

NHẬN DẠNG HÌNH ẢNH NẤM DỰA TRÊN MÔ HÌNH HỌC SÂU EFFICIENTNETV2-S

Nguyễn Quốc Nhựt ⁽¹⁾, Dương Thị Kim Chi ^(2,*)

(1) Viện Công nghệ Số, Trường Đại học Thủ Dầu Một

Email: 2124802010364@student.tdmu.edu.vn

(2) Viện Công nghệ Số, Trường Đại học Thủ Dầu Một

Email: chidtk@tdmu.edu.vn

Ngày nhận bài: 24/5/2026; Chấp nhận đăng: 27/6/2025

DOI: 10.37550/tdmu.EJS/V2026.03.763

Tóm tắt

Nhu cầu tự động phân tích và xử lý hình ảnh đã trở thành một yêu cầu quan trọng trong nhiều lĩnh vực như y tế, nông nghiệp hiện đại. Tuy nhiên, việc phân loại hình ảnh bằng phương pháp truyền thống thường gặp nhiều khó khăn do sự đa dạng về màu sắc, kích thước, góc chụp và điều kiện ánh sáng của dữ liệu ảnh. Trong nghiên cứu này, chúng tôi đề xuất mô hình học sâu EfficientNetV2-S kết hợp với kỹ thuật Grad-CAM nhằm tăng cường khả năng trực quan hóa các vùng đặc trưng quan trọng trên ảnh, qua đó cải thiện hiệu quả nhận dạng. Phương pháp đề xuất áp dụng trên bộ dữ liệu hình ảnh nấm được công bố công khai trên Kaggle. Kết quả cho thấy mô hình đề xuất đạt hiệu năng vượt trội trên nhiều chỉ số đánh giá quan trọng, bao gồm độ chính xác, độ nhạy (Recall), độ đặc hiệu (Specificity), F1-score và ROC, với các giá trị đều đạt mức 0,98. Những kết quả này khẳng định tính khả thi cũng như tiềm năng ứng dụng thực tiễn của mô hình trong bài toán nhận dạng hình ảnh nấm. Trong tương lai, mô hình từ nghiên cứu này có thể được phát triển thành một công cụ hỗ trợ nhận dạng nấm nhanh chóng và chính xác hơn, qua đó góp phần hạn chế nguy cơ ngộ độc nấm trong thực tế.

Từ khóa: Artificial Intelligence, Edible Mushrooms, EfficientNetV2-S, phân loại hình ảnh.

Abstract

IMAGE RECOGNITION OF FUNGI BASED ON THE EFFICIENTNETV2-S DEEP LEARNING MODEL

The need for automated image analysis and processing has become a crucial requirement in many fields such as healthcare and modern agriculture. However, traditional image classification methods often face difficulties due to the diversity of colors, sizes, angles, and lighting conditions in image data. In this study, we propose the EfficientNetV2-S deep learning model combined with Grad-CAM techniques to enhance the visualization of important feature regions in images, thereby improving recognition efficiency. The proposed method is applied to a publicly available dataset of fungal images on Kaggle. The results show that the proposed model achieves superior performance across several important evaluation metrics, including accuracy, sensitivity (Recall), specificity, F1-score, and ROC, with all values reaching 0.98. These results confirm the feasibility and practical application potential of the model in the problem of fungal image recognition. In the future, the model from this research could be developed into a tool to help identify mushrooms more quickly and accurately, thereby contributing to reducing the risk of mushroom poisoning in practice.

Keywords: Artificial Intelligence, Edible Mushrooms, EfficientNetV2-S, image classification.

1. Đặt vấn đề

Nhu cầu tự động phân tích và xử lý hình ảnh đang ngày càng trở nên quan trọng trong nhiều lĩnh vực như y tế, nông nghiệp thông minh và giám sát môi trường. Tuy nhiên, việc phân loại hình ảnh bằng các phương pháp truyền thống vẫn gặp nhiều thách thức do dữ liệu ảnh có sự biến đổi lớn về màu sắc, kích thước, góc chụp và điều kiện ánh sáng. Đặc biệt, trong các bài toán nhận dạng hình ảnh thực tế như hình ảnh y khoa hay thực vật trong tự nhiên, độ phức tạp của dữ liệu càng làm giảm hiệu quả của các phương pháp nhận dạng thông thường.

Một trong những bài toán tiêu biểu là nhận dạng nấm hoang dã, vốn có ý nghĩa quan trọng trong lĩnh vực sức khỏe cộng đồng và an toàn thực phẩm. Theo thống kê, mỗi năm trên thế giới ghi nhận hàng nghìn trường hợp ngộ độc nấm, trong đó khoảng 80% liên quan đến việc định danh không chính xác các loài nấm (Lu et al., 2020). Các loài nấm độc có thể gây ra nhiều hội chứng lâm sàng nghiêm trọng như suy gan cấp, suy thận cấp, tiêu cơ vân, viêm dạ dày ruột, rối loạn thần kinh, tan máu và viêm da nhày cảm. Cần có một giải pháp hiện đại hỗ trợ phân biệt hình ảnh giúp nâng cao hiệu quả phân biệt các loại nấm bằng hình ảnh để hạn chế các tình trạng ngộ độc nấm.

Trong thực tiễn, các chuyên gia thường sử dụng phương pháp truyền thống để phân loại và định danh nấm dựa trên các đặc điểm hình thái như mũ, thịt nấm, phiến nấm và cuống; một số loài còn có thêm vòng và bao gốc. Nhìn chung, các đặc điểm của mũ nấm như hình dạng, kích thước, màu sắc và lớp phủ bề mặt là những căn cứ quan trọng trong quá trình nhận dạng (Hebert et al., 2003) (Somervuo et al., 2016). Các dấu hiệu hình thái này đóng vai trò thiết yếu trong việc phân biệt các loài nấm khác nhau (Duong et al., 2023). Tuy nhiên, do thiếu kiến thức chuyên môn, kỹ năng nhận diện cũng như sự hướng dẫn từ các chuyên gia về nấm, nhiều người dân bản địa vẫn phải đối mặt với nguy cơ sử dụng nhầm nấm độc, đặc biệt khi các loài nấm này có hình thái tương đồng với nấm ăn được.

Trước thực trạng đó, nhiều nghiên cứu đã tập trung khai thác trí tuệ nhân tạo nhằm hỗ trợ nhận diện và phân loại nấm, qua đó góp phần hạn chế nguy cơ ngộ độc. Một trong những hướng tiếp cận tiêu biểu là xác định tên loài dựa trên dữ liệu hình ảnh của nấm. Phương pháp này tận dụng nguồn dữ liệu ảnh được xây dựng từ các nghiên cứu về phân loại hình ảnh, từ đó xây dựng quy trình tự động phân loại hình ảnh tự động trong các lĩnh vực khác nhau như an toàn thực phẩm, y tế, giám sát môi trường.

2. Tổng quan tình hình nghiên cứu

Có nhiều hướng tiếp cận khác đang ngày càng được quan tâm là tự động học và trích xuất đặc trưng trực tiếp từ ảnh nấm bằng các mô hình học sâu (Dong et al., 2020). Kỹ thuật học sâu (Deep Learning, DL) đã được đề xuất như một giải pháp hiệu quả cho bài toán trích xuất đặc trưng tự động và phân loại hình ảnh, tiêu biểu là mạng nơ-ron tích chập (Convolutional Neural Network, CNN) (Panchbhai, K.G. et al., 2024). Trong bối cảnh đó, Bảng 1 trình bày tổng hợp một số nghiên cứu tiêu biểu về ứng dụng học sâu trong nhận dạng loài nấm độc và nấm ăn được.

Bảng 1. Thông tin về các nghiên cứu sử dụng học sâu dự đoán tên loài nấm dựa trên dữ liệu hình ảnh.

Tác giả	Mục tiêu	Công nghệ	Hiệu năng (Accuracy)
(Preechasuk, et al., 2019)	Phân loại 45 loại nấm, trong đó 35 loại là nấm ăn được và 10 loại còn lại là nấm độc	Mô hình mạng neuron tích chập CNN	0,78
(Devika, G., and Asha Gowda Karegowda, 2021)	Phân loại nấm độc và nấm ăn được	Mô hình mạng neuron tích chập sâu (DCNN)	0,93
(Wang, Boyuan, 2022)	Phân loại 11 loại nấm khác nhau	Mô hình kiến trúc học sâu kết hợp với Sự chuyển đổi	0,95
(Zhao, Kexin, et al., 2023)	Phân loại trực quan của nấm đại	Mô hình học kết hợp tích chập và Transformer	0,96

Nguồn: tác giả, 2026

Nhìn chung, các phương pháp được đề xuất trong các nghiên cứu trước đây đều đạt hiệu năng dự đoán khả quan và có tiềm năng triển khai trên các thiết bị di động. Điều này cho thấy định danh loài nấm dựa trên dữ liệu hình ảnh là một hướng tiếp cận hiệu quả, đặc biệt có ý nghĩa thực tiễn đối với người dân trong quá trình thu hái nấm, góp phần hỗ trợ nhận diện các loài nấm độc và giảm thiểu nguy cơ ngộ độc. Bên cạnh các phương pháp phân loại trực tiếp, một hướng tiếp cận khác cũng đang nhận được sự quan tâm là chú thích hình ảnh (image captioning). Phương pháp này tập trung khai thác mối quan hệ ngữ nghĩa và không gian giữa các đối tượng xuất hiện trong ảnh nhằm nâng cao chất lượng mô tả hình ảnh. Nhờ đó, hệ thống không chỉ hỗ trợ nhận diện đối tượng mà còn cung cấp thêm thông tin ngữ cảnh, giúp người dùng đưa ra quyết định chính xác hơn trong quá trình phân biệt và đánh giá loài nấm.

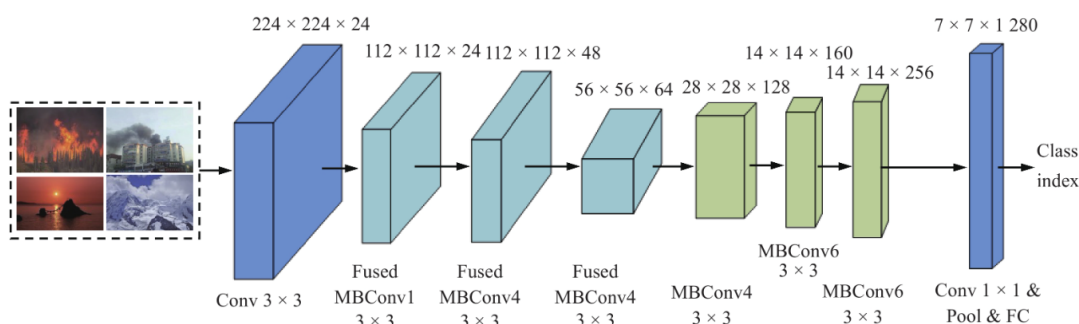
Bài báo này áp dụng các giải pháp học sâu kết hợp với cơ chế tự chú ý để cung cấp thông tin bổ sung cho các mẫu nấm được dự đoán. Mặc dù tập dữ liệu còn hạn chế gồm 2.820 hình ảnh nấm lớn thuộc 9 chi nấm, mô hình đề xuất đạt được độ chính xác cao và có thể được huấn luyện trong thời gian ngắn. Mô hình ViT được cải tiến của chúng tôi đặc biệt phù hợp cho việc phân loại nấm trong các bối cảnh hạn chế tài nguyên. Cụ thể, nghiên cứu này sử dụng EfficientNetV2-S và ViT-Large-Patch16-224 để phân loại 9 chi nấm. Sự tích hợp này tăng cường độ nhạy của mô hình đối với các đặc điểm hình ảnh tinh tế trong khi vẫn duy trì thiết kế nhẹ phù hợp cho các ứng dụng thời gian thực và điện toán biên. Hơn nữa, mô hình được huấn luyện trên một tập dữ liệu được tăng cường nghiêm ngặt để đảm bảo tính mạnh mẽ chống lại các biến thể về ánh sáng, hướng và nhiễu - các yếu tố thường gặp trong môi trường thực địa. Trong quy trình tính toán được đề xuất từ nghiên cứu này, bài viết thực hiện dựa trên 4 giai đoạn chính: (i) Thu thập dữ liệu hình ảnh nấm từ thực địa; (ii) Khảo sát và xử lý các thuộc tính hình ảnh, chuẩn hóa và làm giàu hình ảnh. (iii) Xây dựng bộ phân loại chi nấm dựa trên thuật toán EfficientNetV2_S; (iv) đánh giá mô hình đề xuất.

3. Phương pháp nghiên cứu

3.1. Kiến trúc của mô hình EfficientNet

EfficientNet, được Mingxing Tan và Quoc V. Le đề xuất vào năm 2019, là một kiến trúc mạng neuron tích chập (CNN) đồng thời cũng là một khung mở rộng mô hình

thống nhất theo ba chiều gồm độ sâu, độ rộng và độ phân giải đầu vào. Khác với các phương pháp mở rộng truyền thống thường chỉ tập trung vào một chiều duy nhất, vốn có thể cải thiện độ chính xác nhưng kéo theo chi phí tính toán lớn, EfficientNet hướng đến việc tối ưu hóa đồng thời sự cân bằng giữa kích thước mô hình, độ chính xác và chi phí tính toán. Cụ thể, mô hình áp dụng cơ chế mở rộng kết hợp thông qua một hệ số điều khiển duy nhất ϕ , cùng với các hệ số tỷ lệ α , β và γ để phân bổ hợp lý tài nguyên cho từng chiều mở rộng. Nhờ chiến lược này, EfficientNet đạt được hiệu quả Pareto vượt trội, tức là có thể mang lại độ chính xác cao hơn trong khi vẫn duy trì chi phí tính toán thấp hơn so với các phương pháp mở rộng theo một chiều riêng lẻ. Thông qua quá trình phân tích và loại trừ có hệ thống các chiến lược chỉ mở rộng theo độ sâu, độ rộng hoặc độ phân giải, các tác giả đã chỉ ra rằng, trong cùng một giới hạn tính toán, việc mở rộng đồng thời và cân đối cả ba chiều là giải pháp đem lại sự đánh đổi tối ưu giữa hiệu năng và hiệu quả tính toán. Nói cách khác, chiến lược mở rộng kết hợp của EfficientNet cho phép phân bổ tài nguyên một cách hài hòa giữa độ sâu, độ rộng và độ phân giải đầu vào thông qua một cơ chế điều khiển thống nhất. Sơ đồ của cơ chế này được minh họa trong Hình 1.

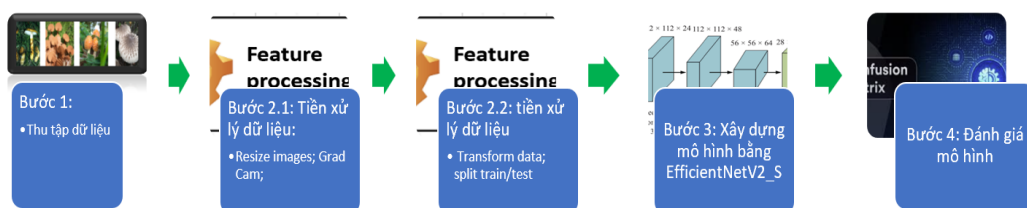


Hình 1. Cấu trúc mô hình EfficientNetV2_S

Nguồn: Tan, Mingxing, and Quoc Le., 2019

3.2. Tổng quan về kiến trúc đề xuất

Trong phần này, bài báo trình bày khung quy trình tổng quát của mô hình dự đoán tên chi nấm. Cụ thể, toàn bộ quy trình của phương pháp đề xuất được minh họa trong Hình 2



Hình 2. Khung quy trình tổng quát của mô hình dự đoán tên chi nấm lớn

Nguồn: tác giả, 2026

Bước 1: Thu thập dữ liệu

Cụ thể, toàn bộ quy trình của phương pháp đề xuất được minh họa trong Hình 6. Bộ dữ liệu được sử dụng trong nghiên cứu được tham khảo từ Kaggle (Kaggle, 2018), bao gồm cả nấm độc và nấm ăn được, với tổng cộng 2.820 hình ảnh được phân thành 9 lớp. Các hình ảnh nấm trong cơ sở dữ liệu được minh họa như Hình 3.

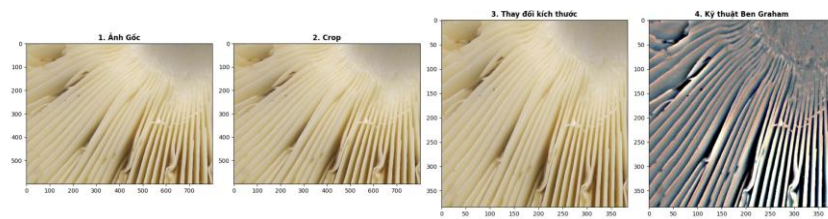


Hình 3. Hình ảnh các chi nấm lớn trong cơ sở dữ liệu huấn luyện

Nguồn: tác giả, 2026

Bước 2.1: Tiền xử lý dữ liệu hình ảnh

Hình ảnh nấm điều chỉnh kích thước, áp dụng biến đổi làm nổi bật các mô cạnh được mô tả như hình 4.

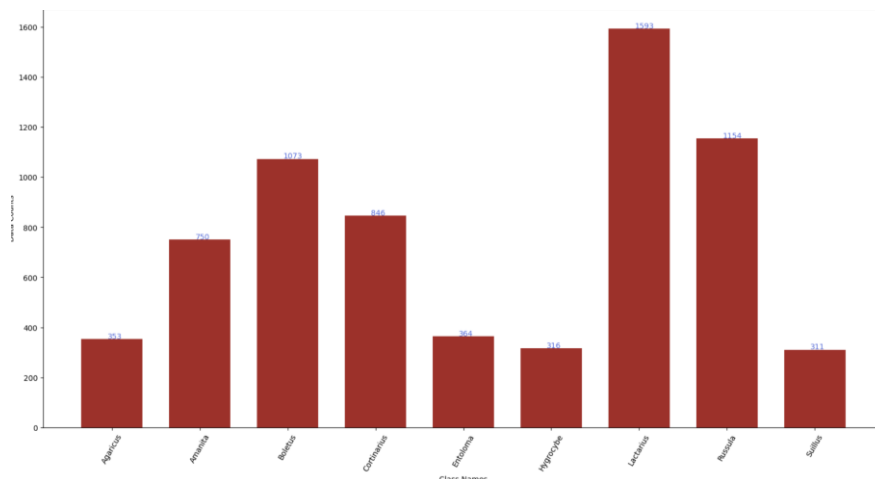


Hình 4. Minh họa các kỹ thuật dùng trong tiền xử lý dữ liệu

Nguồn: tác giả, 2026

Bước 2.2: Biến đổi dữ liệu ảnh, làm giàu ảnh và chia dữ liệu:

Sau giai đoạn 2.1, tập dữ liệu ảnh đã chuẩn hóa được chia theo tỷ lệ 80:10:10, gồm 80% huấn luyện, 10% validation và 10% testing. Phân bố dữ liệu của các lớp được mô tả như Hình 4

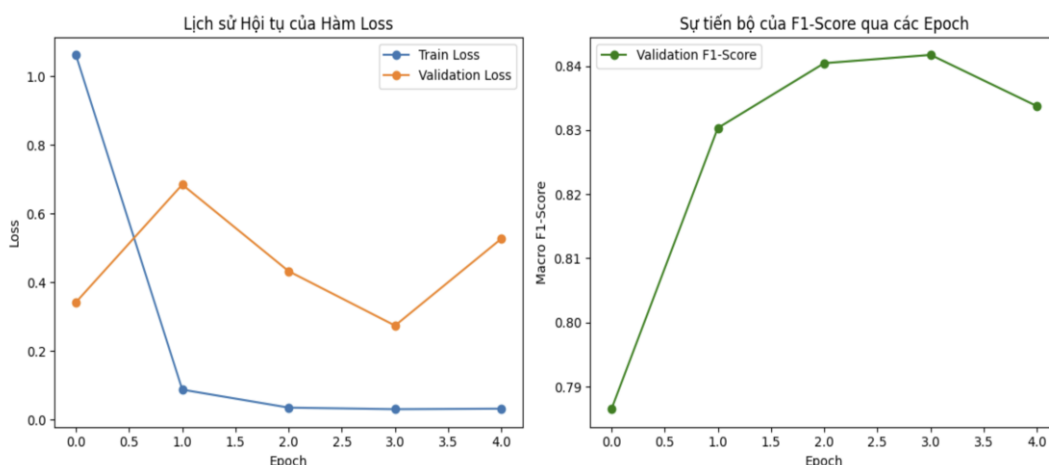


Hình 5. Số mẫu dữ liệu của các chi nấm trong bộ dữ liệu

Nguồn: tác giả, 2026

Bước 3: Xây dựng mô hình học sâu dự đoán tên chi nấm:

Khởi tạo mô hình EfficientNetV2-S. Để giải quyết sự mất cân bằng dữ liệu, bài viết sử dụng Focal Loss thay vì Cross-Entropy tiêu chuẩn. Hàm này ép mô hình tập trung học các mẫu khó hoặc các lớp có ít mẫu bằng cách điều chỉnh trọng số thông qua tham số γ . Bộ tối ưu hóa AdamW được sử dụng với tốc độ học là 1×10^{-4}



Hình 6. Kết quả mô hình huấn luyện của các chỉ năm trong bộ dữ liệu

Nguồn: tác giả, 2026

Bước 4: Đánh giá hiệu năng mô hình

Để đánh giá hiệu quả của mô hình đề xuất, nghiên cứu sử dụng các chỉ số đánh giá phổ biến trong lĩnh vực học máy [8,14], nhằm phản ánh khả năng phân loại và độ tin cậy của mô hình trên tập dữ liệu thực nghiệm.

Accuracy được sử dụng để đo lường tỷ lệ số mẫu được mô hình dự đoán chính xác trên tổng số mẫu trong tập dữ liệu, và được xác định theo công thức:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Trong đó TP - True Positive: chẩn đoán đúng, số mẫu Positive được phân loại đúng

TN - True Negative: loại trừ đúng

FP - False Positive: chẩn đoán sai, số mẫu Negative bị phân loại nhầm thành Positive

FN - False negative: bỏ sót, loại trừ nhầm

Sensitivity (độ nhạy), còn được gọi là Recall hoặc True Positive Rate (TPR), là chỉ số đo lường khả năng của mô hình trong việc nhận diện đúng các mẫu thuộc lớp Positive. Chỉ số này được tính bằng tỷ lệ giữa số trường hợp được dự đoán Positive đúng (True Positive) trên tổng số các trường hợp Positive thực tế. Sensitivity phản ánh mức độ mô hình có thể phát hiện chính xác các đối tượng cần nhận diện, và được xác định theo công thức:

$$TPR = Recall = \frac{TP}{TP + FN}$$

Precision là chỉ số đánh giá khả năng dự đoán chính xác các mẫu thuộc lớp Positive của mô hình. Trong bối cảnh bộ dữ liệu có sự mất cân bằng giữa số lượng mẫu của các lớp (data imbalance), Precision trở thành một chỉ số quan trọng nhằm đo lường mức độ tin cậy của các dự đoán Positive. Chỉ số này được xác định bằng tỷ lệ giữa số

mẫu được dự đoán Positive đúng (True Positive) trên tổng số mẫu được mô hình dự đoán là Positive. Precision được tính theo công thức:

$$Precision = \frac{TP}{TP + FP}$$

F1 score là chỉ số được xác định dựa trên trung bình điều hòa (harmonic mean) giữa hai đại lượng Precision và Recall. Chỉ số này phản ánh sự cân bằng giữa khả năng dự đoán chính xác các mẫu Positive và khả năng phát hiện đầy đủ các mẫu Positive của mô hình. F1-score đặc biệt hữu ích trong các bài toán có dữ liệu mất cân bằng, khi cần đánh giá đồng thời cả Precision và Recall thay vì chỉ sử dụng riêng lẻ từng chỉ số. F1-score được tính theo công thức:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

ROC (Receiver Operating Characteristics) là đường cong được sử dụng để đánh giá hiệu suất phân loại của mô hình. Đường cong ROC biểu diễn mối quan hệ giữa True Positive Rate (TPR) và False Positive Rate (FPR) tại nhiều ngưỡng phân loại khác nhau. Mỗi điểm trên đường cong tương ứng với một cặp giá trị (TPR, FPR) tại một ngưỡng xác định, qua đó phản ánh khả năng phân biệt giữa các lớp của mô hình. Dựa trên đường cong ROC, chỉ số AUC (Area Under the Curve) được tính bằng diện tích nằm dưới đường cong ROC. Chỉ số AUC được sử dụng để đánh giá tổng quát hiệu suất phân loại của mô hình; giá trị AUC càng gần 1 cho thấy mô hình có khả năng phân biệt giữa các lớp càng tốt, trong khi giá trị gần 0,5 phản ánh hiệu suất phân loại gần tương đương với dự đoán ngẫu nhiên.

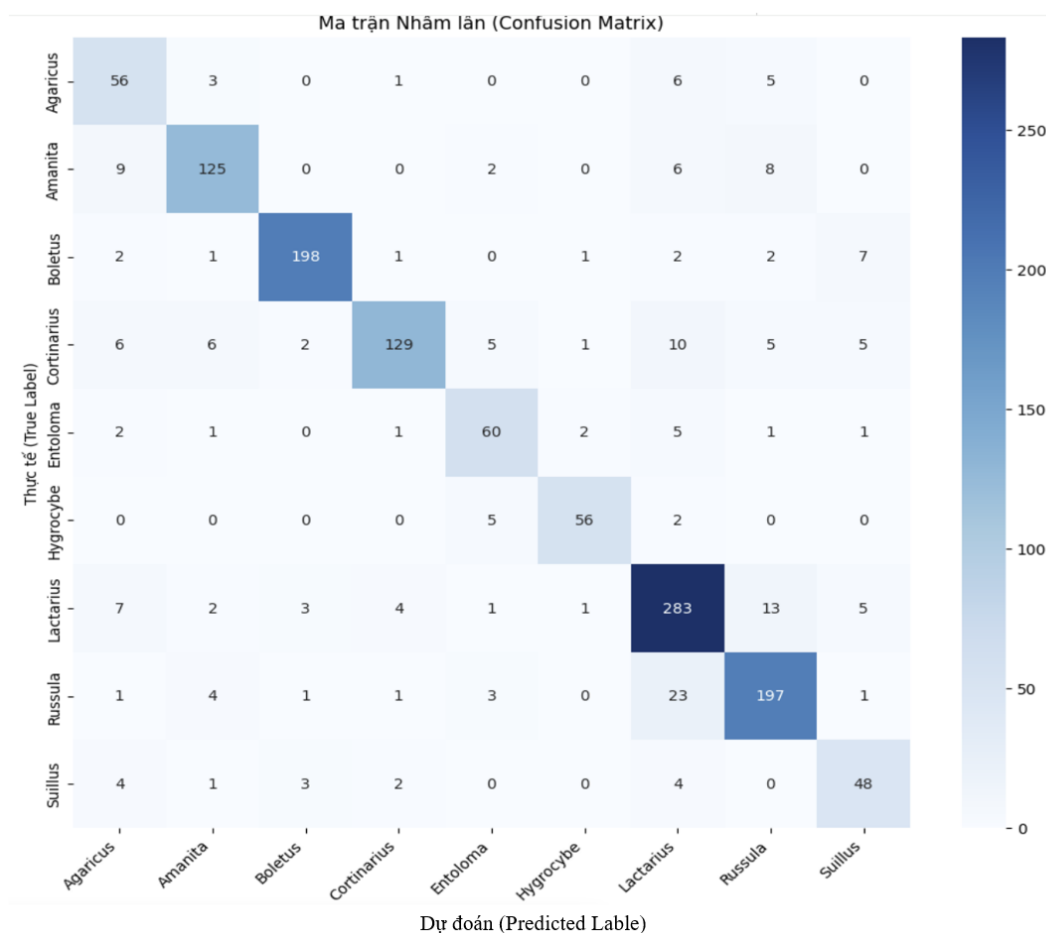
4. Kết quả, thảo luận

Trong nghiên cứu này, chúng tôi tập trung xây dựng mô hình trí tuệ nhân tạo phục vụ bài toán nhận dạng và phân loại chỉ nắm dựa trên dữ liệu hình ảnh. Mô hình được phát triển bằng ngôn ngữ lập trình Python kết hợp với các thư viện học máy và đánh giá mô hình như scikit-learn nhằm tối ưu hóa quá trình huấn luyện và đánh giá hiệu suất. Dữ liệu thực nghiệm bao gồm hình ảnh của 9 chi nắm đã được các chuyên gia nắm học gán nhãn và công bố trên nền tảng Kaggle (kaggle, 2018). Hiệu quả của mô hình AI được kiểm chứng thông qua các thí nghiệm thực nghiệm và các chỉ số đánh giá phân loại. Kết quả nghiên cứu sẽ được trình bày, phân tích và đối chiếu với các công trình liên quan nhằm làm rõ khả năng học đặc trưng hình ảnh, hiệu suất nhận dạng cũng như tiềm năng ứng dụng thực tế của mô hình đề xuất trong hỗ trợ nhận dạng nắm.

Ma trận nhầm lẫn (confusion matrix)

Ma trận nhầm lẫn được dùng để đánh giá hiệu suất của mô hình bằng cách so sánh các giá trị mục tiêu thực tế với các giá trị dự đoán. Nó cung cấp phân tích chi tiết về các dự đoán chính xác (đường chéo) và các phân loại sai (ngoài đường chéo), cho thấy những lớp nào thường bị nhầm lẫn. Ma trận nhầm lẫn chứng minh rằng mô hình có khả năng học được các đặc trưng phân biệt quan trọng của từng chi nắm, đặc biệt ở các lớp có số lượng mẫu lớn và đặc điểm hình thái rõ ràng. Tuy nhiên, các lỗi nhầm lẫn giữa

Lactarius và *Russula*, cũng như giữa *Cortinarius* và *Lactarius*, cho thấy cần tiếp tục cải thiện khả năng trích xuất đặc trưng ở những nhóm nấm có hình thái tương đồng. Việc kết hợp thêm các kỹ thuật trực quan hóa như Grad-CAM có thể giúp xác định vùng ảnh mà mô hình tập trung khi dự đoán, từ đó hỗ trợ đánh giá liệu mô hình đang dựa vào các đặc điểm sinh học phù hợp như mũ nấm, phiến nấm, cuống nấm. Chi tiết ma trận nhầm lẫn được minh họa trong Hình 7.

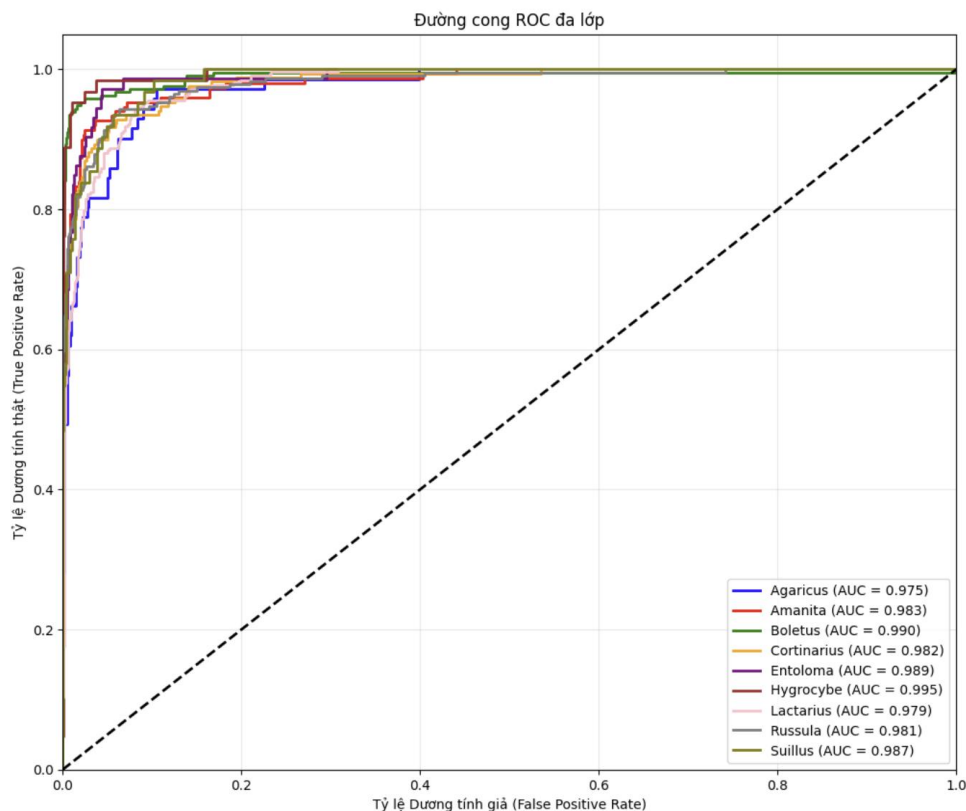


Hình 7. Confusion matrix của mô hình định danh chi nấm

Nguồn: tác giả, 2026

Kết quả mô hình phân loại từ dữ liệu hình ảnh

Kết quả ROC được thể hiện ở Hình 8 cho thấy mô hình có khả năng phân loại tên chi nấm rất tốt. Hầu hết các đường cong ROC đều nằm gần góc trên bên trái và cách xa đường chéo ngẫu nhiên, chứng tỏ mô hình đạt tỷ lệ dự đoán đúng cao và tỷ lệ dương tính giả thấp. Các giá trị AUC của các lớp đều lớn hơn 0,97, trong đó *Hygrocybe* đạt cao nhất với 0.995, tiếp theo là *Boletus* với 0,99, *Entoloma* với 0,989 và *Suillus* với 0,987. Điều này cho thấy mô hình có khả năng phân biệt tốt giữa các chi nấm khác nhau. Nhìn chung, kết quả ROC khẳng định mô hình có hiệu suất phân loại ổn định, đáng tin cậy và phù hợp cho bài toán dự báo tên chi nấm dựa trên hình ảnh

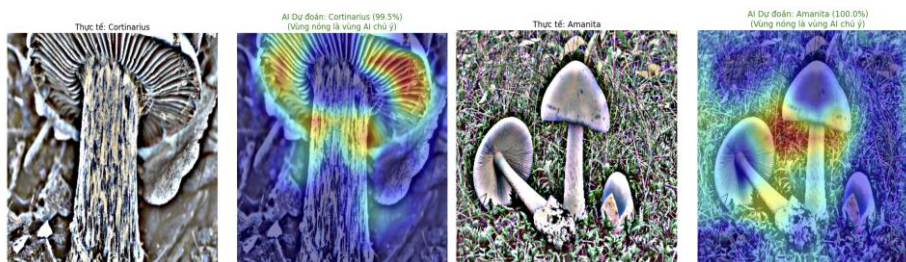


Hình 8. Kết quả hiệu năng của mô hình đề xuất.

Nguồn: tác giả, 2026

Hình 9 cho thấy mô hình đã được sử dụng để dự đoán tên chi nấm dựa trên ảnh đầu vào. Ở ảnh bên trái là hình ảnh thực tế của mẫu nấm, trong khi ảnh bên phải thể hiện kết quả dự đoán của mô hình kết hợp với bản đồ chú ý trực quan. Mô hình đã dự đoán mẫu nấm thuộc chi *Cortinarius* là, và mẫu nấm thuộc chi *Amanita* với độ tin cậy rất cao, đạt 100%. Bản đồ màu trên ảnh bên phải cho thấy mô hình tập trung chủ yếu vào các vùng đặc trưng quan trọng của cây nấm như mũ nấm, thân nấm và phần gốc, đây là những đặc điểm hình thái có ý nghĩa trong việc nhận diện chi nấm. Điều này chứng minh rằng mô hình không chỉ đưa ra kết quả phân loại chính xác mà còn có khả năng học và khai thác các đặc trưng thị giác liên quan đến cấu trúc sinh học của nấm.

Nhìn chung, kết quả minh họa cho thấy mô hình xây dựng có khả năng dự đoán tên chi nấm hiệu quả, đồng thời bản đồ chú ý giúp tăng tính giải thích và độ tin cậy của mô hình trong quá trình phân loại hình ảnh nấm.



Hình 9. Kết quả nhận dạng hình ảnh các chi nấm

Nguồn: tác giả, 2026

So sánh hiệu năng mô hình thực nghiệm với các nghiên cứu khác.

Kết quả thực nghiệm trên bộ dữ liệu hình ảnh nấm thực tế cho thấy mô hình đề xuất đạt độ chính xác phân loại lên đến 98%, cho thấy hiệu năng vượt trội so với nhiều nghiên cứu trước đây trong lĩnh vực nhận dạng và phân loại nấm bằng thị giác máy tính. Việc kết hợp giữa kiến trúc EfficientNetV2-S và cơ chế trực quan hóa Grad-CAM không chỉ cải thiện hiệu suất phân loại mà còn cung cấp bằng chứng trực quan về cách mô hình nhận diện các đặc điểm phân biệt giữa các chi nấm, các Hình ảnh 7, 8 và 9 mô tả chi tiết các khía cạnh này. Bên cạnh đó, các kết quả của mô hình đề xuất còn được so sánh với kết quả dự đoán cùng chức năng như nghiên cứu của tác giả (Zhao, Kexin, et al, 2023) và (Ozsari et al., 2024) Các độ đo hiệu năng của mô hình đề xuất được liệt kê chi tiết ở Bảng 2.

Bảng 2. So sánh hiệu năng của mô hình đề xuất với các tác giả trên cùng tập dữ liệu nghiên cứu

Mô hình tham khảo	Kỹ thuật sử dụng	Accuracy	Precision	Recall	F1 score
(Zhao, Kexin, et al, 2023)	Swin Transformer	0.87	-	-	-
(Ozsari, et al, 2024)	DenseNet121	0.92	0.94	0.90	0.92
Mô hình đề xuất	EfficientNetV2_S	0.98	0.95	0.98	0.98

5. Kết luận

Bài báo này đề xuất một khung học sâu gọn nhẹ, hiệu quả và có khả năng diễn giải dựa trên kiến trúc EfficientNetV2-S nhằm phân loại hình ảnh nấm với độ chính xác cao. Điểm nổi bật của nghiên cứu là việc tích hợp kỹ thuật Gradient-weighted Class Activation Mapping (Grad-CAM), Grad-CAM trong giai đoạn tiền xử lý và phân tích ảnh, giúp xác định các vùng đặc trưng quan trọng mà mô hình sử dụng để đưa ra quyết định phân loại. Thay vì chỉ khai thác toàn bộ ảnh đầu vào một cách thụ động, Grad-CAM hỗ trợ làm nổi bật các khu vực có giá trị sinh học và hình thái học cao, chẳng hạn như mũ nấm, cuống nấm, phiến nấm hoặc các đặc điểm bề mặt đặc trưng. Nhờ đó, quá trình học đặc trưng của mô hình trở nên tập trung hơn, đồng thời tăng cường khả năng giải thích kết quả dự đoán.

Nhìn chung, nghiên cứu chứng minh rằng các mô hình học sâu gọn nhẹ, khi được kết hợp với kỹ thuật tiền xử lý và diễn giải ảnh dựa trên Grad-CAM, có thể đạt được ba giá trị quan trọng: hiệu suất cao, khả năng triển khai thực tiễn và tính minh bạch trong dự đoán. Những kết quả này không chỉ đóng góp nền tảng học thuật cho các nghiên cứu tiếp theo về phân loại nấm mà còn mở ra tiềm năng ứng dụng trong nông nghiệp thông minh, an toàn thực phẩm và các hệ thống nhận dạng sinh học được hỗ trợ bởi trí tuệ nhân tạo.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Thủ Dầu Một trong đề tài mã số DT.23.2-064

TÀI LIỆU THAM KHẢO

[1] Devika, G., and Asha Gowda Karegowda. (2021). Identification of edible and non-edible mushroom through convolution neural network. In 3rd International Conference on Integrated Intelligent Computing Communication & Security, 312-321.

- [2] Dong Jian E, et al. (2020). Deep learning for species identification of bolete mushrooms with two-dimensional correlation spectral (2DCOS) images. *Spectrochimica Acta Part A: Molecular and Biomolecular*.
- [3] Dosovitskiy, et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint*.
- [4] Duong, Thi Kim Chi, et al. (2023). Ensemble learning-based approach for automatic classification of termite mushrooms. *Frontiers in Genetics 14*.
- [5] Feng, J., Ling, X., Wang, Y., Yang, J.:(2023). Mushroom classification based on deep residual network. In: *2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML)*, Urumqi, China, 128–133. doi.: <https://doi.org/10.1109/PRML5>
- [6] Hebert PD, C. A. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270(1512):313–21.
- [7] Kaggle. (2018).2018 FGVCx Fungi Classification Challenge, <https://www.kaggle.com/code/kill92/90-acc-edible-and-poisonous-mushrooms-classifier/input>.
- [8] Lu H., Lou H., et al. (2020). Macrofungi: A Review of Cultivation Strategies, Bioactivity, and Application of Mushrooms. *Food Sci. Food Saf. 2020*, 19:2333–2356.
- [9] Chawla, N. V., et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- [10] Ozsari, et al. (2024). Deep Learning-Based Classification of Macrofungi: Comparative Analysis of Advanced Models for Accurate Fungi Identification. *Sensors 24*.
- [11] Panchbhair, K.G. et al. (2024). Small-size cnn (cas-cnn), and modified mobilenetv2 (cas-modmobnet) to identify cashew nut and fruit diseases. *Multimed. Tools Appl.*, 1–21.
- [12] Preechasuk, et al. (2019). Image analysis of mushroom types classification by convolution neural networks. *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*.
- [13] Somervuo P, K. S. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics*, 32(19):2920–7.
- [14] Tan, Mingxing, and Quoc Le. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, 6105–6114.
- [15] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery*.
- [16] Valter Augusto de Freitas Barbosaa, Juliana Carneiro Gomesb, Maira Araujo de Santana Clarisse Lins de et al. (2021). Covid-19 rapid test by combining a Random Forest-based web system. *JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS*. doi: DOI:10.1080/07391102.2021.1966509
- [17] Wang, Boyuan. (2022). Automatic Mushroom Species Classification Model for Foodborne Disease Prevention Based on Vision Transformer. *Journal of Food Quality*.
- [18] Zhao, Kexin, et al. (2023). Mushroom Image Classification and Recognition Based on Improved Swin Transformer. *IEEE 6th International Conference on Information Systems and Computer Aided Education*.