

AI trong chăm sóc sức khỏe: phân loại ung thư da và giải thích thông qua xác định vùng tổn thương

Phan Mạnh Thường, Đoàn Thiện Minh*, Nguyễn Đức Thông, Nguyễn Thị Hồng Dịu

Trường Đại học Lạc Hồng, TP Biên Hòa, Việt Nam

*dtminh@lhu.edu.vn

Tóm tắt

Với tỷ lệ ung thư da ngày càng gia tăng, nghiên cứu này tập trung nâng cao hiệu quả và tính minh bạch trong chẩn đoán ung thư da tự động thông qua sự kết hợp hai yếu tố: phân loại tổn thương da (lành tính hay ác tính) và xác định vị trí tổn thương bằng khoanh vùng (bounding box) và phân đoạn (segmentation). Phân đoạn không chỉ xác định chính xác vùng da bị ảnh hưởng mà còn hỗ trợ giải thích lý do mô hình phân loại đưa ra kết luận, từ đó tăng độ tin cậy cho mô hình. Phương pháp sử dụng các mô hình học sâu tiên tiến như Segment Anything Model và YOLOv8 để phát hiện và phân đoạn tổn thương, kết hợp với các mạng CNN như MobileNetV3, VGG16, ResNet50 và DenseNet121 để phân loại. Các mô hình được huấn luyện trên tập dữ liệu da liễu công khai ISIC 2023. Kết quả thử nghiệm cho thấy DenseNet121 đạt độ chính xác 89 % và điểm F1 (F1-Score) 89 % trong phân loại, trong khi YOLOv8 đạt độ chính xác 92,25 % và điểm F1 91 %. Sự kết hợp giữa phân loại, khoanh vùng và phân đoạn không chỉ tạo ra một hệ thống AI giải thích được (Explainable AI) mà còn tăng cường độ tin cậy, hỗ trợ phát hiện sớm và đưa ra quyết định chính xác hơn trong chăm sóc sức khỏe da liễu.

Nhận 15/06/2025
Được duyệt 26/07/2025
Công bố 30/10/2025

Từ khóa

ung thư da; học sâu; khoanh vùng; phân loại; trí tuệ nhân tạo

© 2025 Journal of Science and Technology - NTTU

1. Giới thiệu

Sự gia tăng nhanh chóng của dữ liệu hình ảnh trong những năm qua đã đặt ra nhu cầu cấp thiết về các phương pháp xử lý tự động hiệu quả và chính xác. Cùng với sự ra đời của các bộ xử lý mạnh mẽ như bộ xử lý đồ họa (Graphics Processing Unit-GPU), bộ xử lý tensor (Tensor Processing Unit-TPU), bên cạnh đó, sự phát triển của các kiến trúc hạ tầng đám mây, đại diện tiêu biểu cho trí thông minh nhân tạo, mạng nơ-ron tích chập (Convolutional Neural Network – CNN) đã nổi lên như một bước tiến cách mạng trong lĩnh vực học sâu, nhờ vào kiến trúc đặc thù mô phỏng cơ chế xử lý thị giác của con người. Với các lớp tích chập chuyên biệt, mạng CNN cho phép hệ thống học và nhận diện các đặc trưng cục bộ trong ảnh thông qua cơ chế chia sẻ trọng số là một điểm khác biệt quan trọng so với các mạng nơ-ron truyền thống. Cách tiếp cận này không chỉ giúp giảm đáng kể độ phức tạp tính toán mà còn tăng cường khả năng khái quát hóa, đặc biệt hiệu quả khi xử lý ảnh có độ phân giải cao. Tính hiệu quả và khả năng mở rộng của CNN đã được minh chứng qua nhiều ứng dụng thực tiễn, tiêu biểu như các hệ thống chẩn đoán y tế sử dụng AI [1] nơi mà việc chuyển hóa dữ liệu ảnh đầu vào thành thông tin có thể hành động là yếu tố then chốt.

Cùng sự phát triển hữu ích của khoa học kỹ thuật, sự bất lợi do ô nhiễm môi trường, suy giảm tầng ozone, biến đổi khí hậu và lối sống thiếu lành mạnh cũng ra đời làm gia tăng tỷ lệ đột biến gen, một trong những nguyên nhân chính gây ung thư. Ung thư da, một trong những loại ung thư phổ biến nhất, đang có xu hướng tăng nhanh trên toàn cầu do sự suy giảm nồng độ ozone, làm tăng tiếp xúc với bức xạ tử ngoại (UV) là yếu tố có liên hệ trực tiếp đến sự phát triển ung thư da [2]. Hàng triệu ca ung thư da được phát hiện mỗi năm, với tỷ lệ gia tăng đáng kể. Loại ung thư này thường tiến triển âm thầm ở giai đoạn đầu, gây khó khăn trong chẩn đoán và điều trị nếu không được phát hiện sớm. Các phương pháp chẩn đoán truyền thống dựa trên đánh giá hình ảnh bằng mắt thường của bác sĩ da liễu, đòi hỏi kinh nghiệm cao và dễ bị ảnh hưởng bởi yếu tố chủ quan. Do đó, nhu cầu về các phương pháp chẩn đoán tự động, chính xác và nhất quán ngày càng cấp thiết. Mạng CNN đã được ứng dụng hiệu quả trong thị giác máy tính y tế [3, 4], mang lại tiềm năng lớn trong việc phát hiện và phân loại tổn thương da dựa trên hình ảnh lâm sàng. Nghiên cứu này nhằm đề xuất một hệ thống chẩn đoán tự động hóa tiên tiến, tích hợp đồng thời hai nhiệm vụ: phân loại nhị phân (tổn thương lành tính hoặc ác tính),

khoanh vùng và phân đoạn chính xác các khu vực nghi ngờ trên ảnh da liễu. Mô hình được thiết kế với mục tiêu hỗ trợ các chuyên viên xử lý ảnh y tế, bác sĩ da liễu và các chuyên gia lâm sàng trong việc đưa ra quyết định chẩn đoán nhanh chóng và chính xác. Đặc biệt, hệ thống không chỉ cung cấp kết quả phân loại, mà còn trực quan hóa vùng tổn thương đáng ngờ thông qua khả năng khoanh vùng, phân đoạn giải thích được — một yếu tố giúp tăng độ tin cậy trong thực tế lâm sàng.

Với độ chính xác cao và khả năng giải thích rõ ràng, giải pháp được kỳ vọng sẽ góp phần cải thiện tiên lượng bệnh và hỗ trợ xây dựng phác đồ điều trị hiệu quả. Nghiên cứu hướng tới việc phát triển các công cụ chẩn đoán thông minh, có khả năng tích hợp vào quy trình chăm sóc y tế hiện đại, phù hợp với nhu cầu ứng dụng trong thực tiễn lâm sàng.

2. Phương pháp nghiên cứu

Sự gia tăng nhanh chóng của dữ liệu hình ảnh trong thập kỷ qua đã đặt ra nhu cầu cấp thiết về các phương pháp xử lý tự động hiệu quả và chính xác. Trong bối cảnh đó, CNN đã nổi lên như một bước tiến cách mạng trong lĩnh vực học sâu, nhờ vào kiến trúc đặc thù mô phỏng cơ chế xử lý thị giác.

Để đạt được mục tiêu phát triển hệ thống chẩn đoán ung thư da tự động, nghiên cứu này đề xuất một quy trình xử lý hình ảnh y tế dựa trên kỹ thuật học sâu, tập trung vào hai nhiệm vụ chính: phân loại tổn thương da (lành tính hoặc ác tính) và khoanh vùng các khu vực nghi ngờ trên hình ảnh lâm sàng. Nghiên cứu được tiến hành tại phòng thực hành, Khoa Công nghệ thông tin, Trường Đại học Lạc Hồng từ tháng 6 năm 2024 đến nay.

Quy trình nghiên cứu bao gồm các bước sau: (1) Thu thập và tiền xử lý tập dữ liệu hình ảnh da từ cơ sở dữ liệu ảnh công khai, bao gồm hơn 3 000 ảnh nhãn Lành tính (Benign) và ác tính (Malignant) được chú thích bởi chuyên gia về da liễu. (2) Huấn luyện mô hình phân loại sử dụng CNN như DenseNet121 [5], Resnet50 [6, 7], VGG16 [8, 9], Mobilenetv3 [10] để phân biệt tổn thương lành tính và ác tính. (3) Triển khai hai kỹ thuật khoanh vùng: You Only Look Once (YOLO) phiên bản 8 và Segment Anything Model (SAM) để phát hiện nhanh các vùng tổn thương, tinh chỉnh biên dạng tổn thương với độ chi tiết cao. Chỉ tiêu đánh giá bao gồm độ chính xác, độ nhạy, độ đặc hiệu cho phân loại, và chỉ số IoU (Intersection over Union) cho khoanh vùng.

Sự kết hợp giữa phân loại và khoanh vùng nhằm nâng cao độ chính xác trong nhận diện và trực quan hóa

vùng tổn thương, hỗ trợ chuyên gia xử lý hình ảnh, bác sĩ da liễu trong việc chẩn đoán nhanh và hiệu quả. Phương pháp này được thiết kế để phù hợp với thực tiễn lâm sàng, với tiềm năng ứng dụng tại các cơ sở y tế.

2.1. Bộ dữ liệu về da

Bộ dữ liệu được sử dụng trong nghiên cứu này được xây dựng từ kho lưu trữ hình ảnh da liễu thuộc sáng kiến International Skin Imaging Collaboration (ISIC) 2023 – một chương trình hợp tác toàn cầu nhằm chuẩn hóa việc chẩn đoán và phát hiện sớm ung thư da. Từ hơn 13 000 ảnh soi da (dermoscopy), 2 983 hình ảnh đã được chọn lọc, đảm bảo tỷ lệ phân bố cân bằng giữa hai nhóm tổn thương: lành tính (benign) và ác tính (malignant).

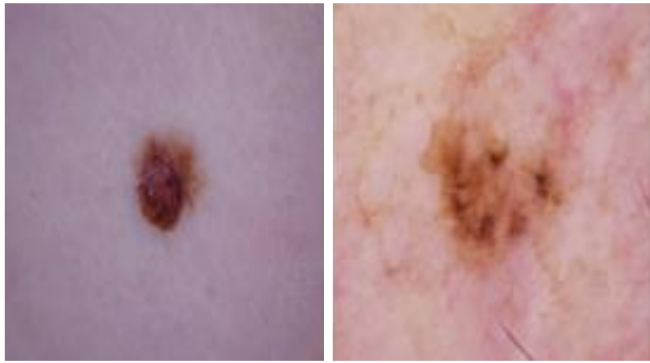
Tất cả các ảnh đều có độ phân giải tối thiểu $1\,024 \times 1\,024$ pixel, được kiểm duyệt chất lượng theo các tiêu chí: độ sắc nét, điều kiện ánh sáng, tiêu cự, và đã được loại bỏ các thông tin nhận dạng cá nhân. Tiêu chí loại trừ bao gồm các ảnh mờ, ảnh có vùng tổn thương bị che khuất trên 15 %, hoặc thiếu xác nhận chẩn đoán lâm sàng.

Trước khi đưa vào huấn luyện, dữ liệu được tiền xử lý bằng các kỹ thuật như chuẩn hóa kích thước, tăng cường độ tương phản, và tăng cường dữ liệu (data augmentation) với các phép biến đổi như xoay, lật và điều chỉnh độ sáng. Toàn bộ quy trình được thực hiện bằng ngôn ngữ lập trình Python, nhằm đảm bảo tính nhất quán và tối ưu hóa dữ liệu đầu vào cho các mô hình học sâu CNN như DenseNet121, Resnet50, VGG16, Mobilenetv3 dùng trong phân loại, cũng như You Only Look Once (YOLO) và SAM dùng trong khoanh vùng, phân đoạn.

Với gần 3 000 ảnh được gán nhãn rõ ràng, bộ dữ liệu này đáp ứng được các khuyến nghị trong huấn luyện mô hình học sâu (tối thiểu 1 000 ảnh cho mỗi lớp), giúp đảm bảo tính đa dạng, cân bằng, và tin cậy trong quá trình huấn luyện và đánh giá mô hình.

2.2. Mô hình phân loại

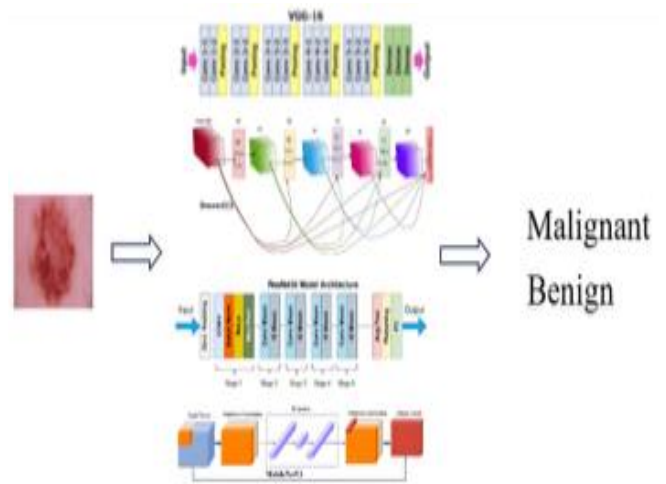
Phương pháp phân loại triển khai trong nghiên cứu dùng phương pháp phân loại nhị phân để xác định tổn thương da là lành tính hoặc ác tính, sử dụng tập dữ liệu hình ảnh da liễu công khai (ISIC) 2023 [11]. Phương pháp áp dụng kỹ thuật học chuyển giao (Transfer Learning) để khai thác các mô hình mạng nơ-ron tích chập CNN được huấn luyện trước, nhằm tối ưu hóa hiệu suất phân loại trên tập dữ liệu giới hạn.



Hình 1. Hình ảnh tổn thương da ác tính [11]

Quy trình phân loại bao gồm các bước sau: (1) Tiền xử lý hình ảnh đầu vào, bao gồm chuẩn hóa kích thước, điều chỉnh độ tương phản, và loại bỏ nhiễu để tối ưu hóa dữ liệu cho bài toán phân loại. (2) Huấn luyện mô hình sử dụng các kiến trúc CNN tiên tiến, bao gồm Visual Geometry Group 16 (VGG16), Residual Network 50 (ResNet50), Dense Convolutional Network (DenseNet121), và MobileNetV3. Trong đó, VGG16 có cấu trúc sâu với các lớp tích chập tuần tự, ResNet50 sử dụng các khối dư (residual blocks) để giảm thiểu mất mát gradient, DenseNet121 tận dụng kết nối dày đặc giữa các lớp để tăng cường truyền tải đặc trưng, còn MobileNetV3 được thiết kế nhẹ với các khối tích chập phân tách theo chiều sâu (depthwise separable convolutions), phù hợp cho các ứng dụng trên thiết bị có tài nguyên hạn chế. (3) Đánh giá mô hình dựa trên các chỉ số hiệu suất như độ chính xác (accuracy), độ nhạy (sensitivity), và độ đặc hiệu (specificity). Mô hình tối ưu được lựa chọn dựa trên kết quả đánh giá để triển khai phân loại, cung cấp dự đoán về loại tổn thương da từ hình ảnh đầu vào.

Hình 2 minh họa một quy trình trực quan và minh bạch, trình bày việc áp dụng các mô hình học chuyển giao (transfer learning) bao gồm VGG16, DenseNet121, ResNet50 và MobileNetV3, được triển khai lần lượt để so sánh và đánh giá độ chính xác trong phân loại tổn thương da. Mỗi mô hình được mô tả như một cấu trúc phân tích tinh vi, trong đó hình ảnh đầu vào được xử lý qua các lớp tích chập và khối đặc trưng được tối ưu hóa, dẫn đến kết quả phân loại chính xác, phân biệt rõ ràng giữa "Ác tính" và "Lành tính".



Hình 2. Quy trình thực hiện phân loại nhị phân

2.3. Mô hình khoanh vùng, phân đoạn nhằm giải thích

Để hỗ trợ phân loại nhị phân và xác định chính xác các khu vực tổn thương da từ bộ dữ liệu (ISIC) 2023, nghiên cứu này triển khai hai kỹ thuật phân đoạn YOLO phiên bản 8 và SAM. Các kỹ thuật này được tích hợp vào quy trình thực nghiệm để phân đoạn và xác định các vùng nghi ngờ, cung cấp thông tin bổ sung cho hệ thống phân loại tổn thương lành tính hoặc ác tính.

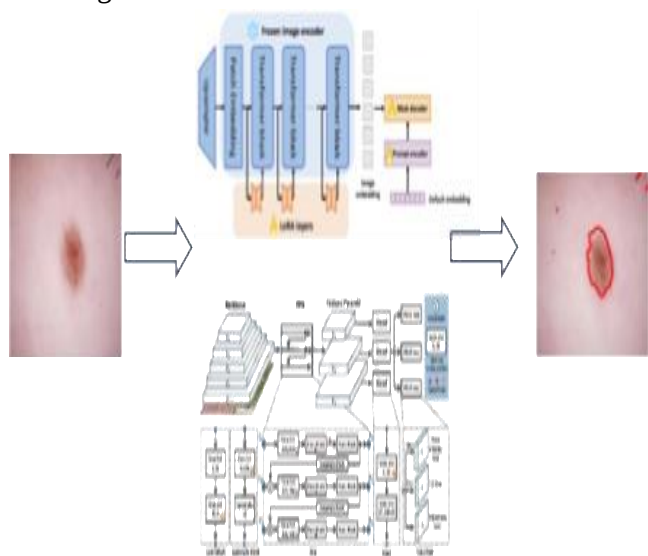
YOLO phiên bản 8 được sử dụng để phát hiện nhanh các vùng tổn thương tiềm năng trong thời gian thực, nhờ kiến trúc tối ưu hóa tốc độ và độ chính xác. SAM, với khả năng phân đoạn cấp độ pixel, được áp dụng để tinh chỉnh biên dạng tổn thương, đảm bảo chi tiết cao trong các trường hợp phức tạp. Cả hai mô hình được huấn luyện trên tập dữ liệu bệnh về da ISIC 2023. Kết quả vùng da tổn thương được khoanh vùng từ YOLO, SAM được tích hợp với kết quả phân loại nhị phân để đánh giá mức độ tương quan giữa vùng tổn thương và nhãn chẩn đoán (lành tính/ác tính).

Hiệu suất của YOLO và SAM được so sánh thông qua các chỉ số định lượng, bao gồm IoU, độ chính xác phân đoạn (segmentation accuracy), và mức độ cải thiện hiệu suất phân loại tổng thể.

Hình 3 bên dưới minh họa một quy trình trực quan và minh bạch, trình bày việc áp dụng các mô hình học chuyển giao (transfer learning) như SAM và YOLOv8 để thực hiện khoanh vùng và phân đoạn vị trí tổn thương da, nhằm giải thích kết quả phân loại. Hình ảnh đầu vào được xử lý qua một cấu trúc tinh vi, với bộ mã hóa hình ảnh cố định (Frozen Image Encoder) và các lớp Transformer kết hợp cùng các lớp LoRA để tinh chỉnh đặc trưng. Dữ liệu tiếp tục được mã hóa qua Prompt Encoder và Default Embedding, sau đó được giải mã bởi Mask Decoder để tạo ra đường viền đỏ

chính xác quanh vùng tổn thương. Quy trình này tối ưu hóa luồng dữ liệu qua từng giai đoạn, giúp người đọc dễ dàng nhận diện cách các mô hình SAM và YOLOv8 phối hợp để xác định và giải thích vị trí da bệnh một cách khoa học và đáng tin cậy.

Mô hình có độ chính xác vượt trội hơn cả sẽ được lựa chọn làm phương pháp khoanh vùng chính thức để triển khai trong hệ thống ứng dụng thực tế. Cách tiếp cận này đảm bảo tính minh bạch, khả năng giải thích và tối ưu hóa hiệu quả chẩn đoán trong môi trường lâm sàng.



Hình 3. Quy trình thực hiện khoanh vùng giải thích lý do bệnh.

3. Kết quả và thảo luận

Qua kết quả đánh giá hiệu suất của bốn kiến trúc mạng: VGG16, ResNet50, DenseNet121 và MobileNetV3 để phân loại nhị phân các tổn thương da thành các loại ác tính và lành tính bằng cách sử dụng tập dữ liệu hình ảnh y tế về ung thư da. Tăng cường dữ liệu và học chuyển giao đã được áp dụng để tối ưu hóa hiệu suất mô hình. Hiệu quả phân loại đã được đánh giá thông qua số liệu chính xác, mang lại kết quả sau: VGG16 đạt độ chính xác 88 %, chứng tỏ khả năng trích xuất tính năng mạnh mẽ nhưng đòi hỏi nhiều tài nguyên tính toán do kiến trúc sâu của nó.

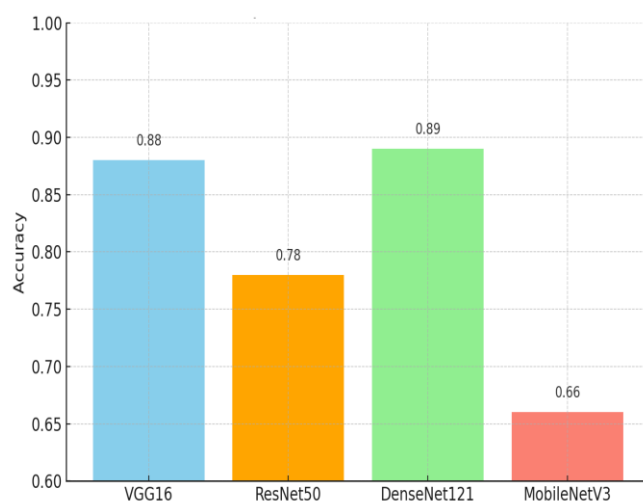
ResNet50 đạt độ chính xác là 78 %, cho thấy những hạn chế trong việc nắm bắt các đặc điểm bệnh lý phức tạp, có thể là do khuôn khổ học tập còn sót lại của nó. DenseNet121 đạt độ chính xác cao nhất với mức 89 %, nhờ lợi thế vượt trội từ kết nối dày đặc giữa các lớp, giúp tối ưu hóa việc truyền bá và học tập các đặc trưng một cách hiệu quả. Mặc dù yêu cầu tính toán cao hơn, cấu trúc này cho phép giảm số lượng tham số cần thiết, từ đó tăng cường hiệu suất và độ tin cậy trong phân loại tổn thương da.

MobileNetV3 ghi nhận độ chính xác là 67 %, phản ánh thiết kế nhẹ được tối ưu hóa cho môi trường hạn chế tài nguyên, chẳng hạn như thiết bị di động, với cái giá phải trả là khả năng phân biệt giảm.

Bảng 1: Hiệu suất phân loại của các mô hình phân loại trong chẩn đoán ung thư da

Model	Accuracy	Specificity	F1-Score	AUC
Vgg16	0,88	0,85	0,89	0,96
Resnet50	0,78	0,7	0,8	0,86
Densenet121	0,89	0,87	0,89	0,97
MobileNetv3	0,66	0,925	0,55	0,84

Những kết quả phân loại làm nổi bật sự đánh đổi giữa độ chính xác và hiệu quả tính toán. DenseNet121 và VGG16, với độ chính xác lần lượt là 89 % và 88 %, rất phù hợp để triển khai trên các hệ thống điện toán hiệu suất cao, nơi tính khả dụng của tài nguyên hỗ trợ kiến trúc phức tạp của chúng. Ngược lại, MobileNetV3, mặc dù có độ chính xác thấp hơn, nhưng lại lý tưởng cho các thiết bị di động và thiết bị biên do thiết kế hiệu quả của nó, cho phép triển khai thực tế trong các cài đặt hạn chế về tài nguyên. Hiệu suất thấp hơn của ResNet50 cho thấy cần phải tối ưu hóa thêm để cân bằng giữa độ chính xác và nhu cầu về tài nguyên. Kết quả việc vận dụng các mô hình phân loại này nhấn mạnh tầm quan trọng của việc lựa chọn kiến trúc dựa trên bối cảnh triển khai, với các mô hình sử dụng nhiều tài nguyên ưu tiên các ứng dụng có độ chính xác cao trên phần cứng mạnh mẽ và các mô hình nhẹ hỗ trợ chẩn đoán dựa trên thiết bị di động.



Hình 4. Biểu đồ so sánh độ chính xác (Accuracy) giữa các mô hình phân loại ung thư da

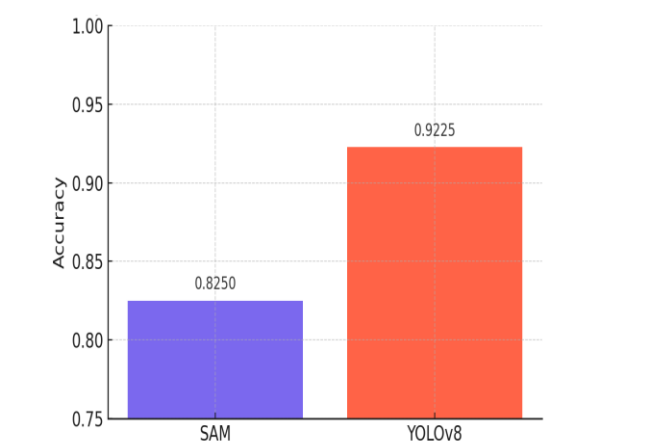
Trong nhiệm vụ khoanh vùng tổn thương ác tính trên ảnh da liễu, nghiên cứu tiến hành so sánh hiệu suất

giữa hai mô hình SAM và YOLOv8. Các mô hình được đánh giá dựa trên các chỉ số chuẩn bao gồm độ chính xác, độ nhạy, độ đặc hiệu, F1-score, diện tích dưới đường cong ROC (AUC) và IoU.

Bảng 2: Hiệu suất phân đoạn của mô hình SAM trong khoanh vùng u ác tính ở da

Model	Accuracy	Sensitivity	Specificity	F1-score	AUC	IoU
SAM	0,825	0,700	0,950	0,800	0,9008	0,6667
YOLOv8	0,9225	0,865	0,980	0,9178	0,9225	0,9995

Kết quả xác nhận YOLOv8 là mô hình vượt trội SAM trong nhiệm vụ khoanh vùng, đặc biệt phù hợp với các ứng dụng lâm sàng đòi hỏi độ chính xác cao, tốc độ xử lý nhanh và khả năng tổng hợp mạnh. Điều đặc biệt đáng kể trong IoU và các số khác cho thấy YOLOv8 có ưu thế về khả năng định vị biên dạng phức tạp. Việc tích hợp YOLOv8 với mô hình phân loại DenseNet121 mở ra tiềm năng phát triển các hệ thống hỗ trợ mong đợi sự hợp nhất, giải pháp thích hợp và khả thi trong việc giải thích mô hình hộp đen của việc phân loại.



Hình 5. Biểu đồ so sánh độ chính xác (Accuracy) giữa các mô hình giải thích Sam, YoloV8.

4. Kết luận

Nghiên cứu này trình bày một phương pháp chẩn đoán ung thư da tự động, tích hợp các kỹ thuật học sâu tiên tiến để thực hiện đồng thời phân đoạn và phân loại tổn thương từ hình ảnh lâm sàng. Trong số các mô hình mạng nơ-ron tích chập được đánh giá (VGG16,

ResNet50, MobileNetV3, và DenseNet121), DenseNet121 đạt hiệu suất vượt trội với độ chính xác 89 %, nhờ vào thiết kế kết nối dày đặc tối ưu hóa việc trích xuất đặc trưng từ dữ liệu hình ảnh y tế phức tạp. Đồng thời, YOLOv8 thể hiện ưu thế với IoU đạt 99,95 % và độ chính xác 92,25 %, phù hợp để giải thích lý do phân loại.

Phương pháp đề xuất nâng cao tính minh bạch và độ tin cậy trong hệ thống, đặc biệt trong việc phát hiện các tổn thương ác tính phức tạp, từ đó hỗ trợ hiệu quả quá trình ra quyết định lâm sàng. Với tiềm năng ứng dụng thực tiễn cao, nghiên cứu này hứa hẹn mang lại giá trị thiết thực cho các bác sĩ da liễu và chuyên gia chăm sóc sức khỏe thông qua các bằng chứng định lượng đáng tin cậy.

Dẫu vậy, nghiên cứu vẫn đối mặt với những hạn chế nhất định, bao gồm sự phụ thuộc vào chất lượng dữ liệu đào tạo và nhu cầu tài nguyên tính toán lớn, có thể gây cản trở khi triển khai trên các thiết bị công suất thấp trong thực tế. Để vượt qua những hạn chế hiện tại, cần thực hiện chiến lược thu thập dữ liệu lâm sàng đa dạng và thường xuyên từ nhiều cơ sở y tế, nhằm tạo ra một bộ dữ liệu phong phú không chỉ về số lượng bệnh nhân mà còn về số lượng hình ảnh của từng cá nhân. Việc ghi lại nhiều hình ảnh ở các thời điểm khác nhau cho cùng một bệnh nhân sẽ hỗ trợ mô hình học sâu nhận diện hiệu quả hơn các biến đổi của tổn thương theo thời gian và điều kiện ánh sáng, từ đó cải thiện độ chính xác, khả năng tổng quát hóa và độ tin cậy của hệ thống chẩn đoán. Về triển vọng nghiên cứu, có thể xem xét tích hợp mô hình với các hệ thống chẩn đoán thời gian thực hoặc thử nghiệm mở rộng trên các loại ung thư khác để nâng cao hiệu quả và phạm vi ứng dụng trong thực tiễn lâm sàng.

Lời cảm ơn

Chúng tôi xin bày tỏ lòng biết ơn sâu sắc đến Trường Đại học Lạc Hồng đã tạo điều kiện thuận lợi góp phần giúp chúng tôi hoàn thiện bài báo khoa học này. Đồng thời, chúng tôi trân trọng cảm ơn Trường Đại học Nguyễn Tất Thành, Thành phố Hồ Chí Minh tạo cơ hội cho chúng tôi được trình bày, chia sẻ kết quả nghiên cứu trước cộng đồng học thuật.

Tài liệu tham khảo

1. Johnson Oyeniyi, Paul Oluwaseyi. (2024). Emerging Trends in AI-Powered Medical Imaging: Enhancing Diagnostic Accuracy and Treatment Decisions. International Journal of Enhanced Research in Science, Technology & Engineering ISSN: 2319-7463, Vol. 13 Issue 4, April-2024, Impact Factor: 8.375.



2. Amdad Hossain Roky và cộng sự. (2025). Overview of skin cancer types and prevalence rates across continents. ELSEVIER, Cancer Pathogenesis and Therapy 3 (2025) 89-100.
3. Ahmad Waleed Salehi và cộng sự. (2023). A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. MDPI. Sustainability 2023, 15(7), 5930.
4. Jang Hyung Lee, Kwang Gi Kim. (2018). Applying Deep Learning in Medical Images: The Case of Bone Age Estimation. Healthc Inform Res. 24(1):86-92, PMID: 29503757 PMCID: PMC5820091.
5. Pooja Pradeep Dalvi, Damodar Reddy Edla, B. R. Purushothama. (2024). DenseNet-121 Model for Diagnosis of COVID-19 Using Nearest Neighbour Interpolation and Adam Optimizer. Wireless Personal Communications Wireless Personal Communications 137(3):1-19.
6. Suvaditya Mukherjee. (2022). The Annotated ResNet – 50.
7. V. Lakide and V. Ganesan. (2024). Precise Lung Cancer Prediction using ResNet – 50 Deep Neural Network Architecture. j.electron.electromedical.eng.med.inform, Vol. 7, No. 1, pp. 38-46, Nov.
8. Sachin Mohan. (2020). Keras Implementation of VGG16 Architecture from Scratch with Dogs Vs Cat Data Set
9. Marjan Kia và cộng sự. (2024). Innovative fusion of VGG16, MobileNet, EfficientNet, AlexNet, and ResNet50 for MRI-based brain tumor identification. Iran J Comput Sci 8, 185-215.
10. Ahmed Firas Majeed và cộng sự. (2024). Multi-Class Brain Lesion Classification Using Deep Transfer Learning With MobileNetV3. IEEE Access PP(99):1-1.
11. Hiệp hội hình ảnh da quốc tế ISIC. (2023). Skin Cancer: Malignant vs. Benign.

AI in Healthcare: Skin Cancer Classification and Explanation through Lesion Localization

Phan Manh Thuong, Doan Thien Minh*, Nguyen Duc Thong, Nguyen Thi Hong Diu

Lac Hong University, Bien Hoa City, Vietnam

*dtminh@lhu.edu.vn

Abstract With the rising incidence of skin cancer, this study focuses on enhancing the efficiency and transparency of automated skin cancer diagnosis by integrating two key components: skin lesion classification (benign or malignant) and lesion localization through bounding box detection and segmentation. Segmentation not only precisely identifies the affected skin areas but also aids in explaining the reasoning behind the classification model's conclusions, thereby increasing model reliability. The approach employs advanced deep learning models such as the Segment Anything Model and YOLOv8 for lesion detection and segmentation, combined with CNN architectures like MobileNetV3, VGG16, ResNet50, and DenseNet121 for classification. The models were trained on the publicly available ISIC 2023 dermatology dataset. Experimental results show that DenseNet121 achieved an accuracy of 89 % and an F1-Score of 89 % in classification, while YOLOv8 attained an accuracy of 92.25 % and an F1-Score of 91 %. The integration of classification, bounding box localization, and segmentation not only creates an explainable AI system but also enhances reliability, enabling early detection and more accurate decision-making in dermatological healthcare.

Keywords: skin cancer; deep learning; segmentation; classification; artificial intelligence