ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA

NGUYỄN THỊ PHƯƠNG THẢO

PHÁT HIỆN BỆNH CÂY CÀ PHÊ DỰA TRÊN HỌC MÁY

Chuyên ngành: Khoa học máy tính
Mã số: 8480101

TÓM TẮT LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Đà Nẵng – Năm 2025

Công trình được hoàn thành tại
TRƯỜNG ĐẠI HỌC BÁCH KHOA

Người hướng dẫn khoa học:
PGS.TS. NGUYỄN THANH BÌNH

Phản biện 1: PGS.TS Phan Trần Đăng Khoa

Phản biện 2: TS. Phạm Anh Phương

Luận văn sẽ được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ khoa học máy tính họp tại Trường Đại học Bách khoa vào ngày 29 tháng 11 năm 2025

Có thể tìm hiểu luận văn tại:

- Trung tâm Học liệu và truyền thông trường Đại học Bách khoa - Đại học Đà Nẵng
- Khoa công nghệ thông tin, trường Đại học Bách khoa - Đại học Đà Nẵng

MỞ ĐẦU

1. LÝ DO CHỌN ĐỀ TÀI

Cà phê có đóng góp to lớn vào nền kinh tế của nhiều quốc gia trên thế giới. Tại Việt Nam, chỉ riêng xuất khẩu cà phê đã đóng góp hơn 3% vào tổng sản lượng quốc nội - theo số liệu năm 2023 của Hiệp hội Cà phê và Ca cao Việt Nam (VICOFA). Tuy nhiên, sản lượng và chất lượng cà phê đang chịu ảnh hưởng nghiêm trọng từ các loại bệnh trên lá. Việc phát hiện bệnh sớm là yếu tố then chốt để bảo vệ mùa vụ và giảm thiểu thiệt hại.

Hiện nay, việc phát hiện bệnh trên lá cây cà phê chủ yếu dựa vào kinh nghiệm của người nông dân hoặc các chuyên gia nông nghiệp, dẫn đến hạn chế về tốc độ, đặc biệt khi quy mô canh tác ngày càng mở rộng. Trong bối cảnh đó, việc ứng dụng công nghệ học máy vào nhận diện bệnh lá cà phê là tiềm năng.

Xuất phát từ thực tiễn trên, đề tài "Phát hiện bệnh trên lá cây cà phê dựa trên học máy" được thực hiện với mục tiêu xây dựng và kiểm chứng mô hình phát hiện bệnh tự động, sử dụng dữ liệu công khai kết hợp dữ liệu thu thập trong nước. Với lợi thế về hạ tầng tính toán, nguồn thư viện mã nguồn mở và sự hỗ trợ từ các chương trình nghiên cứu trong lĩnh vực nông nghiệp – công nghệ, việc triển khai mô hình học sâu cho bài toán này là hoàn toàn khả thi và có ý nghĩa thực tiễn cao.

2. MỤC TIÊU VÀ NHIỆM VỤ

2.1. Mục tiêu

Các mục tiêu cụ thể của nghiên cứu bao gồm:

(i) Phát triển mô hình để phát hiện và phân loại các loại bệnh gỉ sắt ở lá, bệnh sâu đục lá, bệnh đốm nâu và bệnh đốm Cercospora của cây cà phê thông qua nhận diện hình ảnh.

(ii) Đánh giá hiệu quả và độ chính xác của hệ thống phát hiện bệnh trong điều kiện thực tế.

2.2. Nhiệm vụ

Các nhiệm vụ cụ thể cần thực hiện để đạt được mục tiêu nghiên cứu bao gồm:

(i) Nghiên cứu các giải pháp và công nghệ phục vụ dự đoán bệnh trên lá, cụ thể là bệnh gỉ sắt ở lá, bệnh sâu đục lá, bệnh đốm nâu và bệnh đốm Cercospora.

(ii) Triển khai thu thập dữ liệu hình ảnh bao gồm các lá không bệnh và các lá bị các bệnh được nêu trên.

(iii) Triển khai khoanh vùng bệnh trên hình ảnh lá và gán nhãn các loại bệnh trên lá.

(iv) Xây dựng chức năng tiền xử lý ảnh thu thập.

(v) Xây dựng mô hình học sâu để nhận diện lá và chẩn đoán bệnh trên lá cà phê.

(vi) Huấn luyện mô hình học sâu nhận diện lá và chẩn đoán bệnh trên lá cà phê.

(vii) Đánh giá và thử nghiệm các mô hình chẩn đoán trên dữ liệu kiểm tra để đánh giá hiệu suất dự đoán và phát hiện của mô hình.

(viii) Tinh chỉnh và cải tiến hiệu suất của mô hình đề xuất dựa trên dữ liệu thực tế thu thập trong trong các điều kiện khác nhau.

3. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU

3.1. Đối tượng nghiên cứu

(i) Mô hình học sâu: Mô hình phân tích và nhận diện bệnh từ hình ảnh lá cây.

(ii) Dữ liệu nghiên cứu: Dữ liệu bao gồm hình ảnh lá cây bệnh và bình thường.

3.2. Phạm vi nghiên cứu

Không gian nghiên cứu:

- Khu vực nghiên cứu: Các vườn cà phê tại Việt Nam.
- Địa điểm thực nghiệm: Các trang trại cà phê sẵn sàng hợp tác để thu thập dữ liệu hình ảnh thực tế.

Thời gian nghiên cứu: Tập trung vào giai đoạn cây cà phê dễ bị bệnh nhất

Ngoài ra các nguồn dữ liệu được công khai cũng được sử dụng để tối ưu hóa quá trình nghiên cứu.

4. PHƯƠNG PHÁP NGHIÊN CỨU

Phương pháp luận trong nghiên cứu của luận văn là sự kết hợp giữa các phương pháp lý thuyết và thực nghiệm.

5. Ý NGHĨA KHOA HỌC VÀ THỰC TIỄN

5.1. Ý nghĩa khoa học

(i) Cung cấp một phương pháp mới để chẩn đoán bệnh của cây cà phê thông qua hình ảnh.

(ii) Góp phần hoàn thiện cơ sở lý luận về nhận diện bệnh trên lá cây cà phê.

5.2. Ý nghĩa thực tiễn

(i) Hỗ trợ việc phát hiện sớm các loại bệnh trên lá cây cà phê, từ đó kịp thời triển khai các biện pháp xử lý, giảm thiểu thiệt hại trong sản xuất.

(ii) Đóng góp vào việc hiện đại hóa ngành nông nghiệp, tăng khả năng áp dụng công nghệ vào thực tiễn sản xuất.

6. BỐ CỤC LUẬN VĂN

Đề tài được tổ chức thành các phần như sau:

MỞ ĐẦU

CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI

Trong chương 1, những thách thức về vấn đề nhận diện sâu bệnh một cách kịp thời được trình bày. Tiếp theo, chương này sẽ tổng hợp các công nghệ và nghiên cứu liên quan đến việc áp dụng xử lý hình ảnh vào nông nghiệp, đồng thời đưa ra các đánh giá của phương pháp đang được ứng dụng.

CHƯƠNG 2: ĐỀ XUẤT MÔ HÌNH PHÁT HIỆN BỆNH TRÊN LÁ CÂY CÀ PHÊ

Chương 2 mô tả chi tiết các bước xử lý hình ảnh từ việc tiền xử lý dữ liệu đến việc sử dụng các kỹ thuật nâng cao để cải thiện chất lượng ảnh đầu vào. Sau đó, mô hình CNN sẽ được áp dụng để phân tích các đặc điểm của lá cây cà phê, xác định các dấu hiệu của bệnh thông qua việc huấn luyện trên bộ dữ liệu đã được chuẩn bị. Cuối cùng, một mô hình tổng thể, mô tả cách thức các bước xử lý hình ảnh và mạng học máy được kết hợp để đạt được hiệu quả cao nhất trong việc phát hiện bệnh sẽ được đưa ra.

CHƯƠNG 3: CÀI ĐẶT, KẾT QUẢ THỬ NGHIỆM VÀ ĐÁNH GIÁ

Chương này trình bày chi tiết thiết kế và triển khai mô hình. Các kết quả đạt được sẽ được trình bày cụ thể và được phân tích, so sánh với các nghiên cứu trước đây. Bên cạnh đó, luận văn cũng sẽ thảo luận về các yếu tố ảnh hưởng đến hiệu quả của mô hình, bao gồm chất lượng hình ảnh, độ phức tạp của mô hình và các kỹ thuật tiền xử lý dữ liệu. Những kết quả và hạn chế trong quá trình thực hiện mô hình sẽ được đưa ra để làm cơ sở cho các nghiên cứu tiếp theo.

KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO

CHƯƠNG 1: TỔNG QUAN ĐỀ TÀI

1.1. GIỚI THIỆU

Nhu cầu khai thác và phân tích thông tin từ hình ảnh, video ngày càng gia tăng nhờ sự phát triển mạnh mẽ của trí tuệ nhân tạo. Trong đó, thị giác máy tính đóng vai trò quan trọng giúp máy tính nhận biết và hiểu dữ liệu hình ảnh tương tự con người, hỗ trợ quá trình ra quyết định thông minh. Học máy và học sâu là nền tảng cốt lõi cho các ứng dụng này. Vì vậy, chương này trình bày tổng quan về công nghệ, phương pháp và các nghiên cứu liên quan, làm cơ sở cho việc phát triển và ứng dụng các giải pháp AI trong thực tiễn.

1.2. HỌC MÁY VÀ HỌC SÂU

1.2.1. Học máy (Machine Learning)

Học máy là quá trình giúp máy tính học từ dữ liệu để đưa ra quyết định mà không cần lập trình cứng. Học máy có thể được phân loại dựa trên mức độ gắn nhãn của dữ liệu huấn luyện thành các hướng tiếp cận chính sau: Học có giám sát (Supervised learning), Học không giám sát (Unsupervised learning), Học bán giám sát (Semi-supervised learning), và Học tự giám sát (Self-supervised learning).

1.2.2. Học sâu (Deep Learning)

Học sâu là một nhánh của học máy, sử dụng mạng nơ-ron sâu để tự động trích xuất đặc trưng phức tạp từ dữ liệu thô thay vì trích xuất thủ công như học máy truyền thống. Học sâu yêu cầu dữ liệu lớn và tài nguyên tính toán mạnh. Học sâu được ứng dụng chủ yếu vào các lĩnh vực thị giác máy tính, xử lý ngôn ngữ tự nhiên (natural language processing) và nhận dạng giọng nói (voice processing).

1.3. THỊ GIÁC MÁY TÍNH

1.3.1. Tổng quan về thị giác máy tính

1.3.1.1. Khái niệm và mục tiêu

Thị giác máy tính là lĩnh vực nghiên cứu giúp máy tính có khả năng hiểu và phân tích hình ảnh hoặc video tương tự như con người. Các tác vụ chính bao gồm phân loại ảnh, phát hiện đối tượng và phân đoạn ảnh. Đây là nền tảng quan trọng cho nhiều ứng dụng thực tiễn như nông nghiệp, y tế và xe tự lái.

1.3.1.2. Tổng quan quy trình xử lý ánh

Cơ bản, quy trình xử lý ảnh bao gồm: (i) Thu thập và xử lý dữ liệu ảnh, (ii) Trích xuất và học đặc trưng và (iii) Huấn luyện và đánh giá mô hình.

a) **Tiền xử lý dữ liệu và kỹ thuật tăng cường dữ liệu (Data Preprocessing and Augmentation)**

Trong thực tế, dữ liệu được thu thập thường chứa nhiễu hoặc không đồng nhất, dẫn đến các khó khăn trong việc phân tích dữ liệu cũng như xây dựng một mô hình học máy phù hợp. Vì vậy cần các bước tiền xử lý như chuẩn hóa dữ liệu (chuẩn hóa giá trị pixel, resize ảnh) và cân bằng dữ liệu để đảm bảo dữ liệu đầu vào nhất quán, giảm sai lệch, từ đó tăng độ chính xác cho các mô hình.

b) **Phát hiện đặc trưng trong ảnh (Feature Detection)**

Phát hiện đặc trưng có mục tiêu chính là xác định các điểm, cạnh hoặc vùng trong ảnh có khả năng phân biệt cao. Các đặc trưng này cần có tính bất biến đối với phép biến đổi như dịch chuyển, xoay, phóng to, thu nhỏ và thay đổi độ sáng.

Trước khi học sâu phát triển, các phương pháp phát hiện đặc trưng truyền thống chủ yếu dựa trên kỹ thuật xử lý ảnh và toán học, tiêu biểu như Corner Detection, SIFT, SURF, ORB và Canny Edge Detection.

Tổng thể, các phương pháp truyền thống này giúp thiết lập cơ sở lý thuyết cho việc phát hiện, mô tả và so khớp đặc trưng trong ảnh, tạo tiền đề cho các phương pháp trích xuất đặc trưng tự động trong mạng nơ-ron tích chập (CNN) sau này

c) **Liên hệ với học sâu hiện đại**

Trong các phương pháp học sâu hiện đại, đặc trưng không còn được trích xuất thủ công bằng các thuật toán truyền thống, mà được học tự động thông qua mạng nơ-ron tích chập (Convolutional Neural Networks – CNN). Các lớp tích chập trong CNN có khả năng học được những đặc trưng từ cơ bản (cạnh, góc) đến phức tạp (hình dạng, cấu trúc đối tượng) một cách tự động từ dữ liệu huấn luyện. Điều này giúp loại bỏ sự phụ thuộc vào việc thiết kế thủ công và tối ưu hóa đặc trưng.

1.3.1.3. Các bài toán trong thị giác máy tính

a) **Phân loại ảnh (Image Classification)**

Giới thiệu bài toán phân loại ảnh

Bài toán phân loại ảnh đặt mục tiêu là gán cho toàn bộ ảnh đầu vào một nhãn duy nhất đại diện cho loại đối tượng chính trong ảnh. Đầu ra của bài toán thường là một vector xác suất phân bố trên các lớp, trong đó lớp có xác suất cao nhất sẽ được chọn làm nhãn dự đoán cuối cùng.

Giới thiệu các mô hình

- Resnet (Residual Network) do He và cộng sự [1] đề xuất là kiến trúc CNN nổi bật nhờ cơ chế học dư (Residual Learning) với kết nối tắt (skip connection), giúp khắc phục hiện tượng mất hoặc bùng nổ gradient khi mạng quá sâu. ResNet cho phép huấn luyện mạng có độ sâu và độ chính xác cao, tuy nhiên, ResNet cũng tồn tại hạn chế như kích thước mô hình lớn, chi phí tính toán cao và khi tăng số tầng quá nhiều thì hiệu quả cải thiện dần bão hòa.
- MobileNet được Howard và cộng sự [2] giới thiệu như một kiến trúc mạng nơ-ron tích chập tối ưu cho các thiết bị di động và nhúng. MobileNet áp dụng cơ chế *Depthwise Separable Convolution* giúp giảm đáng kể số lượng tham số và phép tính toán so với tích chập chuẩn, trong khi vẫn duy trì được khả năng biểu diễn đặc trưng của mô hình. MobileNet có ưu điểm nhẹ, suy luận nhanh và dễ triển khai trên thiết bị di động, nhưng độ chính xác kém hơn các mô hình lớn như ResNet hay Inception.

b) **Phát hiện đối tượng (Object Detection)**

Bài toán phát hiện đối tượng nhằm mục đích tìm vị trí chứa đối tượng đang được quan tâm. Đầu ra bài toán là mỗi đối tượng được biểu diễn bởi một nhãn lớp đi kèm với một hộp giới hạn (rectangle/window) bao quanh vị trí của nó trong ảnh.

Giới thiệu các mô hình

- R-CNN: Được Girshick và cộng sự [3] đề xuất, R-CNN là một trong những kiến trúc tiên phong trong lĩnh vực phát hiện đối tượng. Kiến trúc này bao gồm ba thành phần chính: Vùng đề xuất hình ảnh (Region Proposal), Trích xuất đặc trưng (Feature Extraction) và Phân loại (Classification). Mặc dù đạt độ chính xác cao, R-CNN chậm do phải xử lý riêng từng vùng đề xuất.
- Faster R-CNN: Được phát triển nhằm khắc phục hạn chế về tốc độ của R-CNN, Faster R-CNN không dùng thuật toán selective search để lấy ra các region proposal, mà nó thêm một mạng CNN mới gọi là Region Proposal Network (RPN) để tìm các region proposal [4]. Kiến trúc này để tự động sinh vùng đề xuất từ feature map, cùng RoI Pooling giúp chuẩn hóa kích thước đầu vào. Quá trình phân loại và hồi quy bounding box được thực hiện trong cùng một mạng, giúp mô hình thống nhất, nhanh và chính xác hơn.

c) **Bài toán phân đoạn ảnh (Image Segmentation)**

Bài toán phân đoạn ảnh (Image Segmentation) với mục tiêu là chia ảnh thành các vùng có ý nghĩa, trong đó mỗi pixel được gán nhãn theo lớp mà nó thuộc về. Điều này cho phép mô hình biết chính xác hình dạng và diện tích của đối tượng. Đầu ra bài toán là một mask phân đoạn (ảnh nhị phân hoặc đa kênh) biểu diễn đối tượng. Có hai loại trong phân đoạn ảnh là Semantic segmentation và Instance segmentation. Semantic segmentation gán nhãn cho từng pixel theo lớp, không phân biệt các cá thể cùng lớp. Trong khi đó Instance segmentation phân biệt từng cá thể riêng biệt trong cùng một lớp.

Giới thiệu các mô hình

- U-net: Đây là một kiến trúc mạng nơ-ron tích chập được thiết kế chuyên biệt cho các tác vụ phân đoạn ảnh, được Ronneberger và cộng sự [5] đề xuất. Mạng có cấu trúc hình chữ U đối xứng gồm hai phần: Encoder (thu hẹp) và Decoder (mở rộng). U-Net thường tối ưu bằng creoss-entropy loss trên từng điểm ảnh, có hiệu quả vượt trội trong y học và các lĩnh vực thị giác máy tính khác. Tuy nhiên, hạn chế đáng chú ý của mô hình này là tốc độ xử lý còn chậm khi áp dụng trên dữ liệu quy mô lớn.
- Segment Anything Model (SAM): Đây là một mô hình nền tảng trong lĩnh vực phân đoạn ảnh, được phát triển bởi Meta AI [6]. SAM được thiết kế để phân đoạn bất kỳ đối tượng nào trong ảnh hoặc video mà không cần huấn luyện bổ sung. SAM bao gồm ba thành phần cốt lõi hoạt động phối hợp bao gồm: Image Encoder, Prompt Encoder và Mask Decoder. Nhược điểm là kích thước lớn, tiêu tốn tài nguyên và phụ thuộc vào chất lượng prompt cũng như miền dữ liệu.

1.3.2. Các phương pháp đánh giá mô hình (Model Evaluation)

Trong quá trình xây dựng một mô hình học máy, để biết được chất lượng của mô hình chúng ta cần có các chỉ số để đánh giá mô hình.

1.3.2.1. Các chỉ số cho bài toán phân loại hình ảnh

Các chỉ số cơ bản bao gồm accuracy (Độ chính xác tổng thể), precision (Độ chính xác dự đoán dương tính), recall (Độ nhạy/ Tỷ lệ dương tính thật), F1-score. Trong phân loại nhiều lớp, các chỉ số trên vẫn áp dụng nhưng cần có phương pháp mở rộng, chẳng hạn như micro-averaging và macro-averaging.

1.3.2.2. Các chỉ số cho bài toán object detection

Bài toán phát hiện đối tượng yêu cầu mô hình dự đoán chính xác cả vị trí (bounding box) và nhãn lớp của đối tượng. Các chỉ số đánh giá thường dùng bao gồm precision, recall, intersection over Union (IoU), average Precision (AP) và mean Average Precision (mAP).

1.3.2.3. Các chỉ số cho bài toán segmentation

Phân đoạn ảnh là bài toán phức tạp hơn, đòi hỏi mô hình xác định chính xác ranh giới và vùng đối tượng trong ảnh. Các chỉ số thường dùng bao gồm: pixel accuracy, Intersection over Union (IoU) / Jaccard Index, Dice Coefficient (F1 Score cho phân đoạn), mean IoU (mIoU), Boundary F1-score (BF Score).

1.4. BÀI TOÁN NHẬN DIỆN LÁ CÀ PHÊ

1.4.1. Giới thiệu bài toán

Trong lĩnh vực thị giác máy tính ứng dụng cho nông nghiệp, bài toán nhận diện lá cà phê là một ví dụ cho việc chuyển đổi dữ liệu hình ảnh thực vật thành thông tin chẩn đoán có giá trị. Mục tiêu của bài toán là xây dựng một hệ thống có khả năng tự động nhận biết và phân loại tình trạng bệnh lý của lá cà phê thông qua hình ảnh, nhằm thay thế hoặc hỗ trợ quá trình kiểm tra thủ công truyền thống. Bản chất của bài toán nằm ở việc phân tích hình thái và sắc tố trên bề mặt lá đặc trưng cho từng loại bệnh, do đó việc trích xuất chính xác các đặc trưng này đóng vai trò quyết định đối với hiệu quả nhận diện.

Dưới góc độ kỹ thuật, bài toán nhận diện lá cà phê là sự kết hợp của ba nhóm tác vụ chính trong thị giác máy tính, bao gồm:

- Phân loại ảnh: Xác định xem lá cà phê thuộc nhóm khỏe mạnh hay mắc một loại bệnh cụ thể.
- Phát hiện đối tượng: Khoanh vùng và nhận diện vị trí của lá trong ảnh, đặc biệt quan trọng khi ảnh được chụp trong môi trường tự nhiên có nhiều yếu tố nhiễu.
- Phân đoạn ảnh: Xác định ranh giới chi tiết của lá hoặc vùng bệnh, giúp mô hình tập trung vào các khu vực mang thông tin hữu ích.

Dữ liệu hình ảnh trong nông nghiệp có nhiều đặc thù khiến việc nhận diện trở nên phức tạp, có thể khái quát qua các yếu tố sau: tính đa dạng cao, đặc điểm bệnh biến đổi linh hoạt, mất cân bằng dữ liệu, nền ảnh phức tạp. Do đó, việc giải quyết bài toán này đòi hỏi sự kết hợp giữa các kỹ thuật xử lý ảnh truyền thống và phương pháp học sâu hiện đại, nhằm trích xuất hiệu quả các đặc trưng quan trọng, giảm nhiễu và nâng cao độ chính xác của mô hình.

1.4.2. Các nghiên cứu liên quan

Các mô hình học sâu, đặc biệt là mạng nơ-ron tích chập, đã chứng minh hiệu quả trong phân tích hình ảnh và được ứng dụng rộng rãi trong nông nghiệp, đặc biệt là trong việc phát hiện và phân loại bệnh trên cây trồng. Sorte và cộng sự [7] áp dụng các phương pháp tính toán để nhận biết bệnh đốm lá (Cercospora) và gỉ sắt (Rust) trên lá cà phê. Do hai căn bệnh này không có hình dạng nên phương pháp trích xuất thuộc tính kết cấu (bao gồm thuộc tính thống kê và mẫu nhị phân cục bộ) để nhận dạng mẫu đã được áp dụng. Nghiên cứu so sánh kỹ thuật đào tạo từ đầu (training from scratch) và học chuyển giao (transfer learning) của Paulos và Woldeyohannis [8] trên bộ dữ liệu 1.120 ảnh thu thập từ trung tâm nghiên cứu nông nghiệp tại Ethiopia cho ra kết quả đáng chú ý: học chuyển giao thông qua Resnet50 đạt tỷ lệ chính xác là 99,89% vớiđộ lệch (test loss)

chỉ 1.40%. Eric Hitimana và cộng sự [9] xây dựng bộ dữ liệu gồm 37.939 hình ảnh về cây cà phê ở Rwanda, với sự xuất hiện của bệnh gỉ sắt, thợ mỏ và nhện đỏ. Nhóm tác giả đã áp dụng các mô hình học sâu như InceptionV3, ResNet50, Xception, VGG16 và DenseNet để phát hiện các bệnh trên cây cà phê. Thử nghiệm đã chỉ ra mô hình DenseNet có độ chính xác tốt nhất, đạt 99.57%. Martinez và cộng sự [10] cũng áp dụng Resnet50 để nhận diện các bệnh thường gặp trên lá cây cà phê với bộ dữ liệu có sẵn gồm 1250 lá cây ở Columbia. Nghiên cứu của Yamashita và Leite [11] được thực hiện trên hai bộ dữ liệu về lá cà phê là BRACOL và RoCoLe, bao gồm các bệnh như gỉ sắt, nấm bồ hóng, Cercospora, Phoma và sâu đục lá. Các tác giả đã phát triển một ứng dụng điện thoại thông minh để phân đoạn ngữ nghĩa (semantic segmentation) và phân loại các triệu chứng bệnh bằng hình ảnh lá cà phê. Marcos và cộng sự [12] đã sử dụng thuật toán di truyền để phát hiện bệnh gỉ sắt nhưng phương pháp của họ gặp khó khăn do độ sáng không đồng nhất và dựa vào một tập dữ liệu nhỏ. Trong một nghiên cứu song song, Marcos và cộng sự [13] đã sử dụng CNN để phát hiện bệnh gỉ sắt nhưng không tinh chỉnh mô hình, hạn chế hiệu suất của nó trên các tập dữ liệu đa dạng. Tassis và cộng sự [14] cũng nghiên cứu trên bộ dữ liệu BRACOL sử dụng phương pháp học sâu kết hợp giữa instance segmentation và sematic segmentation để xác định bệnh và sâu hại trên lá cây cà phê. Trong giai đoạn đầu, tác giả sử dụng mạng Mask R-CNN để instance segmentaion. Tiếp theo, mạng UNet và PSPNet được sử dụng để semantic segmentation và ResNet được sử dụng cuối cùng cho giai đoạn phân loại. Nghiên cứu đạt được MIoU 94,25% và 93,54% khi sử dụng lần lượt mạng UNet và PSPNet. Nghiên cứu của Faisal và cộng sự [15] trên bộ dữ liệu RoCoLe đề xuất phương pháp hợp nhất đặc trưng lai giữa MobinetV3 và Swin Transformer. Phương pháp kết hợp này đạt được độ chính xác là 84,29%.

Tổng quan lại, các thách thức chính bao gồm: (i) Phụ thuộc vào dữ liệu gán nhãn thủ công; (ii) Thiếu tính tổng quát hóa trên các tập dữ liệu mới, (iii) Chưa tích hợp hiệu quả giữa các giai đoạn phân đoạn và phân loại.

Nghiên cứu hiện tại khắc phục các hạn chế trên bằng cách tích hợp kỹ thuật phân đoạn mạnh mẽ (SAM2 + Faster RCNN) và học tự giám sát (SSL), giúp nâng cao độ chính xác và khả năng áp dụng thực tế trong điều kiện nông nghiệp tại hiện trường.

1.5. KẾT LUẬN CHƯƠNG

Chương này trình bày cơ sở lý thuyết và tổng quan về các phương pháp, mô hình cũng như hướng nghiên cứu liên quan đến bài toán nhận diện hình ảnh, đặc biệt trong lĩnh vực nông nghiệp. Các khái niệm nền tảng như học máy, học sâu và thị giác máy tính đã được phân tích nhằm làm rõ vai trò của chúng trong việc xây dựng các hệ thống nhận diện tự động. Bên cạnh đó, chương cũng đã giới thiệu đặc điểm và thách thức của bài toán nhận diện lá cà phê. Những nội dung này đóng vai trò là nền tảng lý luận và thực tiễn cho quá trình đề xuất và xây dựng mô hình phát hiện bệnh trên lá cây cà phê, sẽ được trình bày chi tiết trong Chương 2.

CHƯƠNG 2: ĐỀ XUẤT MÔ HÌNH PHÁT HIỆN BỆNH TRÊN LÁ CÂY CÀ PHÊ

2.1. GIỚI THIỆU

Chương này trình bày mô hình phát hiện bệnh trên lá cà phê được đề xuất, bao gồm: nguồn dữ liệu sử dụng, quy trình tiền xử lý ảnh để chuẩn hóa và nâng cao chất lượng dữ liệu, cùng kiến trúc hệ thống tích hợp các mô hình và phương pháp học sâu. Mục tiêu là xây dựng một quy trình phát hiện và phân loại bệnh chính xác, hiệu quả và có khả năng ứng dụng thực tiễn.

2.2. QUY TRÌNH THU THẬP VÀ CHUẨN BỊ DỮ LIỆU

Trong lĩnh vực phát hiện và phân loại bệnh trên cây cà phê, dữ liệu ảnh đóng vai trò then chốt, quyết định trực tiếp đến hiệu suất của các mô hình học sâu. Việc lựa chọn bộ dữ liệu phù hợp giúp đảm bảo độ tin cậy và phản ánh sự đa dạng về môi trường, loại bệnh và hình thái lá. Nghiên cứu này sử dụng hai bộ dữ liệu công khai RoCoLe và BRACOL để phục vụ quá trình huấn luyện và đánh giá mô hình.

2.2.1. Bộ dữ liệu RoCoLe

RoCoLe (Robusta Coffee Leaf Dataset) [16] gồm 1.560 ảnh lá cà phlà bộ dữ liệu ảnh lá cà phê Robusta được gán nhãn theo ba lớp: lá khỏe mạnh (*healthy*), lá không khỏe (*unhealthy*), và lá bị bệnh (*diseased*). Nhóm lá bệnh chủ yếu thể hiện triệu chứng của gỉ sắt lá cà phê (*coffee leaf rust*) và sâu hại nhện đỏ (*red spider mite infestation*). Dữ liệu được thu thập tại hiện trường từ 390 cây cà phê trong điều kiện nền phức tạp, giúp đánh giá khả năng tổng quát hóa và độ bền vững của mô hình trong môi trường thực tế.

Điểm nổi bật của RoCoLe là đi kèm với tệp chú thích JSON, trong đó các lá được đánh dấu bằng đa giác (polygon) dựa trên các điểm bao quanh phần biên lá

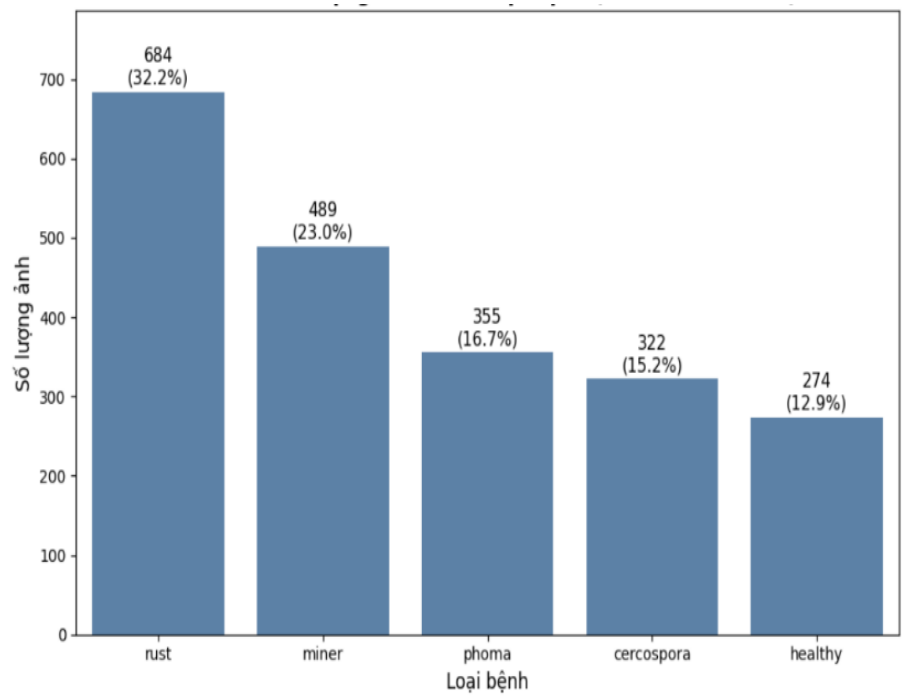
Hình 2.1 Ví dụ về polygon trong tập dữ liệu RoCoLe

Trong khuôn khổ nghiên cứu này, RoCoLe được khai thác theo mục tiêu bổ trợ, cụ thể là huấn luyện mô hình phát hiện đối tượng (object detection) nhằm sinh ra bounding box bao quanh lá cà phê, từ đó cung cấp prompt tự động cho mô hình SAM2 trong giai đoạn phân đoạn.

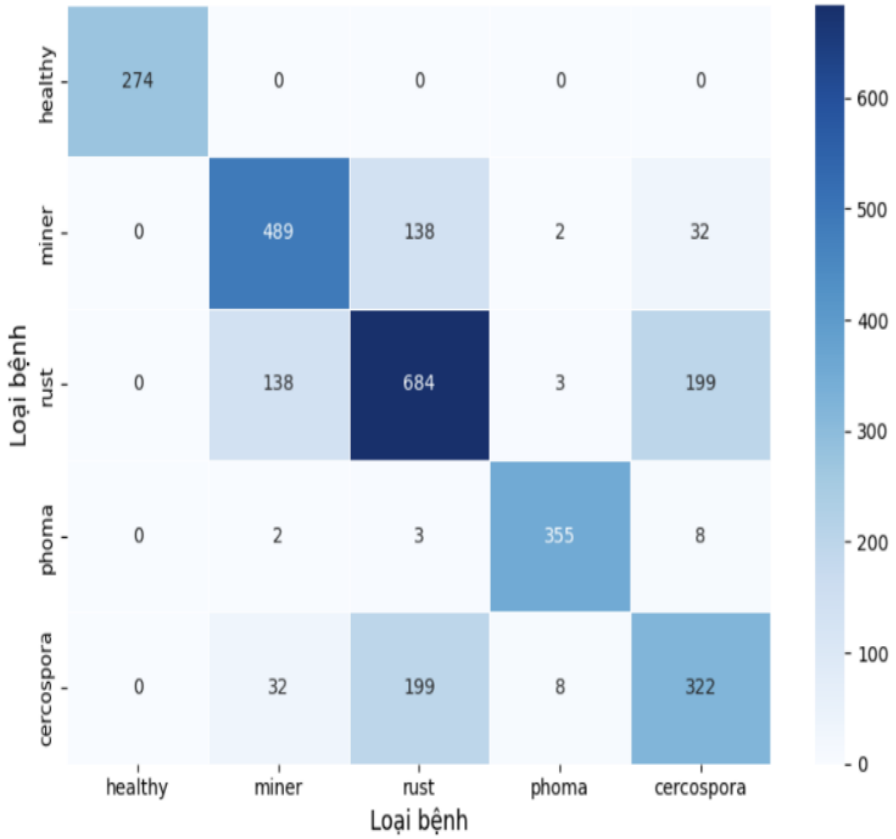
2.2.2. Bộ dữ liệu BRACOL

BRACOL (Brazilian Arabica Coffee Leaf Dataset) [17] bao gồm 1.747 ảnh lá cà phê Arabica, được ghi nhận bằng điện thoại thông minh trong điều kiện môi trường được kiểm soát, tập trung vào vùng lá để nhấn mạnh đặc trưng hình thái và màu sắc. Tập dữ liệu bao phủ cả lá khỏe mạnh và các trường hợp stress sinh học (biotic stress) gồm: Leafminer (sâu đục lá), Rust (gỉ sắt), Brown leaf spot (đốm nâu) và Cercospora leaf spot (đốm Cercospora).

Các ảnh được gán nhãn chi tiết, thể hiện rõ các loại stress sinh học ảnh hưởng đến lá cà phê. BRACOL phù hợp cho tác vụ phân loại bệnh. Tuy nhiên hiện tượng mất cân bằng dữ liệu vẫn tồn tại (xem hình 2.3).



Hình 2.3 Phân bố số lượng ảnh theo loại bệnh trong bộ BRACOL



Hình 2.4 Ma trận đồng xuất hiện giữa các bệnh ở lá

Trong nghiên cứu này, BRACOL được sử dụng cho giai đoạn huấn luyện và đánh giá mô hình phân loại bệnh, khai thác hiệu quả các đặc điểm hỗ trợ nhận dạng bệnh.

2.2.3. Vai trò của các bộ dữ liệu

Sự kết hợp giữa hai bộ dữ liệu mang lại các lợi thế bổ sung cho nhau. Cụ thể, RoCoLe bao gồm các ảnh chụp trong môi trường tự nhiên kèm theo tệp chú thích định dạng *JSON* cung cấp tọa độ vùng bao quanh lá, phù hợp cho huấn luyện mô hình phát hiện và phân đoạn lá cà phê. Trong khi BRACOL chứa các ảnh được gán nhãn chi tiết theo từng loại bệnh và thu thập trong điều kiện môi trường được kiểm soát, thích hợp cho giai đoạn huấn luyện mô hình phân loại bệnh.

2.3. QUY TRÌNH TIỀN XỬ LÝ DỮ LIỆU ẢNH

2.3.1. Bộ dữ liệu RoCoLe

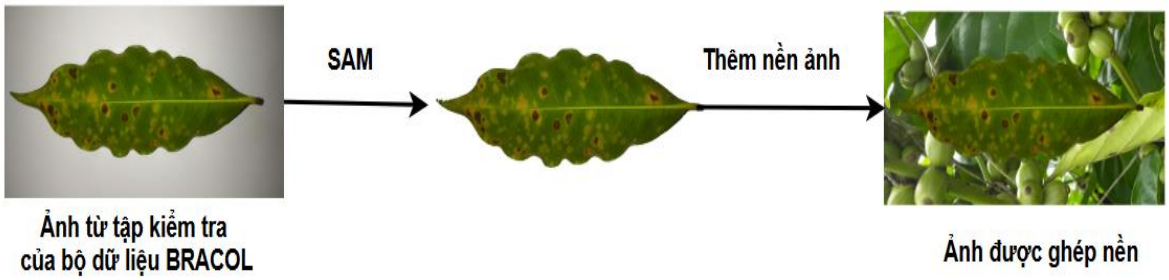
Các file JSON chứa các tọa độ polyon được tính toán để tạo thành các file JSON chứa annotation có dạng [xmin, ymin, xmax, ymax], mô tả danh sách các hộp giới hạn tương ứng với đối tượng trong ảnh. Hệ thống tiến hành đọc toàn bộ ảnh từ thư mục gốc và đối chiếu từng ảnh với tệp chú thích cùng tên để đảm bảo mỗi ảnh đều có dữ liệu nhãn hợp lệ, loại bỏ các trường hợp thiếu hoặc sai định dạng.

2.3.2. Bộ dữ liệu BRACOL

Hệ thống đối chiếu ảnh với tệp nhãn CSV để loại bỏ dữ liệu không hợp lệ, giữ lại 1.747 ảnh hợp lệ thuộc 5 lớp: bốn loại bệnh (gỉ sắt, sâu đục lá, đốm nâu, đốm Cercospora) và lá khỏe mạnh. Dữ liệu được chia theo tỷ lệ 80% cho huấn luyện và 20% cho đánh giá hệ thống. Các ảnh được chuẩn hóa về định dạng RGB, kích thước 224×224 pixel và áp dụng tăng cường dữ liệu như xoay, lật, chuyển đổi tensor nhằm tăng tính đa dạng và giảm overfitting.

2.3.3. Bộ dữ liệu kiểm thử

Bộ dữ liệu kiểm thử được trích từ 20% tập BRACOL, dùng để mô phỏng điều kiện môi trường thực tế nhằm đánh giá tính bền vững của hệ thống. Ảnh lá cà phê sau khi được SAM2 phân đoạn sẽ loại bỏ nền gốc, giữ lại vùng lá và chèn vào các nền phức tạp nhân tạo với mức nhiễu khác nhau để tạo bộ dữ liệu thử nghiệm giả lập.



Hình 2.7 Minh họa quy trình tạo ảnh mô phỏng

Quy trình này giúp tái hiện các tình huống phổ biến mà người nông dân có thể gặp khi chụp ảnh lá bằng điện thoại trong điều kiện tự nhiên, nơi ánh sáng, màu sắc nền và độ tương phản có thể làm suy giảm hiệu quả nhận dạng bệnh.

2.4. LỰA CHỌN MÔ HÌNH VÀ XÂY DỰNG HỆ THỐNG TỔNG THỂ

2.4.1. Cơ sở lựa chọn mô hình

Trong nghiên cứu này, việc lựa chọn mô hình và công nghệ được cân nhắc dựa trên ba tiêu chí chính:

(i) Tính phù hợp với từng giai đoạn xử lý trong hệ thống.

(ii) Khả năng tổng quát hóa trong môi trường thực tế có nhiều yếu tố nhiễu.

(iii) Hiệu suất đã được chứng minh trong các nghiên cứu liên quan đến bệnh lá cà phê.

Theo đó, hệ thống đề xuất kết hợp các mô hình sau:

- Faster R-CNN kết hợp MobileNetV3: Phát hiện và khoanh vùng vị trí lá cà phê trong ảnh.
- SAM2: Phân đoạn chính xác vùng lá dựa trên cơ chế gợi ý (prompting) tự động.
- Tự giám sát qua tô màu ảnh: Đóng vai trò tiền huấn luyện, giúp mô hình tự học đặc trưng màu sắc tự nhiên của lá mà không cần dữ liệu gán nhãn.
- Mạng ResNet50: Phân loại bệnh dựa trên đặc trưng màu sắc và hình thái học đã được học.

2.4.2. Giai đoạn 1 – Phát hiện và phân đoạn lá cà phê

2.4.2.1. Phát hiện lá bằng Faster R-CNN kết hợp MobileNetV3

Mô hình Faster R-CNN được sử dụng để xác định vị trí lá cà phê, với MobileNetV3 làm backbone nhằm giảm chi phí tính toán và tăng tốc xử lý. Sự kết hợp này giúp đạt cân bằng giữa độ chính xác và hiệu năng, tạo ra bounding box chất lượng cao phục vụ bước phân đoạn bằng SAM2.

2.4.2.2. Phân đoạn lá bằng SAM2

Các bounding box từ Faster R-CNN được dùng làm prompt cho mô hình SAM2, giúp tách chính xác lá cà phê khỏi nền. Cách tiếp cận này loại bỏ nhiễu nền và bảo toàn chi tiết lá, tạo đầu vào sạch và nhất quán cho giai đoạn phân loại.

2.4.3. Giai đoạn 2 – Phân loại bệnh trên lá cà phê

Sau khi thu được ảnh lá đã phân đoạn, hệ thống tiến hành phân loại tình trạng bệnh lý của lá thông qua hai bước kế tiếp nhau: tiền huấn luyện tự giám sát và huấn luyện có giám sát.

2.4.3.1. Tiền huấn luyện tự giám sát bằng tô màu ảnh (Colorization)

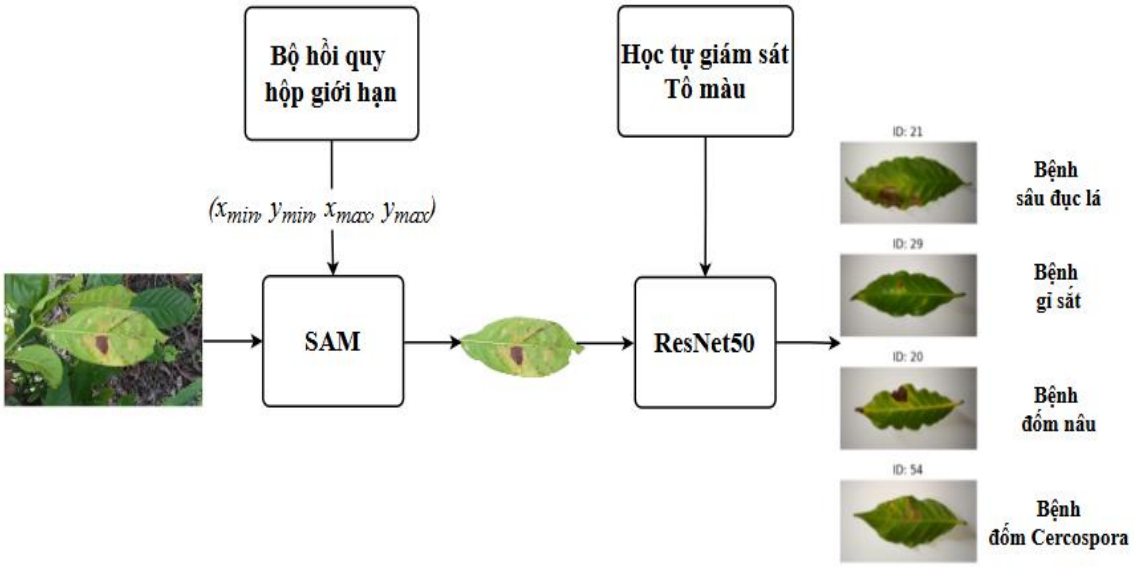
Mô hình encoder–decoder CNN học cách dự đoán hai kênh màu (a, b) từ kênh sáng (L) trong không gian Lab, giúp nắm bắt đặc trưng sắc độ, kết cấu và hình thái, giảm phụ thuộc vào dữ liệu có nhãn.

2.4.3.2. Huấn luyện có giám sát và fine-tune với ResNet50

Encoder từ bước trước được dùng để khởi tạo ResNet50, giúp mô hình hội tụ nhanh hơn, ổn định hơn và tổng quát hóa tốt hơn khi huấn luyện trên bộ dữ liệu BRACOL.

2.4.4. Tổng thể hệ thống đề xuất

Tổng thể hệ thống được thiết kế theo pipeline hai giai đoạn như minh họa ở Hình 2.8. Cụ thể từng giai đoạn được sơ đồ hóa ở hình 2.9 và 2.10.



Hình 2.8 Sơ đồ tổng thể hệ thống đề xuất

2.5. KẾT LUẬN CHƯƠNG

Chương 2 đã mô tả quy trình chuẩn bị dữ liệu, tiền xử lý ảnh và xây dựng pipeline hai giai đoạn cho hệ thống phát hiện bệnh trên lá cà phê. Việc kết hợp RoCoLe và BRACOL vừa hỗ trợ hệ thống phát hiện phần cần phân đoạn, vừa giúp hệ thống học đặc trưng hình thái và nhận dạng chính xác các loại bệnh. Chương 3 sẽ trình bày chi tiết quá trình thực nghiệm và đánh giá kết quả của hệ thống này.

CHƯƠNG 3: CÀI ĐẶT, KẾT QUẢ THỬ NGHIỆM VÀ ĐÁNH GIÁ

3.1. GIỚI THIỆU

Chương này trình bày quá trình cài đặt, thực nghiệm và đánh giá hiệu quả của hệ thống phát hiện và phân loại bệnh lá cà phê được đề xuất cũng như mô tả môi trường thực nghiệm, dữ liệu, thông số huấn luyện và so sánh kết quả của hệ thống với các phương pháp truyền thống, qua đó đánh giá hiệu quả và ưu thế của mô hình đề xuất.

3.2. MÔI TRƯỜNG THỰC NGHIỆM VÀ DỮ LIỆU

3.2.1. Môi trường thực nghiệm

Quá trình triển khai hệ thống phát hiện bệnh lá cà phê được thực hiện theo hai giai đoạn chính: (1) phân đoạn lá và (2) phân loại bệnh. Toàn bộ quy trình được triển khai trên nền tảng Kaggle với GPU Tesla P100 và bộ nhớ 16GB sử dụng thư viện PyTorch, openCV để xây dựng và huấn luyện mô hình.

3.2.2. Dữ liệu

Các thí nghiệm được thực hiện trên các ảnh hợp lệ từ bộ dữ liệu RoCoLe và BRACOL và tập dữ liệu thu thập được, sử dụng các chỉ số đánh giá phổ biến như mIoU cho phân đoạn và accuracy, F1-score, confusion matrix cho phân loại.

3.3. GIAI ĐOẠN PHÂN ĐOẠN

3.3.1. Dữ liệu và thông số huấn luyện

3.3.1.1. Dữ liệu

Giai đoạn huấn luyện hộp giới hạn

Trong giai đoạn này, bộ dữ liệu RoCoLe được sử dụng nhằm huấn luyện và đánh giá mô hình phân đoạn lá cà phê. Ảnh đầu vào được chuyển từ dạng BGR sang RGB để phù hợp với mô hình học sâu. Các tệp chú thích (annotation) ở định dạng JSON chứa thông tin về tọa độ khung giới hạn (bounding boxes) của các lá cà phê, được đọc và chuyển đổi thành tensor phục vụ cho quá trình huấn luyện. Mỗi ảnh được gán nhãn duy nhất (class = 1) để biểu diễn đối tượng lá cà phê trong bài toán phát hiện và phân đoạn.

Dữ liệu được chia ngẫu nhiên có kiểm soát (random_state = 42) thành 85% cho huấn luyện (1326 ảnh) và 15% cho kiểm thử (234 ảnh), nhằm đảm bảo tính tái lập và đánh giá khách quan trong quá trình huấn luyện mô hình.

Giai đoạn phân đoạn

Sau giai đoạn trên, hai tập dữ liệu của tập BRACOL và RoCoLe sẽ được sử dụng để đánh giá và trực quan quá kết quả quá trình phân đoạn kết hợp giữa hộp giới hạn và SAM2. Cụ thể các bộ dữ liệu sẽ được đi qua mô hình đề xuất, ghép nền trắng để so sánh các biểu diễn đặc trưng.

3.3.1.2. Thiết lập huấn luyện

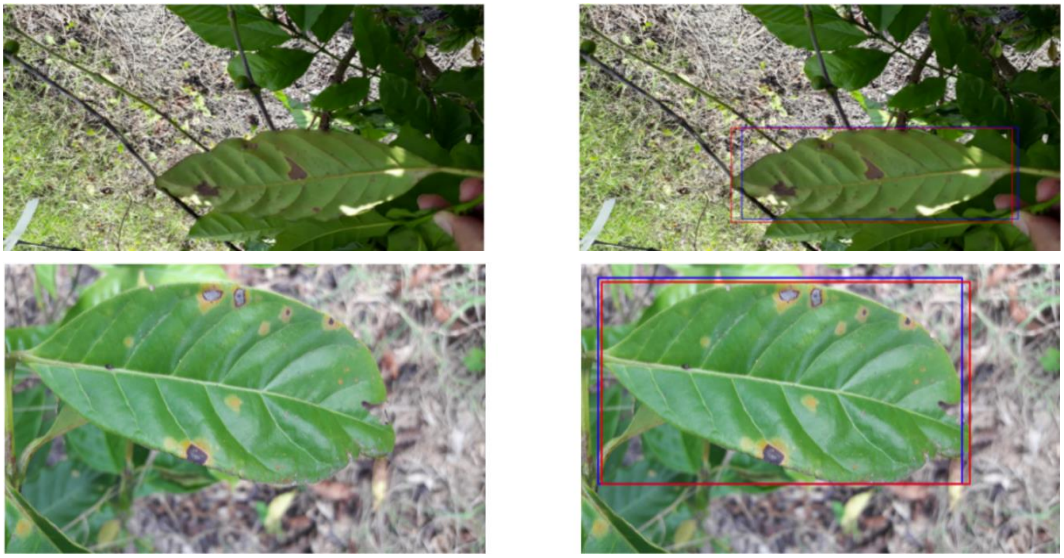
Ở giai đoạn đầu, mô hình Faster R-CNN với backbone MobileNetV3-Large được huấn luyện để dự đoán bounding box của lá cà phê trong ảnh. Bộ dữ liệu RoCoLe được sử dụng để huấn luyện, trong đó các annotation polygon được chuyển thành bounding box làm nhãn đầu ra. Mô hình được huấn luyện trong 5 epoch

với learning rate là 0.005, momentum 0.9, weight decay 0.0005 và batch size là 2. Hàm mất mát Smooth L1 Loss được sử dụng để tối ưu hóa hồi quy bounding box. Các bounding box dự đoán từ mô hình sau đó được dùng làm input prompt cho mô hình phân đoạn SAM2, giúp tạo ra mặt nạ lá một cách chính xác mà không cần huấn luyện lại mô hình SAM2.

Quá trình bên trên cho ra một mô hình dự đoán boundingbox, dùng làm input prompt cho mô hình phân đoạn SAM2, giúp tạo ra mặt nạ lá một cách chính xác mà không cần huấn luyện lại mô hình SAM2.

3.3.2. Kết quả

Trong phương pháp hai giai đoạn được đề xuất, ảnh trải qua quá trình xử lý ban đầu thông qua mô-đun hồi quy hộp giới hạn. Mô-đun này tạo ra tọa độ chính xác để cô lập các lá trong ảnh. Hình 3.1 minh họa một ví dụ về kết quả của giai đoạn xác định boundingbox.



Hình 3.1 Hình ảnh đầu vào có hộp giới hạn được tạo bởi bộ hồi quy hộp giới hạn

Sau 5 epoch, quá trình huấn luyện hộp giới hạn tạo ra một hộp giới hạn dự đoán (hộp màu đỏ) trong gần giống với hộp giới hạn thực tế hộp màu xanh) với Mean IoU 90.52% đây là mô hình đáng tin cậy để làm đầu vào cho SAM.

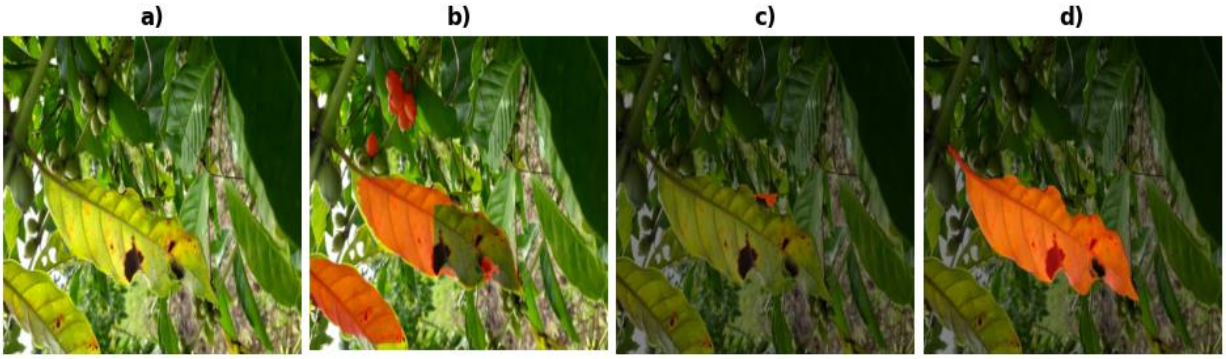
Dưới đây là bảng so sánh hiệu quả phân đoạn giữa ba cấu hình: Mask R-CNN, SAM2 với prompt điểm trung tâm, và SAM2 kết hợp với bounding box tự động do mô hình Faster R-CNN sinh ra.

Bảng 3.1 So sánh kết quả mean IoU của các cấu hình giai đoạn phân đoạn

Mô hình phân đoạn	Mean IoU
Mask RCNN	76.50%
SAM2 với tính năng nhắc nhở điểm trung tâm	87.77%
SAM2 với tính năng nhắc nhở tự động	92.87%

Kết quả cho thấy cấu hình đề xuất (SAM2 + bounding box tự động) đạt Mean IoU lên đến 92.87%, cao hơn rõ rệt so với SAM2 dùng prompt điểm trung tâm (87.77%) và Mask R-CNN (76.50%).

Dưới đây là hình ảnh cung cấp so sánh trực quan giữa các mặt nạ phân đoạn lá được tạo ra bởi Mask R-CNN, SAM2 với dấu nhắc điểm trung tâm và phương pháp được đề xuất trong nghiên cứu, SAM2 với bộ hồi quy hộp giới hạn.

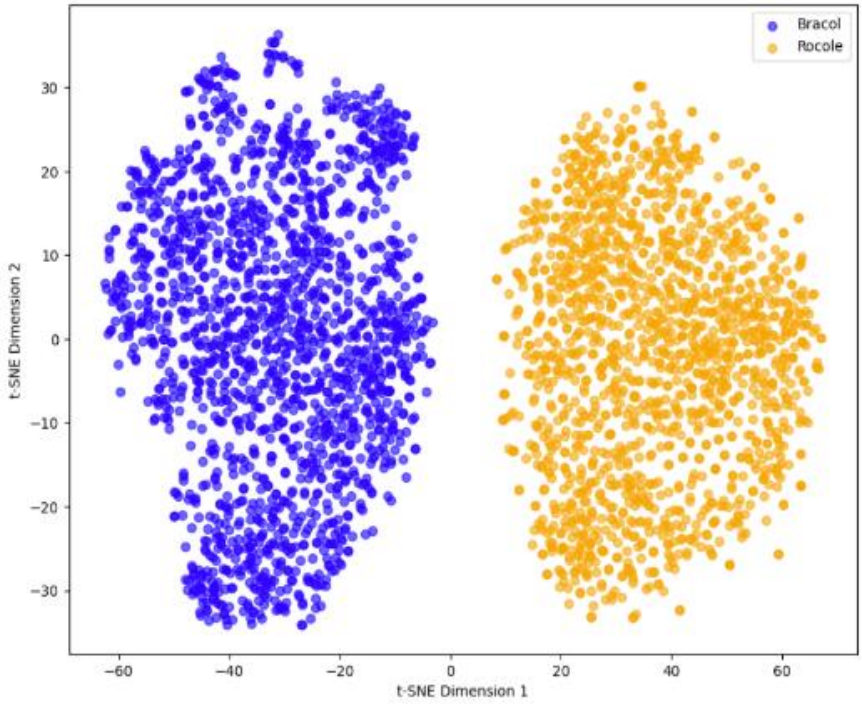


Hình 3.2 So sánh các mặt nạ phân đoạn lá.

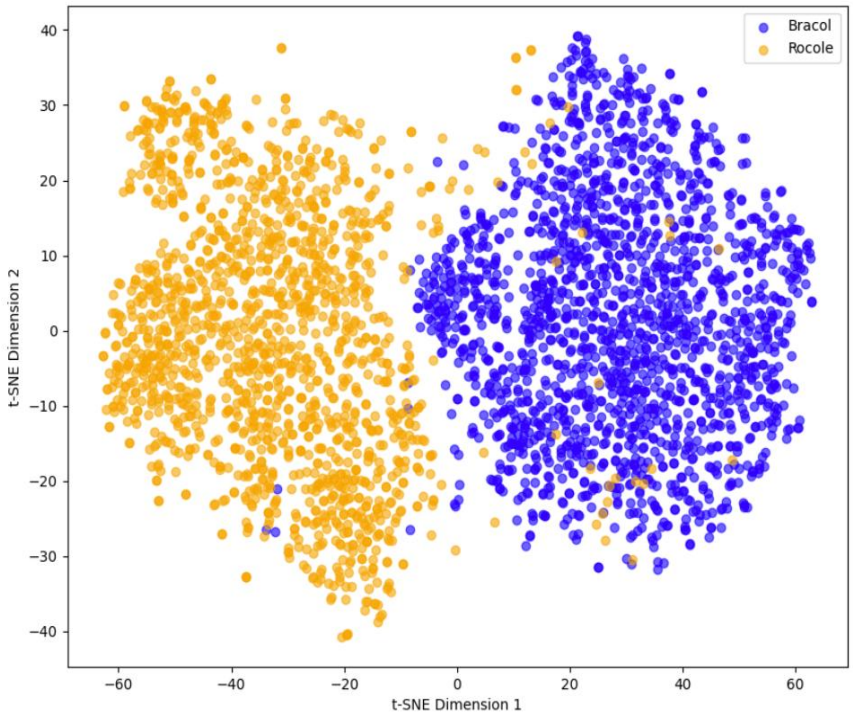
(a) Ảnh gốc, (b) Kết quả từ Mask R-CNN
(c) Kết quả phân đoạn từ SAM2 , (d) Kết quả phân đoạn từ SAM2 có bộ hồi quy bounding box

Các mặt nạ được tạo ra bởi SAM2, đặc biệt khi sử dụng bộ hồi quy hộp giới hạn, thể hiện độ trung thực vượt trội so với ranh giới lá thực tế. Độ chính xác được cải thiện này giúp giảm thiểu các lỗi phân đoạn, chẳng hạn như phân đoạn thiếu (khi một phần lá bị bỏ sót) hoặc phân đoạn thừa (khi bao gồm các vùng không phải lá).

Tiếp theo, kết quả của t-SNE đã được sử dụng để trực quan hóa các biểu diễn đặc trưng trước và sau khi áp dụng mô hình phân đoạn.



Hình 3.3 Trực quan hoá t-SNE của các biểu diễn đặc trưng hình ảnh trước khi áp dụng SAM 2



Hình 3.4 Trực quan hoá t-SNE của các biểu diễn đặc trưng hình ảnh sau khi áp dụng SAM 2

Trong biểu đồ t-SNE điểm màu cam biểu diễn ảnh từ tập dữ liệu RoCoLe, trong khi các điểm màu xanh lam tương ứng với ảnh từ tập dữ liệu BRACOL. Hình 3.3 cho thấy bộ dữ liệu RoCoLe và BRACOL trở nên gần nhau hơn trong không gian đặc trưng, chứng minh khả năng căn chỉnh hiệu quả của mô hình phân đoạn. Các cải tiến từ SAM2, đặc biệt trong ảnh có bối cảnh phức tạp, giúp giảm lỗi phân đoạn, cải thiện độ chính xác và nâng cao độ tin cậy cho các bước phân loại và phân tích sau đó.

3.4. HỌC TỰ GIÁM SÁT VÀ PHÂN LOẠI BỆNH

3.4.1. Dữ liệu và thông số huấn luyện

3.4.1.1. Dữ liệu

Giai đoạn này, bộ dữ liệu BRACOL được dùng để huấn luyện và đánh giá mô hình. Ảnh được chuẩn hóa về 224×224 pixel, chuyển sang không gian màu Lab, và áp dụng tăng cường dữ liệu gồm xoay ±70°, lật ngang – dọc, cùng điều chỉnh độ sáng và tương phản nhằm nâng cao khả năng khái quát.

Dữ liệu được chia theo tỷ lệ 80% huấn luyện và 20% validation, có stratify theo lớp để đảm bảo cân bằng giữa các loại bệnh.

3.4.1.2. Thiết lập huấn luyện

Tiếp theo, phương pháp học tự giám sát (self-supervised learning) được áp dụng để pre-train kiến trúc encoder–decoder dựa trên ResNet, với tác vụ giả định là tô màu ảnh (colorization). Ảnh đầu vào được chuyển sang thang xám, và mô hình học cách khôi phục lại màu sắc gốc dựa trên đặc trưng màu, kết cấu và hình dạng lá. Mô hình được huấn luyện trong 10 epoch với learning rate = 0.0001, batch size = 16 và hàm mất mát MSE, giúp tăng khả năng biểu diễn và giảm phụ thuộc vào dữ liệu có nhãn.

Sau giai đoạn pre-train, encoder được fine-tune cho nhiệm vụ phân loại bệnh sử dụng backbone ResNet50 trên bộ dữ liệu BRACOL. Quá trình huấn luyện diễn ra trong 15 epoch với CrossEntropy Loss, optimizer Adam và learning rate ban đầu 0.0001. Cơ chế ReduceLROnPlateau được dùng để điều chỉnh tốc độ học dựa trên hiệu suất của tập validation.

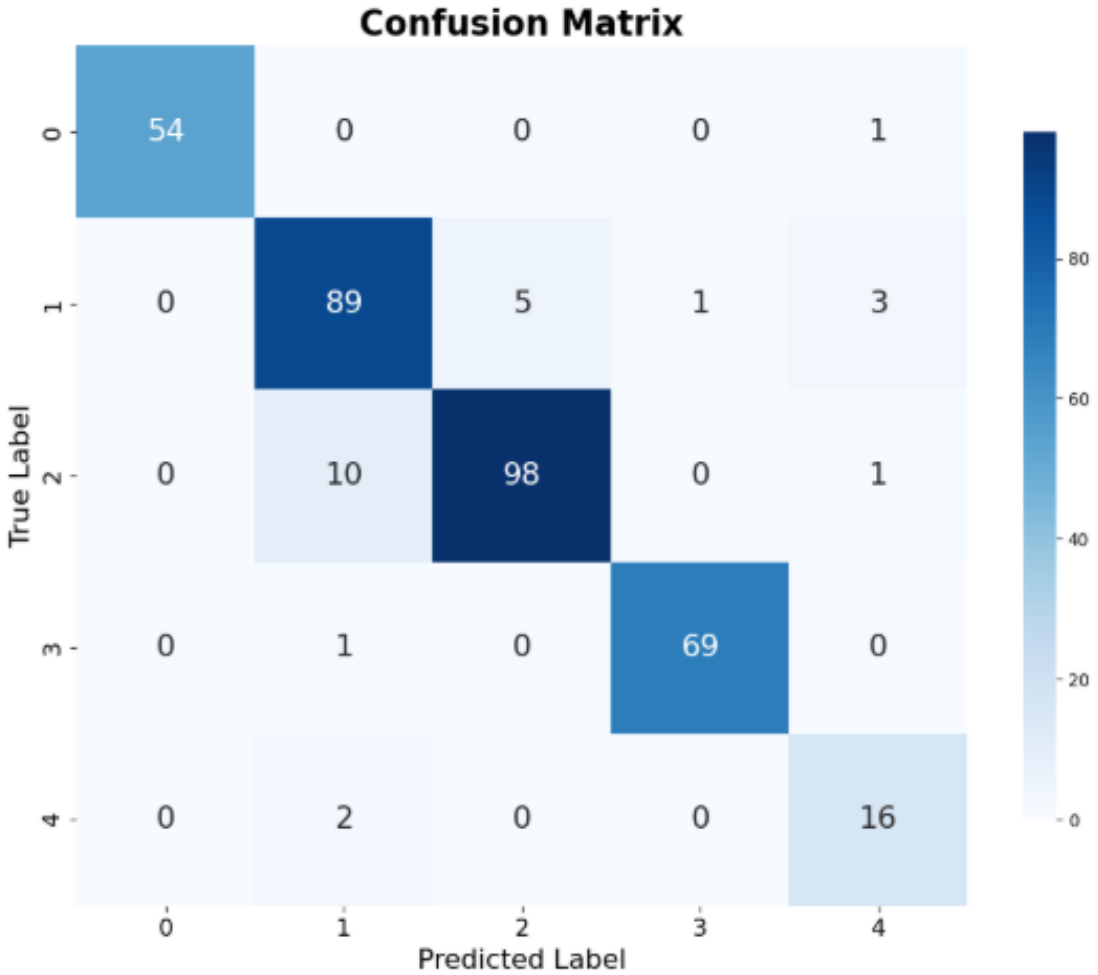
3.4.2. Kết quả

Bảng 3.2 trình bày kết quả so sánh giữa mô hình ResNet50 gốc và mô hình sau khi fine-tune.

Bảng 3.2 So sánh kết quả thử nghiệm các cấu hình giai đoạn phân loại

Cấu hình mô hình	Accuracy	F1-score
ResNet50	89.43%	89.24%%
ResNet50 + SLL (colorization)	93.14%	93.22%

Dưới đây là ma trận nhầm lẫn được đánh giá trên tập validation để hiển thị phân phối của nhãn thực so với nhãn dự đoán trên năm nhóm: khỏe mạnh, bệnh gỉ sắt, bệnh sâu đục lá, bệnh đốm nâu và bệnh đốm Cercospora. Các phần tử đường chéo biểu thị các phân loại chính xác, trong khi các phần tử ngoài đường chéo biểu thị các phân loại sai. Mô hình thể hiện độ chính xác cao, với hầu hết các dự đoán đều xác định đúng lớp bệnh.



Hình 3.6 Ma trận nhầm lẫn minh hoạ hiệu suất phân loại.

3.5. ĐÁNH GIÁ TOÀN BỘ HỆ THỐNG

Tiếp theo là quá trình đánh giá hiệu quả của hệ thống hai giai đoạn được đề xuất, tích hợp phân đoạn SAM2 tinh chỉnh và bộ mã hóa ResNet50 được đào tạo trước thông qua SSL trên tác vụ tô màu.

3.5.1. Dữ liệu và thông số huấn luyện

Tập dữ liệu được sử dụng để đánh giá tổng thể hiệu quả của hệ thống là 20% dữ liệu kiểm thử được tách ra từ bộ BRACOL, tương ứng với 350 ảnh lá cà phê. Tập dữ liệu này bao gồm đầy đủ các lớp bệnh như khỏe mạnh, sâu đục lá, gỉ sắt, phoma, cercospora và được giữ nguyên tỉ lệ phân bố) giữa các lớp, đảm bảo tính đại diện cho toàn bộ tập dữ liệu gốc.

Hình 3.7 Một số hình ảnh lá trong bộ BRACOL được ghép nền

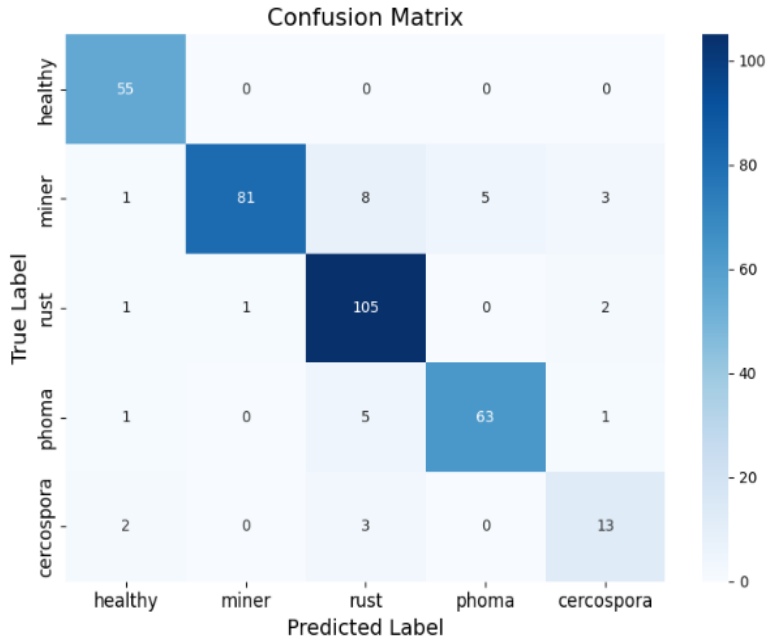
3.5.2. Kết quả

Hiệu suất được so sánh với các cấu hình cơ sở khác: (1) ResNet50 được đào tạo hoàn toàn trên tập dữ liệu mục tiêu; (2) mô hình ResNet50 được đào tạo trước trên ImageNet và sau đó được tinh chỉnh trên tập dữ liệu mục tiêu; (3) cấu hình kết hợp SAM với ResNet50 được đào tạo trên tập dữ liệu mục tiêu và; (4) cấu hình với ResNet50 được đào tạo trước bằng pretext-SSL.

Bảng 3.3 So sánh kết quả thử nghiệm các cấu hình toàn hệ thống

Cấu hình mô hình	Accuracy	F1-score
Resnet50 (Huấn luyện truyền thống)	66.57%	59.06%
SSL + ResNet50	66.00%	58.46%
SAM2 + ResNet50	89.43%	89.38%
SAM2 + SSL + ResNet50	**90.57%**	**90.54%**

Như được thể hiện trong Bảng 3.3, việc tích hợp SAM2 với phân đoạn, đào tạo trước với tô màu bằng SSL và bộ mã hóa ResNet50 cho kết quả chính xác nhất với độ chính xác là 90.57% và F1 là 90.54%. Hệ thống hai giai đoạn được đề xuất, kết hợp cả hai kỹ thuật, đã cải thiện đáng kể hiệu suất tổng thể.



Hình 3.8 Ma trận nhầm lẫn minh hoạ hiệu suất phân loại toàn hệ thống

Hình 3.8 là trận nhầm lẫn để hiển thị phân phối của nhãn thực so với nhãn dự đoán trên năm nhóm: khỏe mạnh, sâu đục lá, gỉ sắt, phoma và cercospora. Các phần tử đường chéo biểu thị các phân loại chính xác, trong khi các phần tử ngoài đường chéo biểu thị các phân loại sai. Mô hình thể hiện độ chính xác cao, với hầu hết các dự đoán đều xác định đúng nhóm bệnh.

3.6. PHÂN TÍCH CÁC YẾU TỐ ẢNH HƯỞNG, THẢO LUẬN HẠN CHẾ

Hệ thống đề xuất cho thấy hiệu quả vượt trội so với các mô hình truyền thống. Việc áp dụng học tự giám sát (SSL) thông qua tác vụ tô màu đạt hiệu quả cao nhất khi dữ liệu đã được phân đoạn và làm sạch. Tuy nhiên, việc chuyển đổi mẫu ảnh đa nhãn thành đơn nhãn trong quá trình gán nhãn làm mất thông tin về sự đồng xuất hiện của bệnh, ảnh hưởng đến khả năng phản ánh thực tế. Ngoài ra, quá trình sử dụng SAM2 để phân đoạn và ghép nền có thể làm sai lệch thông tin hình ảnh, khiến độ chính xác của mô hình trên dữ liệu thực tế giảm so với mong đợi.

3.7. KẾT LUẬN CHƯƠNG

Chương này đã trình bày toàn bộ kết quả hệ thống đề xuất. Kết quả đạt được là:

- Giai đoạn phân đoạn: Đạt Mean IoU 92.87%, vượt trội hơn Mask R-CNN và SAM2 thông thường.
- Giai đoạn phân loại (ResNet50 + SSL): Đạt độ chính xác 93.14%, cải thiện rõ rệt so với huấn luyện truyền thống.
- Hệ thống tích hợp hoàn chỉnh (SAM2 + SSL + ResNet50): Đạt Accuracy 90.57% và F1-score 90.54%, chứng minh hiệu quả vượt trội và ổn định.

Tuy nhiên, mô hình vẫn chịu ảnh hưởng từ việc tinh chỉnh các nhãn bệnh đa lớp thành đơn lớp và quá trình ghép nền có thể làm sai lệch thông tin. Những vấn đề này sẽ được khắc phục trong các hướng nghiên cứu tiếp theo nhằm hoàn thiện hệ thống cho ứng dụng thực tế.

KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO

1. TÓM TẮT KẾT QUẢ ĐẠT ĐƯỢC

Nghiên cứu đã đề xuất mô hình phân loại hai giai đoạn, kết hợp giữa phương pháp phân đoạn bằng SAM2 và cơ chế tô màu tự giám sát, nhằm tăng cường khả năng nhận diện đặc trưng bệnh trên lá cà phê. Kết quả thực nghiệm cho thấy mô hình đạt độ chính xác tổng thể 90.57% và điểm F1 90.54%, vượt trội so với các mô hình cơ sở. Điều này khẳng định tính hiệu quả, độ tin cậy và tiềm năng ứng dụng của mô hình trong thực tiễn phát hiện bệnh cây trồng.

2. ĐÁNH GIÁ MỨC ĐỘ ĐÁP ỨNG MỤC TIÊU BAN ĐẦU

Nghiên cứu đã hoàn thành mục tiêu đề ra là đề xuất được mô hình học sâu hiệu quả cho bài toán phân loại bệnh trên lá cà phê với độ chính xác cao.

3. HẠN CHẾ CỦA NGHIÊN CỨU

Mặc dù đạt được những kết quả khả quan, nghiên cứu vẫn tồn tại một số giới hạn nhất định. Bộ dữ liệu được sử dụng còn chưa phong phú về quy mô và tính đa dạng, chủ yếu bao gồm một số loại bệnh phổ biến, trong khi trên thực tế, cây cà phê có thể mắc nhiều bệnh khác với biểu hiện phức tạp và đa dạng hơn. Bên cạnh đó, tập dữ liệu kiểm thử mới dừng lại ở mức mô phỏng bối cảnh tự nhiên, chưa được đánh giá trực tiếp trong điều kiện thực địa, nên hiệu suất của mô hình có thể có sai lệch nhất định khi áp dụng vào môi trường thực tế.

4. ĐỀ XUẤT HƯỚNG PHÁT TRIỂN TRONG TƯƠNG LAI

Trong giai đoạn tiếp theo, nghiên cứu sẽ được mở rộng thông qua việc thu thập và xây dựng bộ dữ liệu có quy mô lớn hơn, được lấy từ nhiều khu vực trồng cà phê khác nhau, dưới các điều kiện ánh sáng, thời tiết và góc chụp đa dạng. Điều này nhằm nâng cao khả năng khái quát hóa, độ ổn định và tính thích ứng của mô hình.

Bên cạnh đó, hướng nghiên cứu sẽ tập trung vào việc tối ưu hóa kiến trúc mô hình để cải thiện hiệu suất nhận diện, đồng thời mở rộng phạm vi phân loại đối với nhiều loại bệnh khác nhau trên cây cà phê. Mô hình sau khi hoàn thiện sẽ được tích hợp vào các ứng dụng di động hoặc hệ thống giám sát nông nghiệp thông minh, góp phần thúc đẩy chuyển đổi số trong lĩnh vực nông nghiệp bền vững và hỗ trợ người nông dân trong công tác quản lý, phòng ngừa sâu bệnh một cách hiệu quả hơn.

CÔNG TRÌNH KHOA HỌC ĐÃ CÔNG BỐ

N. H. N. Minh, D. C. Cuong, P. V. N. Vinh, N. T. C. Khang, D. Q. Thang, L. D. Phuc, N. T. P. Thao, and N. T. Binh, "Self-supervised colorization driven two-stage coffee leaf disease detection," in *Proc. 14th Conf. Information Technology and Its Applications (CITA)*, 2025.

DANH MỤC TÀI LIỆU THAM KHẢO

[1] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2016.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[6] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment Anything," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2023, pp. 4015–4026.

[7] L. X. B. Sorte, C. T. Ferraz, F. Fambrini, R. dos R. Goulart, and J. H. Saito, "Coffee leaf disease recognition based on deep learning and texture attributes," *Procedia Comput. Sci.*, vol. 159, pp. 135–144, 2019.

[8] E. B. Paulos and M. M. Woldeyohannis, "Detection and classification of coffee leaf disease using deep learning," in *Proc. Int. Conf. Information and Communication Technology for Development for Africa (ICT4DA)*, 2022, pp. 1–6.

[9] E. Hitimana, O. J. Sinayobye, J. C. Ufitinema, J. Mukamugema, P. Rwibasira, T. Murangira, E. Masabo, L. C. Chepkwony, M. C. A. Kamikazi, J. A. U. Uwera, *et al.*, "An intelligent system-based coffee plant leaf disease recognition using deep learning techniques on Rwandan Arabica dataset," *Technologies*, vol. 11, no. 5, p. 116, 2023.

[10] F. Martinez, H. Montiel, and F. Martinez, "A machine learning model for the diagnosis of coffee diseases," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 4, 2022.

[11] J. V. Y. B. Yamashita and J. P. R. R. Leite, "Coffee disease classification at the edge using deep learning," *Smart Agric. Technol.*, vol. 4, p. 100183, 2023.

[12] A. P. Marcos, N. L. S. Rodovalho, and A. R. Backes, "Coffee leaf rust detection using genetic algorithm," in *Proc. XV Workshop de Visão Computacional (WVC)*, 2019, pp. 16–20.

[13] A. P. Marcos, N. L. S. Rodovalho, and A. R. Backes, "Coffee leaf rust detection using convolutional neural network," in *Proc. XV Workshop de Visão Computacional (WVC)*, 2019, pp. 38–42.

[14] L. M. Tassis, J. E. T. de Souza, and R. A. Krohling, “A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images,” *Comput. Electron. Agric.*, vol. 186, p. 106191, 2021.

[15] M. Faisal, J.-S. Leu, and J. T. Darmawan, “Model selection of hybrid feature fusion for coffee leaf disease classification,” *IEEE Access*, vol. 11, pp. 62281–62291, 2023.

[16] Parraga-Alava, K. Cusme, A. Loor, and E. Santander, “RoCoLe: A robusta coffee leaf images dataset,” *Mendeley Data*, vol. 2, 2019, doi: 10.17632/c5yvn32dzg.2.

[17] R. A. Krohling, G. J. M. Esgario, and J. A. Ventura, “BRACOL: A Brazilian Arabica coffee leaf images dataset to identification and quantification of coffee diseases and pests,” *Mendeley Data*, vol. 1, 2019, doi: 10.17632/yy2k5y8mxg.1.