



DOI:10.22144/ctujos.2025.048

NHẬN DẠNG MỘT SỐ LOẠI CÂY THUỐC BẰNG MÔ HÌNH CONVOLUTIONAL NEURAL NETWORK (CNN) TRÊN ỨNG DỤNG DI ĐỘNG

Nguyễn Đức Tấn* và Thái Thuận Thương

Khoa Công nghệ Thông tin, Trường Đại học Yersin Đà Lạt, Việt Nam

*Tác giả liên hệ (Corresponding author): tannd@yersin.edu.vn

Thông tin chung (Article Information)

Nhận bài (Received): 14/09/2024

Sửa bài (Revised): 08/10/2024

Duyệt đăng (Accepted): 24/12/2024

Title: Identifying medicinal plants using convolutional neural network (CNN) on mobile devices

Author(s): Nguyen Duc Tan* and Thai Thuan Thuong

Affiliation(s): Faculty of Information Technology, Yersin University of Da Lat, Viet Nam

TÓM TẮT

Việt Nam sở hữu một hệ thống y học cổ truyền từ lâu đời với các loại cây thuốc và thảo dược đa dạng đã đóng góp lớn vào công tác phòng và chữa bệnh tại địa phương. Tuy nhiên, việc nhận biết chính xác cây thuốc đòi hỏi nhiều kiến thức và kinh nghiệm của người dùng. Do đó, nhu cầu tự động nhận dạng hình ảnh cây thuốc bằng ứng dụng di động là hữu ích với nền y học. Tập dữ liệu hình ảnh 41 loại cây thuốc và dược liệu được sử dụng trong nghiên cứu, tổng hợp từ nhiều nguồn, sau đó được huấn luyện qua các mô hình CNN. Kết quả cho thấy mô hình EfficientB0 với phương pháp tăng cường dữ liệu và học chuyển giao phù hợp để triển khai trên thiết bị di động với tỷ lệ nhận dạng chính xác đạt ~94% cao hơn mô hình MobileNetV2 đạt ~90%. Ứng dụng di động nhận dạng cây thuốc là công cụ hỗ trợ các đối tượng hoạt động trong các lĩnh vực y học cổ truyền, nông nghiệp và giáo dục.

Từ khóa: Học chuyển giao, mạng nơ-ron tích chập, nhận dạng cây thuốc, tăng cường dữ liệu, ứng dụng di động

ABSTRACT

Vietnam possesses a long-standing traditional medicine system with a variety of medicinal plants and herbs that have contributed significantly to local disease prevention and treatment. However, accurately identifying medicinal plants requires sufficient knowledge and experience from users. Therefore, the demand for automatic recognition of medicinal plant images through mobile applications is valuable to the field of medicine. The study used an image dataset of 41 medicinal plants and herbs synthesized from various sources and then trained through CNN models. The results showed that the EfficientB0 model with data augmentation and transfer learning methods is suitable for deployment on mobile devices with an accuracy rate of ~94%, surpassing the ~90% achieved by the MobileNetV2 model. The mobile application for medicinal plant identification is a tool that supports individuals working in the fields of traditional medicine, agriculture, and education.

Keywords: Convolutional neural networks, data augmentation, medicinal plant identification, mobile application, transfer learning

1. GIỚI THIỆU

Việt Nam là một quốc gia nhiệt đới với sự đa dạng về khí hậu và loại địa hình phù hợp cho sự sinh trưởng của nhiều loại dược liệu quý. Theo dữ liệu báo cáo từ World Bank Group (2023), hơn 60% dân số sống và làm việc ở nông thôn, trong đó có một số lượng lớn các nhóm dân tộc thiểu số sinh sống ở vùng miền núi và cao nguyên. Người dân ở đây gặp khó khăn trong việc tiếp cận các loại thuốc hiện đại. Tuy nhiên, Việt Nam lại sở hữu một hệ thống y học cổ truyền từ lâu đời với các loại cây thuốc và thảo dược đã đóng góp lớn vào công tác phòng và chữa bệnh tại địa phương. Dựa trên kinh nghiệm truyền lại, người dân ở vùng nông thôn và miền núi từ xưa đã phân biệt cây thuốc dựa vào đặc điểm về mùi vị, lá, thân, rễ và hoa. Các loại cây thuốc có sẵn tại địa phương này đã được sử dụng để điều trị một số bệnh nói chung, ngoài ra một số loại còn được sử dụng như rau và gia vị hàng ngày để bảo vệ sức khỏe và cung cấp dinh dưỡng. Theo thực tiễn và nghiên cứu khoa học đã chứng minh, cây thuốc và thảo dược ở Việt Nam có hiệu quả trong việc phòng và chữa các loại bệnh thông thường, thậm chí một số bệnh khó chữa. Cây thuốc và dược liệu có ý nghĩa to lớn trong lĩnh vực y tế cộng đồng, nguồn dược liệu này vừa là nguồn cung cấp thuốc cho quốc gia, vừa là mặt hàng xuất khẩu có giá trị. Theo công bố của các nhà nghiên cứu, Việt Nam có hơn 5100 loại cây thuốc và dược liệu, bên cạnh các loại dược liệu phổ biến còn có một số loại rất quý hiếm và có giá trị. Việc xác định chính xác các cây thuốc dược liệu là rất quan trọng để xác định các đặc tính dược liệu và ứng dụng chữa bệnh tiềm năng. Do đó, để phục vụ công tác nghiên cứu, giảng dạy ứng dụng thực tế của cây thuốc, một ấn phẩm “Medical plant in Viet Nam” (World Health Organization Western Pacific, 1990) đã cung cấp thông tin gồm 200 cây thuốc phổ biến ở Việt Nam, kèm theo mô tả chi tiết bằng hình ảnh để nhận biết, phân loại các loại dược liệu và công dụng chữa bệnh của chúng. Một vấn đề tồn tại đối với số lượng dược liệu vô cùng phong phú như vậy là khó khăn trong việc nhận biết một cách chính xác cây thuốc dược liệu dựa trên các đặc điểm hình thái của một số dược liệu do chúng có tính tương đồng, bên cạnh đó việc thực hiện bằng mắt thường sẽ tốn nhiều thời gian, công sức và đòi hỏi nhiều kiến thức và kinh nghiệm của người dùng.

Trong lĩnh vực xử lý ảnh, gần đây có nhiều nghiên cứu về ứng dụng các mô hình Deep learning để nhận dạng hình ảnh trong nông nghiệp, lâm nghiệp, y học,... Cụ thể, nhóm nghiên cứu Trung quốc đã ứng dụng Deep learning với mô hình ResNet26 để nhận dạng hình ảnh của 100 loại cây

cảnh với độ chính xác đạt 91,78% (Sun et al., 2017). Một nghiên cứu khác từ Ấn Độ đã sử dụng mô hình Convolutional Neural Network (CNN) - MobileNetV2 để nhận dạng 13 loại hoa của cây thuốc với độ chính xác cao nhất đạt 98,23% (Bipin Nair et al., 2024). Đặc biệt, với sự phát triển của thiết bị di động thông minh và các thiết bị nhúng, nhóm nghiên cứu Indonesia đã tiến hành thực nghiệm đánh giá độ chính xác của 3 mô hình CNN: VGG-16 (Liu & Deng, 2015), MobileNetV2 (Sandler et al., 2018) và DenseNet-121 (Huang et al., 2017) và sau đó chọn mô hình MobileNetV2 tích hợp vào ứng dụng di động với khả năng nhận dạng 24 loại cây thuốc ở Indonesia dựa trên tập hình ảnh về lá với độ chính xác đạt 97,74% (Sugiarto et al., 2023). Ở Việt Nam, Vo et al. (2019) đã sử dụng mô hình học sâu VGG16 kết hợp kỹ thuật phân loại LightGBM (Ke et al., 2017) để nhận dạng 10 loại cây dược liệu trong tự nhiên đạt độ chính xác 93,6%. Bên cạnh đó, nghiên cứu của nhóm tác giả Nguyễn Quốc Trung (Nguyen & Truong, 2020) về nhận dạng cây thuốc được tiến hành trên các mô hình CNN như: VGG16, Resnet50 (He et al., 2015), InceptionV3 (Szegedy et al., 2016), DenseNet121, Xception (Chollet, 2017) và MobileNet. Kết quả thực nghiệm cho thấy mô hình Xception đạt độ chính xác cao 88,26%.

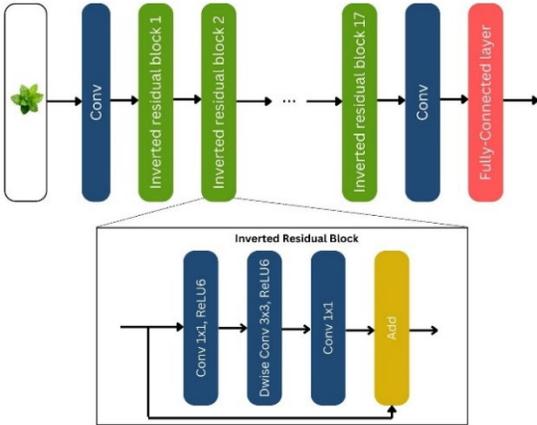
Với tầm quan trọng của việc nhận dạng cây thuốc đặc thù ở Việt Nam, sự phát triển mạnh mẽ của công nghệ học sâu (Deep learning) cùng với sự phổ biến của thiết bị di động thông minh với đặc điểm giới hạn về tài nguyên và khả năng tính toán so với máy tính, giải pháp nhận dạng cây thuốc trên ứng dụng thiết bị di động được đề xuất trong nghiên cứu, tích hợp mô hình mạng nơ-ron tích chập hiện đại EfficientB0 kết hợp với phương pháp tăng cường dữ liệu và học chuyển giao để tăng hiệu suất nhận dạng. Bên cạnh đó, mô hình MobileNetV2 cũng được sử dụng để so sánh kết quả dựa trên bộ dữ liệu hình ảnh của 41 loại dược liệu ở Việt Nam được thu thập từ các nguồn từ internet, cơ sở dữ liệu quốc gia và thực tiễn.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Mô hình MobileNetV2

MobileNetV2 là một mô hình mạng nơ-ron tích chập được giới thiệu vào năm 2018 bởi nhóm nghiên cứu Google (Sandler et al., 2018). Mô hình được thiết kế để giảm thiểu số lượng phép tính và dung lượng mô hình, phù hợp để triển khai trên các thiết bị di động và hệ thống tính toán hạn chế. Mô hình này là sự cải tiến của MobileNetV1 (Howard et al., 2017), MobileNetV2 vẫn sử dụng kỹ thuật tích chập tách biệt chiều sâu (depthwise separable

convolution) thành hai bước tích chập chiều sâu (depthwise convolution) và tích chập điểm (pointwise convolution) để giảm đáng kể số lượng phép tính cần thực hiện (Hình 1).



Hình 1. Kiến trúc mô hình MobileNetV2

Tuy nhiên, mô hình có thêm cải tiến ở khối đảo ngược (Inverted Residual Block). Lớp đầu tiên mở rộng số lượng kênh đầu vào thông qua một lớp pointwise convolution (1x1). Sau khi mở rộng số lượng kênh, dữ liệu đi qua một lớp depthwise convolution - tích chập riêng cho từng kênh để giảm

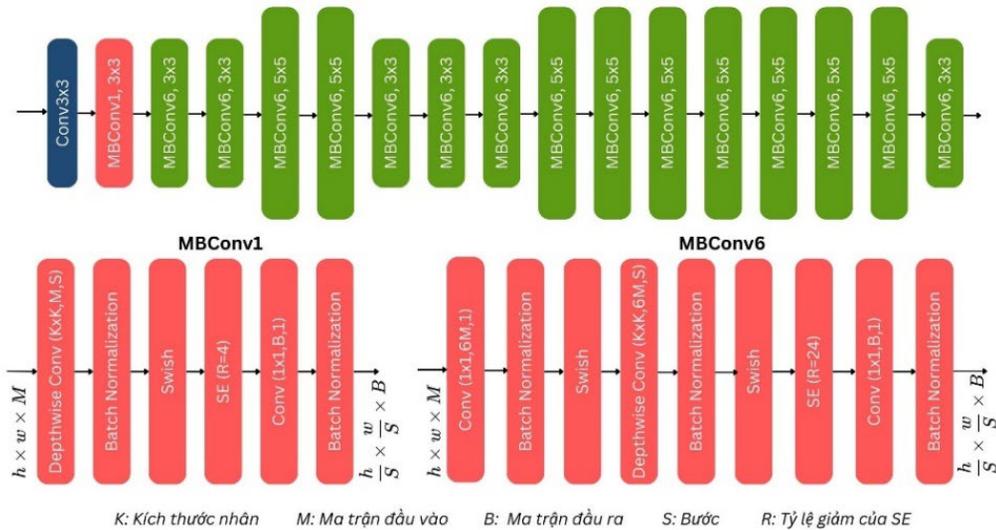
số lượng phép tính và tăng hiệu suất. Tiếp theo, số lượng kênh được thu hẹp về kích thước ban đầu thông qua một lớp pointwise convolution (1x1). Cuối cùng, kết nối tắt chỉ được sử dụng nếu kích thước của đầu vào và đầu ra tương thích (chiều rộng, chiều cao và số lượng kênh giống nhau). Ngoài ra, nút thắt tuyến tính (Linear bottleneck) là thành phần cuối của khối Inverted Residual, thay vì sử dụng hàm kích hoạt phi tuyến tại bước này, Linear bottleneck giữ tuyến tính để tránh mất mát thông tin.

2.2. Mô hình EfficientB0

Kiến trúc mô hình EfficientNet được đề xuất bởi nhóm nghiên cứu của Google AI (Tan & Le, 2019).

Mô hình phát triển dựa trên mạng nơ-ron tích chập bằng cách mở rộng đồng thời chiều rộng, chiều cao và độ phân giải với một tỉ lệ phù hợp, giúp mô hình tăng độ chính xác mà vẫn giữ được hiệu suất tốt. EfficientNet-B0 là một trong những kiến trúc của họ EfficientNet.

EfficientNetB0 sử dụng MBConv (Mobile Inverted Bottleneck Convolution) với các cải tiến thêm so với Inverted Residual Block bao gồm thêm các kỹ thuật như Squeeze-and-Excitation (SE) để tối ưu hóa hiệu suất (Hình 2).



Hình 2. Kiến trúc mô hình EfficientB0

2.3. Học chuyển giao

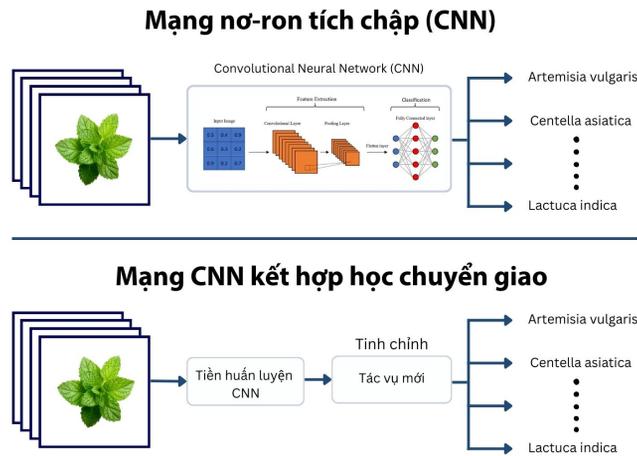
Học chuyển giao có thể được triển khai dựa trên bất kỳ mô hình học máy nào, kỹ thuật được đánh giá là trở nên phổ biến hơn khi sử dụng với học sâu (Ali et al., 2023; Gupta et al., 2022). Một mạng nơ-ron tích chập (CNN) sẽ được đào tạo trên một tập dữ liệu

nhất định để trích xuất các đặc điểm. Dựa trên các đặc điểm đã học, mô hình có thể dự đoán kết quả. Để cải thiện độ chính xác của CNN, cần có một lượng lớn dữ liệu để đào tạo mô hình. Trong thực tế, có một số khó khăn khi thu thập một lượng lớn dữ liệu. Và một trường hợp khác – nếu một phần dữ liệu

đã thu thập bị thiếu thì dữ liệu đó không đủ tin cậy để triển khai với các mô hình CNN. Giải pháp cho những vấn đề này là kỹ thuật chuyển giao học tập.

Hình 3 trình bày về kỹ thuật học chuyển giao, một mô hình có thể duy trì các tham số tối ưu dựa

trên quy trình đào tạo trên một tập dữ liệu phổ biến như ImageNet (Kornblith et al., 2018). Sau đó, các đặc điểm đã học được sử dụng lại cho tác vụ huấn luyện tiếp theo trong quá trình tinh chỉnh (fine-tuning). Nhờ đó, độ chính xác tổng thể của mô hình được cải thiện.



Hình 3. Kỹ thuật học chuyển giao

2.4. Thu thập và tiền xử lý hình ảnh

2.4.1. Thu thập hình ảnh

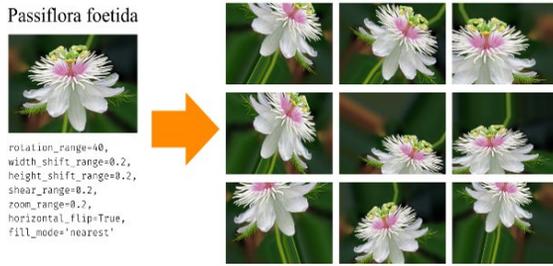
Mục tiêu của bước này là thu thập hình ảnh dữ liệu cây thuốc và thảo dược để dùng cho huấn luyện mô hình. Đây là bước quan trọng và quyết định đến chất lượng của mô hình CNN.

Nghiên cứu này đã thu thập hình ảnh của 41 loại cây thuốc và thảo dược phổ biến, việc thu thập hình ảnh và gắn nhãn loại dược liệu dựa trên các thông tin mô tả chi tiết về tên khoa học, tên dân gian, họ cây thuốc kèm hình vẽ mô tả hình thái tự nhiên của lá hoặc hoa và được đặc tả chi tiết trong ấn phẩm “Medical plant in Viet Nam” (World Health Organization Western Pacific, 1990). Đối với mỗi loại dược liệu thu thập tối thiểu 15 hình ảnh, tối đa 25 hình ảnh để bảo đảm độ cân bằng giữa các lớp trong bài toán phân loại. Các nguồn để thu thập hình ảnh từ các trang web, cơ sở dữ liệu công cộng như ImageNet, Kaggle datasets, Google Images hoặc Flickr... Bên cạnh đó, nếu dữ liệu đặc thù hoặc không có sẵn trực tuyến, có thể tự chụp ảnh bằng máy ảnh kỹ thuật số hoặc điện thoại thông minh. Ngoài ra, cơ sở dữ liệu chuyên ngành, tạp chí cũng là nguồn để cung cấp các hình ảnh cây thuốc liên quan.

Tiêu chí dữ liệu hình ảnh: dùng kỹ thuật quét hoặc chụp rõ nét hình ảnh cây thuốc, đặc biệt phải chứa đặc trưng nhận dạng của loại cây đó, có thể lá hoặc hoa của cây thuốc trên nền đơn sắc hoặc nền tự nhiên.

2.4.2. Tăng cường dữ liệu

Tăng cường hình ảnh (Image Augmentation) là kỹ thuật được sử dụng để tăng số lượng của dữ liệu hình ảnh cho quá trình huấn luyện mà không cần thu thập thêm dữ liệu mới (Mikolajczyk & Grochowski, 2018), kỹ thuật này áp dụng các biến đổi ngẫu nhiên lên hình ảnh gốc. Điều này đặc biệt quan trọng khi dữ liệu huấn luyện ban đầu hạn chế, giúp cải thiện khả năng tổng quát hóa của mô hình học sâu, giảm nguy cơ overfitting và làm cho mô hình trở nên mạnh mẽ hơn khi gặp phải dữ liệu mới, chưa từng thấy trước đó. Một số phép biến đổi hình ảnh phổ biến như: xoay hình ảnh, dịch chuyển hình ảnh, phóng to/thu nhỏ hình ảnh, lật hình ảnh, biến dạng hình học, thay đổi độ sáng, thay đổi độ tương phản, thêm nhiễu, thay đổi hình dạng, thay đổi độ bão hòa (Hình 4).



Hình 4. Ví dụ hình ảnh tăng cường

Trong thực nghiệm này, phép biến đổi hình ngẫu nhiên sử dụng hàm ImageDataGenerator trong thư viện Keras với tham số (Bảng 1) để tăng số lượng dữ liệu hình ảnh lên 9 lần đối với tập dữ liệu huấn luyện (training set).

2.4.3. Phân chia dữ liệu

Trong bài toán CNN nhận dạng hình ảnh, việc phân chia tập dữ liệu là một bước quan trọng để đảm bảo rằng mô hình được huấn luyện và học được các đặc trưng từ dữ liệu, cũng như có khả năng tổng quát hóa tốt trên các dữ liệu mới.

Bảng 1. Tham số tăng cường ảnh áp dụng trong thực nghiệm

Tham số	Giá trị	Ý nghĩa
rotation_range	40	Xoay ngẫu nhiên hình ảnh trong khoảng góc từ -40 độ đến +40 độ.
width_shift_range	0,2	Dịch chuyển ngẫu nhiên hình ảnh theo chiều ngang (trục x) trong phạm vi 20% của chiều rộng hình ảnh.
height_shift_range	0,2	Dịch chuyển ngẫu nhiên hình ảnh theo chiều dọc (trục y) trong phạm vi 20% của chiều cao hình ảnh.
shear_range	0,2	Áp dụng biến dạng hình học (shearing) cho hình ảnh, với cường độ biến dạng tối đa là 20%.
zoom_range	0,2	Phóng to hoặc thu nhỏ hình ảnh ngẫu nhiên trong phạm vi từ 80% đến 120% so với kích thước gốc.
horizontal_flip	True	Lật ngẫu nhiên hình ảnh theo chiều ngang (trục x).
fill_mode	'nearest'	Khi áp dụng các phép biến đổi như xoay, dịch chuyển, shear, hoặc zoom, một số vùng của hình ảnh có thể bị mất đi và cần được điền thêm. fill_mode='nearest' sẽ điền các vùng trống này bằng cách sử dụng màu sắc của các điểm ảnh gần nhất (nearest neighbor).

Bảng 2. Phân chia dữ liệu trước và sau khi tăng cường hình ảnh

Tập dữ liệu	Dữ liệu trước tăng cường			Dữ liệu sau khi tăng cường		
	Train	Validation	Test	Train	Validation	Test
Số lượng hình ảnh	422	258	124	3771	258	124

Sau khi thu thập dữ liệu hình ảnh 41 loại cây thuốc, dữ liệu này được gắn nhãn theo thư mục. Cụ thể, mỗi loại cây thuốc sẽ có một thư mục riêng với tên nhãn là tên khoa học của cây thuốc, trong thư mục chứa các hình ảnh của loại cây thuốc này. Ngoài ra, để thực hiện huấn luyện và kiểm tra, tập dữ liệu cần chia thành các tập: tập huấn luyện (training set) được áp dụng kỹ thuật tăng cường dữ liệu để huấn luyện mô hình; tập xác thực (validation set) được sử dụng để điều chỉnh các siêu tham số (hyperparameters) và kiểm tra khả năng của mô hình trên dữ liệu chưa gặp trong quá trình huấn luyện; tập kiểm tra (test set) chiếm tỷ lệ ~15% bao gồm các hình ảnh chưa được mô hình huấn luyện, được sử dụng để đánh giá cuối cùng về hiệu suất của mô hình sau khi quá trình huấn luyện và tinh chỉnh đã hoàn tất (Bảng 2).

2.4.4. Tiền xử lý hình ảnh

Tiền xử lý dữ liệu hình ảnh là một bước thiết yếu nhằm chuẩn bị và cải thiện chất lượng dữ liệu trước khi đưa vào mô hình học máy hoặc học sâu. Quá trình này giúp làm sạch bao gồm loại bỏ các yếu tố không cần thiết, làm nổi bật các đặc trưng quan trọng, chuẩn hóa, làm giảm thiểu hiện tượng quá khớp (overfitting), tối ưu hóa thời gian huấn luyện và tài nguyên tính toán để mô hình có thể học và nhận dạng các đặc trưng một cách hiệu quả nhất.

Các bước tiền xử lý phổ biến trong nhận dạng hình ảnh áp dụng trong thực nghiệm gồm:

Điều chỉnh kích thước hình ảnh (Resizing): Đưa tất cả các hình ảnh về cùng một kích thước chuẩn mà mô hình yêu cầu (ví dụ: 224x224 pixel). Điều này giúp giảm thiểu sự biến đổi về kích thước, giúp mô hình xử lý dữ liệu đồng nhất và tiết kiệm tài nguyên tính toán.

Chuẩn hóa giá trị pixel (Normalization): Công thức chuẩn hoá (1) được áp dụng để đưa giá trị pixel về cùng một phạm vi [0, 1], bằng cách chia giá trị pixel tại vị trí i cho 255 (Koo & Cha, 2017). Mục đích nhằm giảm thiểu sự chênh lệch về giá trị và giúp mô hình hội tụ nhanh hơn trong quá trình huấn luyện.

$$x_o = \frac{x_i}{255} \in [0,1] \quad (1)$$

Bảng 3. Thông số của các mô hình CNN thực nghiệm

Thông tin	MobileNetV2	EfficientNetB0
Số lượng lớp	53 lớp	237 lớp
Kích thước mô hình	~14 MB	~20 MB
Số lượng tham số	~3,4 triệu (3,4M)	~5,3 triệu (5,3M)
FLOPs	~300 triệu (300M)	~390 triệu (390M)
Kích thước đầu vào mặc định	224x224 pixels	224x224 pixels
Kiến trúc đặc trưng	Depthwise Separable Convolutions, Inverted Residuals với Linear Bottlenecks.	MBConv Blocks (Mobile Inverted Bottleneck Convolution), Compound Scaling

Trước khi bắt đầu quá trình huấn luyện, mô hình cần phải khởi tạo giá trị các siêu tham số vì các tham số này không được học trực tiếp từ dữ liệu mà được sử dụng để điều khiển quá trình huấn luyện mô hình. Một số siêu tham số quan trọng cần thiết lập: tốc độ học (learning rate), số vòng lặp qua toàn bộ tập dữ liệu huấn luyện (epochs), kích thước lô dữ liệu (batch size), thuật toán tối ưu hóa (Optimizer).

2.6. Kiểm tra và đánh giá mô hình

Để đánh giá toàn diện về hiệu suất của mô hình CNN và xác định khả năng tổng quát hóa của mô hình trên dữ liệu chưa từng thấy trước đó, đồng thời so sánh hiệu suất với giữa các mạng CNN, các chỉ số Accuracy, Precision, Recall, F1-score (D. M. W. Powers, 2011) được sử dụng để đánh giá hiệu suất.

Độ chính xác (Accuracy) được tính bằng tỷ lệ giữa số lượng mẫu phân loại đúng (Correct prediction number) trên tổng số lượng mẫu phân loại (sample number).

$$Accuracy = \frac{Correct\ prediction\ number}{Sample\ number} \quad (2)$$

Độ chuẩn xác (Precision): Tỷ lệ giữa số lượng dự đoán dương tính đúng so với tổng số dự đoán dương tính của mô hình. Độ chính xác của các dự đoán dương tính được tập trung (giảm thiểu dự đoán sai dương tính).

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

2.5. Huấn luyện mô hình

Mô hình MobileNetV2 và EfficientNetB0 kết hợp học chuyển giao được sử dụng để huấn luyện mô hình nhận dạng. Thông tin về hai mô hình được mô tả ở Bảng 3. Việc chạy thực nghiệm hai mô hình trên môi trường Google Colab với ngôn ngữ lập trình Python được tiến hành trong nghiên cứu. Google Colab với ưu điểm hỗ trợ GPU Tesla T4 giúp quá trình huấn luyện các mô hình diễn ra nhanh hơn.

Độ phủ (Recall): Tỷ lệ giữa số lượng dự đoán dương tính đúng so với tổng số lượng mẫu thực tế là dương tính, tập trung vào khả năng phát hiện tất cả các mẫu dương tính (giảm thiểu bỏ sót mẫu dương tính).

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F1-score: trung bình cân bằng của độ chuẩn xác và độ phủ, kết hợp cả hai chỉ số này thành một chỉ số duy nhất. F1-Score hữu ích khi quan tâm sự cân bằng giữa độ chuẩn xác và độ phủ của mô hình, đặc biệt trong các tình huống mà dữ liệu không cân bằng.

$$F1 - score = \frac{2xPrecisionxRecall}{PrecisionxRecal} \quad (5)$$

trong đó, TP-True Positive là số dự đoán chính xác, FP-False Positive và FN-False Negative là số dự đoán sai và số dự đoán bị sót.

2.7. Triển khai mô hình và thiết kế ứng dụng di động

Các thiết bị di động, máy tính bảng và hệ thống nhúng IoT thường có hạn chế về phần cứng và khả năng xử lý, do đó các mô hình CNN không thể triển khai trực tiếp trên các thiết bị này. Để khắc phục, việc tối ưu mô hình, giảm kích thước để cải thiện tốc độ xử lý, đồng thời tiết kiệm bộ nhớ và năng lượng là cần thiết. Google đã phát triển TensorFlow Lite, một phiên bản đơn giản hơn của TensorFlow, giúp giảm yêu cầu về tài nguyên bộ nhớ.

Bảng 4. Mô tả thiết kế ứng dụng MePDetect

STT	Tên	Mô tả
1	Vùng chọn ảnh từ Gallery/Chụp ảnh	Chọn hình ảnh từ Gallery hoặc nhấn giữ nút Camera trên màn hình để chụp ảnh trực tiếp
2	Vùng hiển thị hình ảnh	Hiển thị hình ảnh cần nhận dạng
3	Vùng hiển thị kết quả	Hiển thị tên khoa học của cây được nhận dạng và độ tin cậy

Ứng dụng demo trên thiết bị di động được phát triển bằng ngôn ngữ Dart thông qua Flutter - một khung nguồn mở do Google phát triển và hỗ trợ. Giao diện ứng dụng được thiết kế theo như mô tả ở Bảng 4. Thư viện Google ML Kit được sử dụng để tích hợp các tính năng nhận dạng bằng mô hình CNN trên thiết bị mà không cần kết nối internet, giúp tiết kiệm tài nguyên và thời gian xử lý.

3. KẾT QUẢ VÀ THẢO LUẬN

Dựa trên kết quả về các chỉ số hiệu suất thu được từ quá trình huấn luyện các mô hình, kết quả được thể hiện ở Bảng 5 cung cấp cái nhìn về hiệu suất của các mô hình CNN nhận dạng tập dữ liệu cây thuốc.

3.1.1. Về phương pháp tăng cường ảnh

Phương pháp tăng cường ảnh (data augmentation) có tác động lớn đến tập dữ liệu để huấn luyện mô hình CNN. Phương pháp đã cải thiện đáng kể hiệu suất nhận dạng của các mô hình, cụ thể: dữ liệu tăng cường với mô hình MobileNetV2 có các chỉ số hiệu suất trung bình đều tăng ~1 - 2% so với dữ liệu chưa tăng cường. Tương tự, dữ liệu tăng cường với mô hình EfficientNetB0 có các chỉ số tăng ~3 - 4%, cải thiện đáng kể so với dữ liệu chưa tăng cường (Bảng 5). Phương pháp này có ý nghĩa trong việc tăng tính đa dạng của tập dữ liệu hình ảnh cây thuốc, đặc biệt là khó khăn trong quá trình thu thập hình ảnh của những loại cây thuốc hoặc thảo dược hiếm, ít hình ảnh.

Bảng 5. Các chỉ số hiệu suất trên tập dữ liệu kiểm tra với các mô hình CNN epochs = 20

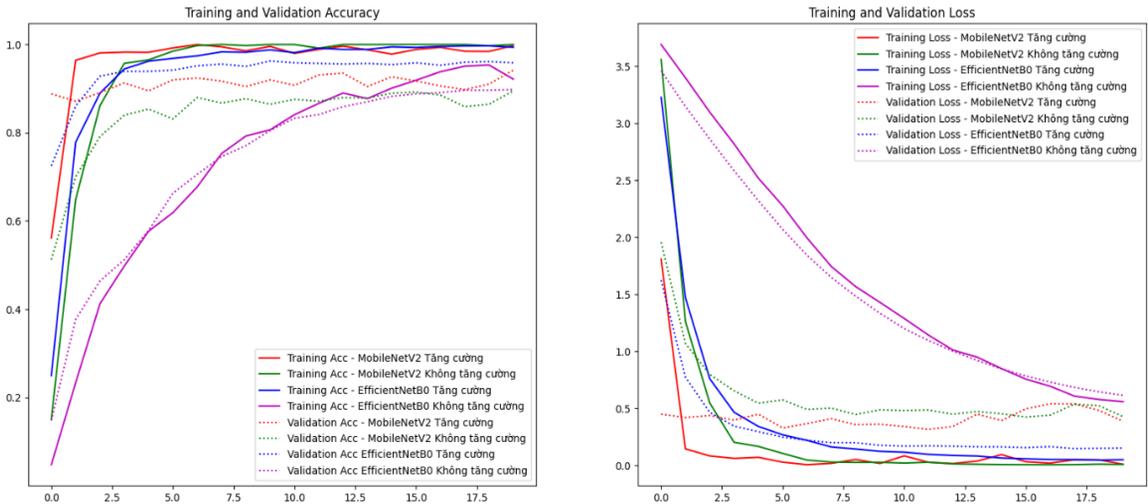
Mô hình	Accuracy	Precision	Recall	F1-Score
MobileNetV2 – Học chuyển giao – Không tăng cường	0,894	0,903	0,894	0,894
EfficientNetB0 - Học chuyển giao – Không tăng cường	0,901	0,912	0,901	0,900
MobileNetV2 - Học chuyển giao – Tăng cường	0,905	0,918	0,905	0,905
EfficientNetB0 - Học chuyển giao – Tăng cường	0,940	0,948	0,940	0,940

3.1.2. So sánh hiệu suất giữa 2 mô hình

Khi bộ dữ liệu hình ảnh hạn chế, trong quá trình huấn luyện, độ chính xác và độ mất mát của dữ liệu huấn luyện và kiểm tra với mô hình EfficientNetB0 ổn định chậm hơn so với MobileNetV2 được thể hiện ở đường học tập accuracy và loss (Hình 5). Nguyên nhân là do sự chênh lệch về số lượng lớp và tham số của kiến trúc mô hình.

Tuy nhiên, khi dữ liệu đa dạng, mặc dù ở những epoch đầu, độ chính xác và độ mất mát của dữ liệu với mô hình EfficientNetB0 ổn định chậm hơn (Hình 5) nhưng các epoch sau đường accuracy và loss của dữ liệu hội tụ tốt và ổn định hơn so với MobileNetV2. Ngoài ra, các chỉ số hiệu suất của EfficientNetB0 cũng vượt trội hơn, cụ thể là F1-score đạt ~94% cao hơn so với MobileNetV2 ~90%.

Nhìn chung, cả 2 mô hình áp dụng học chuyển giao và tăng cường dữ liệu đều đạt các chỉ số hiệu suất cao và phù hợp để triển khai trên thiết bị di động. EfficientNetB0 cho kết quả cao về mặt hiệu suất, nhưng mô hình này có hạn chế là tiêu tốn tài nguyên nhiều hơn MobileNetV2 do EfficientNetB0 có kích thước mô hình lớn và tham số nhiều. Nếu thiết bị di động có khả năng xử lý tốt và có dung lượng lưu trữ đủ lớn thì EfficientNetB0 là một lựa chọn tốt để đảm bảo độ chính xác. Nhưng nếu đối với thiết bị bị giới hạn về tài nguyên thì mô hình này có thể gặp phải các vấn đề về suy luận chậm hơn và tiêu thụ nhiều năng lượng hơn. So với EfficientNetB0, MobileNetV2 đạt sự cân bằng về tốc độ và dung lượng, phù hợp để triển khai trên thiết bị di động bảo đảm hiệu suất tốt nhưng vẫn duy trì tốc độ và mức tiêu thụ tài nguyên hợp lý.



Hình 5. Biểu đồ Loss và Accuracy của 2 mô hình CNN

3.1.3. Ứng dụng di động MePDetect

Trong nghiên cứu này, ứng dụng MePDetect được thiết kế và mô hình CNN EfficientB0 được tích hợp vào ứng dụng di động hoạt động trên hệ

điều hành Android 9.0 Pie chạy trên các thiết bị sử dụng kiến trúc x86. Kết quả nhận dạng ban đầu của một số loại cây dược liệu đạt độ tin cậy cao (Hình 6).



Hình 6. Giao diện ứng dụng và kết quả nhận dạng

4. KẾT LUẬN

Ứng dụng di động sử dụng mô hình EfficientNetB0 được xây dựng để nhận dạng một số loại cây thuốc và dược liệu, kết quả cho thấy khả năng nhận diện tốt. Các mô hình CNN được đánh giá khi được áp dụng kỹ thuật học chuyên giao và tăng cường dữ liệu đã giúp cải thiện hiệu quả huấn luyện với số lượng bộ dữ liệu hạn chế. Đặc biệt, nếu thiết bị di động có khả năng xử lý tốt và có dung lượng lưu trữ đủ lớn thì EfficientNetB0 là một lựa chọn tốt để đảm bảo độ chính xác mô hình. Ngược lại, nếu thiết bị bị giới hạn về tài nguyên, khả năng

tính toán của vi xử lý thì MobileNetV2 là lựa chọn để cân bằng về tốc độ và dung lượng, phù hợp để triển khai trên thiết bị di động bảo đảm tốt về hiệu suất nhưng vẫn duy trì tốc độ xử lý cũng như mức tiêu thụ tài nguyên hợp lý.

Tuy nhiên, nghiên cứu có một số hạn chế như sau: dữ liệu hình ảnh hạn chế do khó khăn trong việc thu thập một số loại thuốc ít gặp, làm giảm khả năng mô hình nhận diện được biến thể của một số cây thuốc trong thực tế; phương pháp triển khai mô hình trực tiếp trên ứng dụng di động hạn chế về khả năng mở rộng do yêu cầu tái huấn luyện mô hình khi nhận diện thêm các loại cây mới hoặc cập nhật dữ liệu

mới trên ứng dụng có thể tốn thời gian và cần tài nguyên xử lý cao hơn.

Đề xuất một số hướng phát triển nghiên cứu như: mở rộng tập dữ liệu đặc biệt đối với một số loại cây thuốc đã có tuy nhiên tỷ lệ dự đoán đúng thấp; cải tiến tối ưu hóa mô hình bằng cách áp dụng thêm các kiến trúc tối ưu hơn hoặc thuật toán phân loại để cải thiện tốc độ và tiết kiệm tài nguyên mà vẫn duy trì

độ chính xác cao; cải thiện khả năng xử lý hình ảnh thực tế trong các môi trường không kiểm soát như ánh sáng yếu, nhiễu động và hình ảnh chất lượng thấp bằng các kỹ thuật xử lý ảnh nâng cao; triển khai hệ thống cập nhật mô hình tự động khi có dữ liệu mới hoặc khi phát hiện các loại cây mới trong hệ thống, đảm bảo mô hình luôn được cải thiện theo thời gian.

TÀI LIỆU THAM KHẢO (REFERENCES)

- Ali, A. H., Yaseen, M. G., Aljanabi, M., & Abed, S. A. (2023). Transfer Learning: A New Promising Techniques. *Mesopotamian Journal of Big Data*, 29–30. <https://doi.org/10.58496/MJBD/2023/004>
- Bipin Nair, B. J., Arjun, B., Abhishek, S., Abhinav, N. M., & Madhavan, V. (2024). Classification of Indian Medicinal Flowers using MobileNetV2. *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, 1512–1518. <https://doi.org/10.23919/INDIACom61295.2024.10498274>
- Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807. <https://doi.org/10.1109/CVPR.2017.195>
- Gupta, J., Pathak, S., & Kumar, G. (2022). Deep Learning (CNN) and Transfer Learning: A Review. *Journal of Physics: Conference Series*, 2273(1), 012029. <https://doi.org/10.1088/1742-6596/2273/1/012029>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. <https://doi.org/10.1109/CVPR.2016.90>
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. https://www.researchgate.net/publication/316184205_MobileNets_Efficient_Convolutional_Neural_Networks_for_Mobile_Vision_Applications
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Neural Information Processing Systems*.
- Koo, K.-M., & Cha, E.-Y. (2017). Image recognition performance enhancements using image normalization. *Human-Centric Computing and Information Sciences*, 7(1), 33. <https://doi.org/10.1186/s13673-017-0114-5>
- Kornblith, S., Shlens, J., & Le, Q. V. (2018). *Do Better ImageNet Models Transfer Better?* <https://doi.org/10.1109/CVPR.2019.00277>
- Liu, S., & Deng, W. (2015). Very deep convolutional neural network based image classification using small training sample size. *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 730–734. <https://doi.org/10.1109/ACPR.2015.7486599>
- Mikolajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. *2018 International Interdisciplinary PhD Workshop (IIPHDW)*, 117–122. <https://doi.org/10.1109/IIPHDW.2018.8388338>
- Nguyen, T. Q., & Truong, V. H. (2020). Medicinal Plant identification in the wild by using CNN. *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 25–29. <https://doi.org/10.1109/ICTC49870.2020.9289480>
- Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Research*, 2(1), 37–63.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>
- Sugiarto, D., Siswanto, J., Naufal, M. F., & Idrus, B. (2023). Mobile Application for Medicinal Plants Recognition from Leaf Image Using Convolutional Neural Network. *Indonesian Journal of Information Systems*, 5(2), 43–56. <https://doi.org/10.24002/ijis.v5i2.6633>
- Sun, Y., Liu, Y., Wang, G., & Zhang, H. (2017). Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, 2017, 1–6. <https://doi.org/10.1155/2017/7361042>

- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning*. <http://proceedings.mlr.press/v97/tan19a.html>
- Vo, A. H., Dang, H. T., Nguyen, B. T., & Pham, V.-H. (2019). Vietnamese Herbal Plant Recognition Using Deep Convolutional Features. *International Journal of Machine Learning and Computing*, 9(3), 363–367. <https://doi.org/10.18178/ijmlc.2019.9.3.811>
- World Bank Group. (2023). *Rural population (% of total population) - Viet Nam*. <https://Data.Worldbank.Org/Indicator/SP.RUR.TOTL.ZS?Locations=VN>.
- World Health Organization Western Pacific. (1990). *Medicinal Plants in Viet Nam*. <https://www.who.int/publications/i/item/9290611014>