiến lớn về ứng dụng thực tiễn thông qua hệ thống Client-Server. Việc triển khai thành công mô hình Meta-learning phức tạp lên máy chủ suy luận với tốc độ phản hồi thời gian thực đã thu hẹp khoảng cách giữa nghiên cứu hàn lâm và triển khai thực tế. Giao diện trực quan của ứng dụng giúp người dùng cuối (nông dân, cán bộ kỹ thuật) dễ dàng tiếp cận công nghệ AI tiên tiến mà không cần kiến thức chuyên sâu, mở ra triển vọng tích hợp vào các hệ thống nông nghiệp thông minh IoT trong tương lai.

Chương 3 không chỉ dừng lại ở các con số khô khan mà đã khẳng định sự thành công trong việc áp dụng tư duy Meta-learning để giải quyết bài toán thực tế. Những kết quả này xác nhận rằng hướng tiếp cận Học dựa trên độ đo (Metric-based Learning) là lời giải tối ưu cho bài toán ít mẫu, tạo tiền đề vững chắc để tác giả tổng hợp các đóng góp và đề xuất hướng phát triển trong phần Kết luận chung của đồ án.

KẾT LUẬN

Như vậy, đề án đã giải quyết bài toàn về Phân loại bệnh trên lá chè trong điều kiện dữ liệu hạn chế (Few-shot Learning). Bài toán này có nhiều thách thức so với cách nhận diện truyền thống do sự đa dạng về hình thái bệnh, sự tương đồng giữa các lớp bệnh khác nhau và đặc biệt là sự khan hiếm dữ liệu mẫu chuẩn hóa trong thực tế sản xuất.

Để giải quyết những khó khăn đó, đề án đã đề xuất một kiến trúc mô hình học sâu ít mẫu tiên tiến, kết hợp giữa mạng nơ-ron tích chập (CNN) trích xuất đặc trưng và khung RBI. Đóng góp cốt lõi của nghiên cứu nằm ở việc tích hợp sáng tạo cơ chế RPE và RRA vào quá trình học. Cách tiếp cận này cho phép mô hình không chỉ học các đặc trưng thị giác đơn lẻ mà còn khai thác mối tương quan ngữ cảnh giữa các mẫu trong cùng một tập huấn luyện, từ đó tinh chỉnh và làm nổi bật các đặc trưng phân biệt quan trọng giữa các loại bệnh dễ nhầm lẫn như Thán thư, Chấm xám hay Đốm mắt cua.

Quá trình thực nghiệm trên tập dữ liệu bệnh lá chè tự thu thập và chuẩn hóa đã chứng minh hiệu quả vượt trội của phương pháp đề xuất. Kết quả cho thấy mô hình Attentive Batch Model sử dụng backbone ConvNeXt-Base đạt độ chính xác cao nhất (trên 97% cho kịch bản 5-shot), vượt qua các kiến trúc cơ sở như ResNet-50 và ResNet-101. Đặc biệt, mô hình thể hiện khả năng thích nghi nhanh chóng và độ ổn định cao ngay cả khi chỉ có rất ít mẫu dữ liệu đầu vào (1-shot, 5-shot), khẳng định tính ưu việt của thuật toán học dựa trên độ đo (Metric-based) được tăng cường bởi cơ chế chú ý quan hệ. Bên cạnh đó, việc xây dựng thành công ứng dụng demo với kiến trúc Client-Server đơn giản cũng đã minh chứng tính khả thi trong việc triển khai mô hình vào thực tế, đáp ứng yêu cầu về tốc độ phản hồi và độ chính xác cho người dùng cuối.

Tuy nhiên, nhìn nhận một cách khách quan, dự án vẫn còn tồn tại một số hạn chế nhất định. Hiệu quả phân loại của mô hình hiện tại vẫn phụ thuộc một phần vào lượng kiến thức tiên nghiệm (inductive bias) được huấn luyện từ trước của Mạng

xương sống để đạt kết quả cao, và khả năng tổng quát hóa trên các biến thể bệnh hoàn toàn lạ lẫm vẫn cần được cải thiện.

Hiệu quả của mô hình vẫn phụ thuộc một phần vào lượng kiến thức được huấn luyện từ trước của Mạng xương sống để đạt kết quả cao. Hướng nghiên cứu trong tương lai sẽ tập trung vào việc mở rộng tập dữ liệu với nhiều biến thể bệnh và điều kiện môi trường hơn, đồng thời khám phá các kỹ thuật học tăng cường (Reinforcement Learning) hoặc học bán giám sát (Semi-supervised Learning) để tận dụng nguồn dữ liệu chưa gán nhãn khổng lồ. Ngoài ra, việc tối ưu hóa mô hình để chạy trên các thiết bị di động (Mobile/Edge Devices) cũng là một hướng đi đầy hứa hẹn, giúp đưa công nghệ này đến gần hơn với người nông dân trên chính cánh đồng của họ.

TÀI LIỆU THAM KHẢO

[1]. *Learning to compare: Relation network for few-shot learning.* F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales. s.l. : IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.

[2] Jake Snell, Kevin Swersky, Richard S. Zemel. *Prototypical Networks for Few-shot Learning.* 2017.

[3] *Deep Learning for Plant Disease Detection: Deep Learning for Plant.* Khalid, Munaf & Karan, Oguz. 2023, International Journal of Mathematics.

[4] *Enhancing fine-grained image classification through attentive batch training.* Le, Duy & Bui, Bao & Tran, Anh & Tran, Cong & Pham, Cuong. 2025, ICT Express.

[5] *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks.* Levine, Chelsea Finn and Pieter Abbeel and Sergey. PMLR, s.l. : PMLR, 2017, Proceedings of the 34th International Conference on Machine Learning, Vol. 70, pp. 1126--1135.

[6] *How to train your MAML.* Antoniou, Antreas and Edwards, Harrison and Storkey, Amos. 2018.

[7] *A Closer Look at Few-shot Classification.* Huang, Wei-Yu Chen and Yen-Cheng Liu and Zsolt Kira and Yu-Chiang Frank Wang and Jia-Bin. arXiv, 2020.

[8] *Dynamic Few-Shot Visual Learning Without Forgetting.* Gidaris, Spyros and Komodakis, Nikos. 2018, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4367-4375.

[9] *Plant Disease Detection Using Deep Learning.* Hamed, Bahaa and Hussein, Mahmoud and Mousa, Afaf. 2023, International Journal of Intelligent Systems and Applications, pp. 38-50.

[10] *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.* Tan, Mingxing and Le, Quoc. 2019.

[11] *ImageNet Classification with Deep Convolutional Neural Networks.* Hinton, Geoffrey and Krizhevsky, Alex and Sutskever, Ilya and Rachmad, Yoesoep. 2012, Advances in Neural Information Processing Systems, pp. 1097-1105.

[12] *Deep Residual Learning for Image Recognition.* He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian. 2016, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.

[13] *Densely Connected Convolutional Networks.* Huang, Gao and Liu, Zhuang and Van Der Maaten, Laurens and Weinberger, Kilian Q. 2017, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261-2269.

[14] *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.* Szegedy, Sergey Ioffe and Christian. 2015.

[15] *Dropout: A Simple Way to Prevent Neural Networks from Overfitting.* Salakhutdinov, Nitish Srivastava and Geoffrey Hinton and Alex Krizhevsky and Ilya Sutskever and Ruslan. 2014, Journal of Machine Learning Research, Vol. 15, pp. 1929-1958.

[16] *A low shot learning method for tea leaf's disease identification.* Wan, Gensheng Hu and Haoyu Wu and Yan Zhang and Mingzhu. 2019, Computers and Electronics in Agriculture, Vol. 163, p. 104852.

[17] *MAF-MixNet: Few-Shot Tea Disease Detection Based on Mixed Attention and Multi-Path Feature Fusion.* Zhang, Wenjing and Tan, Ke and Wang, Han and Hu, Di and Pu, Haibo. 2025, Plants, Vol. 14.

[18] *Few-Shot Learning for Plant-Disease Recognition in the Frequency Domain.* Lin, Hong and Tse, Rita and Tang, Su-Kit and Qiang, Zhenping and Pau, Giovanni. 2022, Plants, Vol. 11.

[19] *Meta-Learning for Few-Shot Plant Disease Detection.* Chen, Liangzhe and Cui, Xiaohui and Li, Wei. 2021, Foods, Vol. 10.

[20] *Semi-supervised few-shot learning approach for plant diseases recognition.* Li, Yang and Chao, Xuewei. 2021, Plant Methods, Vol. 17.

[21] *Few-shot learning for plant disease recognition: A review.* Sun, Jianqiang and Cao, Wei and Fu, Xi and Ochi, Sunao and Yamanaka, Takehiko. 2023, Agronomy Journal, Vol. 116.

[22] *Evaluation of Different Few-Shot Learning Methods in the Plant Disease Classification Domain.* Uzhinskiy, Alexander. 2025, Biology, Vol. 14.

[23] *Few-shot crop disease recognition using sequence- weighted ensemble model-agnostic meta-learning.* Li, Junlong and Feng, Quan and Yang, Junqi and Zhang, Jianhua and Yang, Sen. 2025, Frontiers in Plant Science, Vol. 16.

[24] *Few-shot learning approach with multi-scale feature fusion and attention for plant disease recognition.* Hong Lin and Rita Tse and Tang, \{Su Kit\} and Qiang, \{Zhen Ping\} and Giovanni Pau. s.l. : Frontiers Media S.A., 2022, Frontiers in Plant Science, Vol. 13.

[25] *Few-shot disease recognition algorithm based on supervised contrastive learning.* Mu, Jiawei and Feng, Quan and Yang, Junqi and Zhang, Jianhua and Yang, Sen. 2024, Frontiers in Plant Science, Vols. Volume 15 - 2024.

[26] *Deep learning based computer vision approaches for smart agricultural applications.* Singh, V.G. Dhanya and A. Subeesh and N.L. Kushwaha and Dinesh Kumar Vishwakarma and T. {Nagesh Kumar} and G. Ritika and A.N. 2022, Artificial Intelligence in Agriculture, Vol. 6, pp. 211-229.

[27]*Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review.* Patrício, Diego and Rieder, Rafael. 2018, Computers and Electronics in Agriculture, Vol. 153.

[28] *Tea Leaf Disease Classification Using Deep Learning Techniques.* Srivastav, Somya and Guleria, Kalpna and Sharma, Shagun and Singh, Gurpreet. 2024, 2024 International Conference on Intelligent Computing and Emerging Communication Technologies (ICEC).

[29] *Matching Networks for One Shot Learning.* Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, Daan Wierstra. s.l. : arXiv, 2017.

[30] *A ConvNet for the 2020s.* Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. 2022, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[31] *Decoupled Weight Decay Regularization.* Loshchilov, I., & Hutter, F. 2019, International Conference on Learning Representations (ICLR).

[32] *A survey on image data augmentation for deep learning.* Shorten, C., & Khoshgoftaar, T. M. 2019, Journal of Big Data.

[33] *CBAM: Convolutional block attention module.* Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. 2018, Proceedings of the European conference on computer vision (ECCV).

[34] *Siamese neural networks for one-shot image recognition.* Koch, G., Zemel, R., & Salakhutdinov, R. 2015, ICML deep learning workshop.

[35] *Visual tea leaf disease recognition using a convolutional neural network model.* Chen, J., Liu, Q., & Gao, L. 2019, Symmetry.

BẢN CAM ĐOAN

Tôi cam đoan đã thực hiện việc kiểm tra mức độ tương đồng nội dung đề án qua phần mềm *Turnitin* một cách trung thực và đạt kết quả mức độ tương đồng 20% toàn bộ nội dung đề án. Bản đề án kiểm tra qua phần mềm là bản cứng đề án đã nộp để bảo vệ trước hội đồng. Nếu sai tôi xin chịu các hình thức kỷ luật theo quy định hiện hành của Học viện.

Hà Nội, ngày 5 tháng 2 năm 202

HỌC VIÊN CAO HỌC

Đinh Mạnh Hải

20% Tính tương đồng nói chung

Tổng cộng của tất cả các kết quả trùng khớp, bao gồm cả các nguồn trùng lặp, cho mỗi c...

Đã lọc khỏi Báo cáo

- Mục lục tham khảo
- Văn bản được trích dẫn
- Kết quả trùng khớp nhỏ (ít hơn 8 từ)

Các tiêu chí loại trừ

- 27 kết quả trùng khớp đã loại trừ

Nguồn hàng đầu

17% Nguồn Internet
8% Ấn bản
7% Bài tập được nộp (bài của học sinh)

BỘ KHOA HỌC VÀ CÔNG NGHỆ
HỌC VIỆN CÔNG NGHỆ
BƯU CHÍNH VIỄN THÔNG

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

BÁO CÁO GIẢI TRÌNH
SỬA CHỮA, HOÀN THIỆN ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

Họ và tên học viên: Đinh Mạnh Hải

Chuyên ngành: Hệ thống thông tin

Khóa: Khoá 2024 đợt 2

Tên đề tài: Phân loại một số bệnh trên cây chè dựa trên học sâu ít mẫu

Người hướng dẫn khoa học: PGS.TS. Phạm Văn Cường

Ngày bảo vệ: 24/01/2026

Các nội dung học viên đã sửa chữa, bổ sung trong đề án tốt nghiệp theo ý kiến đóng góp của Hội đồng chấm đề án tốt nghiệp:

TT	Ý kiến hội đồng	Sửa chữa của học viên
1	Bố cục báo cáo, sửa các lỗi chế bản, chính tả	Đã chỉnh sửa lại tên các thuật ngữ việt hoá. Đưa nội dung cơ chế tập trung từ chương 3 sang chương 2. Chỉnh sửa nội dung báo cáo kết thúc chương 2
2	Bổ sung, giải thích hình ảnh và công thức	Đã bổ sung giải thích về tập dữ liệu kiểm thử ở trang 5
3	Chỉnh sửa tài liệu tham khảo đúng định dạng	Đã chỉnh sửa định dạng trích dẫn

Hà Nội, ngày 02 tháng 02 năm 2026

Ký xác nhận của

CHỦ TỊCH HỘI ĐỒNG CHẤM ĐỀ ÁN	THƯ KÝ HỘI ĐỒNG	NGƯỜI HƯỚNG DẪN KHOA HỌC	HỌC VIÊN
PGS.TSKH. Hoàng Đăng Hải	TS. Phạm Hoàng Duy	PGS.TS. Phạm Văn Cường	Đinh Mạnh Hải

BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ
BƯU CHÍNH VIỄN THÔNG

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc
****

BẢN NHẬN XÉT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

(Dùng cho người phản biện)

Tên đề án: Phân loại một số bệnh trên cây chè dựa trên học sâu ít mẫu

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

Tên học viên: ĐINH MẠNH HẢI

Họ và tên người nhận xét: Trần Tiến Công

Học hàm, học vị: Tiến sĩ, Giảng viên

Chuyên ngành: Trí tuệ nhân tạo

Cơ quan công tác: Khoa Trí tuệ nhân tạo, Học viện Công nghệ BC-VT

NỘI DUNG NHẬN XÉT

I/ Cơ sở khoa học và thực tiễn, tính cấp thiết của đề tài:

• Đề tài có ý nghĩa thực tiễn lớn đối với nông nghiệp thông minh, đặc biệt là ngành trồng chè. Việc phát hiện sớm bệnh hại giúp giảm thiểu rủi ro mất mùa và hạn chế lạm dụng thuốc bảo vệ thực vật,.

• Tính cấp thiết được thể hiện rõ qua việc giải quyết "nút thắt cổ chai" về dữ liệu: Các phương pháp học sâu truyền thống cần lượng lớn ảnh gán nhãn (Big Data) rất tốn kém và khó thu thập đủ đối với các bệnh hiếm gặp. Hướng tiếp cận Học sâu ít mẫu (Few-shot Learning - FSL) là hoàn toàn phù hợp và hiện đại,.

II/ Về nội dung, chất lượng của đề án, các kết quả đạt được (so với đề cương đã được duyệt):

*** Nội dung đề án:** Đề án gồm 3 chương chính, cấu trúc logic và chặt chẽ:

• **Chương 1:** Tổng quan về bệnh hại trên lá chè (mô tả chi tiết đặc điểm sinh học, tác hại của các bệnh như thán thư, chấm xám...) và giới thiệu bài toán FSL (N-way K-shot),.

• **Chương 2:** Phân loại bệnh dựa trên học sâu ít mẫu: Đề xuất mô hình "Attentive Batch Model" kết hợp Backbone (ResNet/ConvNeXt) với Khung tích hợp quan hệ (RBI). Điểm nhấn là sử dụng hai mô-đun: Mã hóa vị trí quan hệ (RPE) và Chú ý quan hệ thặng dư (RRA) để khai thác ngữ cảnh trong batch,.

• **Chương 3:** Thực nghiệm và đánh giá: Thiết lập huấn luyện theo tập (Episodic Training), so sánh hiệu quả giữa các backbone và xây dựng ứng dụng Web Client-Server hỗ trợ chẩn đoán,.

*** Chất lượng đề án:** Đạt

*** Các kết quả đạt được:**

• Đề xuất được kiến trúc lai ghép giữa Mạng nguyên mẫu (Prototypical Networks) và cơ chế Attention (RBI).

• Kết quả thực nghiệm trên kịch bản 5-way 5-shot đạt độ chính xác cao (lên tới 98.27% với ConvNeXt-Base).

• Xây dựng được hệ thống demo có tính ứng dụng, giải quyết vấn đề tính toán lại nguyên mẫu bằng cơ chế "Prototype Memory".

III/ Những vấn đề cần giải thích thêm:

Mặc dù đề án có hàm lượng khoa học tốt, nhưng một số vấn đề về phương pháp luận và dữ liệu cần được làm rõ để tăng tính thuyết phục:

• Về Dữ liệu (Sự không nhất quán):

◦ Tại mục "Phương pháp nghiên cứu" (trang 6), học viên nêu: *"Thu thập dữ liệu từ các tập dữ liệu bệnh lá chè được công khai trên Kaggle"*. Tuy nhiên, tại mục 4 "Phạm vi nghiên cứu" (trang 5), học viên lại khẳng định: *"Tập dữ liệu ảnh lá chè do nhóm thực hiện thu thập và xử lý"*. Cần giải thích rõ tỷ lệ giữa dữ liệu tự thu thập thực tế và dữ liệu lấy từ Kaggle? Nếu dùng dữ liệu Kaggle, tính mới và độ khó của dữ liệu có được đảm bảo không?

• Về Phương pháp đề xuất (RBI và PSNR):

◦ Trong mô-đun RPE (Mã hóa vị trí quan hệ), học viên sử dụng chỉ số PSNR (Peak Signal-to-Noise Ratio) làm hàm đo độ tương đồng đặc trưng. PSNR thường dùng để đo chất lượng tái tạo ảnh (pixel-level) so với ảnh gốc. Việc áp dụng PSNR vào không gian đặc trưng (feature vectors) sau khi qua Backbone là một hướng đi lạ. Cần giải thích cơ sở lý thuyết tại sao PSNR lại hiệu quả hơn các độ đo phổ biến như Cosine Similarity hay Euclidean trong việc "khuếch đại sự khác biệt" giữa các véc-tơ đặc trưng trừu tượng?

◦ Công thức (6) tính dựa trên . Cần biện luận về tính ổn định của phép đo này khi các véc-tơ đặc trưng chưa được chuẩn hóa.

• Về Thực nghiệm và So sánh:

◦ Bảng 3 (trang 44) so sánh độ chính xác giữa "Kiến trúc đề xuất (Backbone + RBI)" với "ResNet-50" và "ResNet-101". Cách so sánh này chưa công bằng. Cần làm rõ: Các dòng "ResNet-50/101" ở đây là chạy Prototypical Networks *thuần túy* (không có RBI) hay chỉ là mạng phân loại Softmax thông thường? Để chứng minh hiệu quả của RBI, cần so sánh: *ProtoNet + ResNet (Baseline)* so với *ProtoNet + ResNet + RBI (Đề xuất)*.

◦ Kết quả 98.27% là rất cao. Liệu có hiện tượng rò rỉ dữ liệu (data leakage) giữa tập Support và Query trong quá trình chia Episode ngẫu nhiên không, đặc biệt nếu dữ liệu đầu vào là frame cắt từ video (các frame gần nhau rất giống nhau)?

• Về Ứng dụng:

◦ Hệ thống sử dụng "Prototype Memory" lưu trên RAM. Cần giải thích cơ chế cập nhật: Nếu hệ thống cần nhận diện một loại bệnh *mới* (New Class) chưa từng có trong Memory, người quản trị có cần huấn luyện lại toàn bộ mô hình (Backbone) hay chỉ cần nạp ảnh mẫu mới vào RAM? Đây là ưu điểm cốt lõi của FSL nhưng chưa được mô tả kịch bản thao tác cụ thể trong phần ứng dụng.

Câu hỏi:

1. Giải thích sự mâu thuẫn về nguồn gốc dữ liệu (Kaggle hay tự thu thập)? Quy trình gán nhãn cho dữ liệu tự thu thập được thực hiện như thế nào để đảm bảo độ chính xác chuyên môn?

2. Tại sao lựa chọn chỉ số PSNR để đo độ tương đồng giữa các véc-tơ đặc trưng trong mô-đun RPE thay vì Cosine Similarity?

3. Trong Bảng 3 so sánh độ chính xác, kiến trúc cơ sở (baseline) để so sánh là gì? (Là Prototypical Network thường hay là CNN phân loại truyền thống?).

4. Cơ chế "Prototype Memory" trên Server xử lý thế nào khi người dùng muốn thêm một loại bệnh mới vào hệ thống? Có cần tắt Server để training lại không?

IV/ Kết luận:

Đề án đáp ứng các tiêu chuẩn của Đề án tốt nghiệp thạc sĩ kỹ thuật.

Đồng ý cho phép học viên bảo vệ Đề án tốt nghiệp.

Hà Nội, ngày 20 tháng 01 năm 2026

Người phản biện
(Ký ghi rõ họ tên)



Trần Tiến Công