



Bài báo nghiên cứu

ỨNG DỤNG MÔ HÌNH HỌC SÂU TRONG CHẨN ĐOÁN UNG THƯ PHỔI

Trịnh Huy Hoàng^{1*}, Trần Văn Tuấn¹,

Nguyễn Lê Minh Ngọc¹, Nguyễn Dương Quốc Bảo¹, Lê Hồng Thúy Vũ²

¹Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam

²Khoa Công nghệ thông tin, Trường Đại học Ngoại ngữ – Tin học Thành phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ: Trịnh Huy Hoàng – Email: hoangth@hcmue.edu.vn

Ngày nhận bài: 27-5-2025; Ngày nhận bài sửa: 02-6-2025; Ngày duyệt đăng: 22-10-2025

TÓM TẮT

Ngày nay với sự phát triển của y học hiện đại, bệnh nhân ung thư phổi ngày càng được phát hiện sớm. Nghiên cứu đề xuất một phương pháp mới bằng cách cải tiến kiến trúc U-Net truyền thống, loại bỏ hoàn toàn phần giải mã (decoder) và thay thế bằng một nhánh phân loại chuyên biệt ở cuối phần mã hóa (encoder). Sự thay đổi này giúp mô hình tập trung vào việc trích xuất đặc trưng tổng thể của ảnh cho bài toán phân loại, thay vì tái tạo lại bản đồ phân đoạn ở mức pixel. Điều này không chỉ giúp giảm đáng kể độ phức tạp tính toán mà vẫn duy trì khả năng học hỏi các đặc trưng sâu sắc từ ảnh y khoa để phát hiện và chẩn đoán bệnh ung thư phổi. Dựa trên tập dữ liệu IQ-OTH/NCCD, mô hình U-Net cải tiến cho kết quả vượt trội so với các phương pháp học sâu khác như VGG-16, ResNet-50, NasNet Mobile và ViT.

Từ khóa: phương pháp học sâu; chẩn đoán ung thư phổi; mô hình U-Net

1. Giới thiệu

Ung thư phổi được xem là nguyên nhân hàng đầu gây tử vong liên quan đến ung thư trên phạm vi toàn thế giới. Theo Globocan (IARC, 2022), tại Việt Nam trong năm 2022 ghi nhận 24.426 ca mắc mới và 22.597 ca tử vong vì căn bệnh này ở cả hai giới. Do cấu trúc giải phẫu phức tạp của phổi, việc phát hiện và chẩn đoán chính xác ung thư phổi vẫn là một thách thức lớn đối với ngành y khoa toàn cầu. Vì vậy, việc phát triển các mô hình hỗ trợ chẩn đoán tự động là cần thiết, nhằm giảm tải cho đội ngũ y khoa, đồng thời nâng cao độ chính xác cũng như hiệu quả cho quá trình chẩn đoán. Những năm gần đây, nhiều nghiên cứu đã ứng dụng các mô hình học sâu vào nhiệm vụ nhận diện ung thư phổi từ ảnh y khoa, cho thấy tiềm năng lớn trong việc hỗ trợ bác sĩ đưa ra quyết định lâm sàng nhanh chóng và chính xác hơn.

Cite this article as: Trinh, H. H., Tran, V. T., Nguyen, L. M. N., Nguyen, D. Q. B., & Le, H. T.V. (2025). Use of deep learning models for lung cancer diagnosis. *Ho Chi Minh City University of Education Journal of Science*, 22(11), 1993-2005. [https://doi.org/10.54607/hcmue.js.22.11.5018\(2025\)](https://doi.org/10.54607/hcmue.js.22.11.5018(2025))

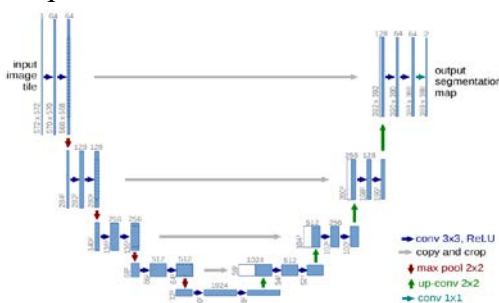
Phân loại ung thư phổi dựa trên đặc điểm mô bệnh học (Howlader et al, 2020) giúp bác sĩ đưa ra chẩn đoán chính xác và định hướng điều trị hiệu quả. Trong đó, ung thư phổi được chia thành hai loại chủ yếu: ung thư phổi tế bào nhỏ (Small Cell Lung Cancer - SCLC) chiếm khoảng 15% tổng số ca bệnh và ung thư phổi không tế bào nhỏ (Non-Small Cell Lung Cancer - NSCLC) chiếm khoảng 85% còn lại. NSCLC tiếp tục được phân thành ba thể chính, trong đó phổ biến nhất là ung thư biểu mô tuyến, chiếm khoảng 40% trường hợp; kế đến là ung thư biểu mô tế bào vảy, chiếm 25-30%; và cuối cùng là ung thư biểu mô tế bào lớn - thể ít gặp nhất, chiếm khoảng 10-15% trường hợp.

Trong nghiên cứu này, chúng tôi đóng góp chủ yếu ở việc cải tiến mô hình U-Net để phân loại ung thư phổi. Sau đó chúng tôi so sánh mô hình cải tiến đề xuất với các mô hình khác để chứng minh hiệu suất của U-Net cải tiến với mô hình phân loại khác.

2. Đối tượng và phương pháp

2.1. Đối tượng

Hiện nay, nhiều giải pháp ứng dụng trí tuệ nhân tạo trong tầm soát ung thư phổi đã được triển khai, giúp phát hiện bệnh nhanh chóng và tối ưu chi phí (Hosseini et al., 2024). Các kiến trúc thường được áp dụng để chẩn đoán có thể kể đến như VGG16 (Prasad et al., 2023; Xu et al., 2024), hay ResNet-50 (Arun et al., 2024; Nasra, 2024; Kumar et al., 2024). Trong đó, mô hình U-Net cũng có những đóng góp to lớn (Kumar et al., 2023; Azad et al., 2024). Trong kiến trúc gốc, U-Net là mạng nơ-ron tích chập có cấu trúc hình chữ U, được thiết kế dành cho các bài toán phân đoạn ảnh y khoa, mục tiêu là phân loại từng pixel trong ảnh đầu vào tương ứng với một vùng giải phẫu hoặc tổn thương cụ thể. Mô hình bao gồm hai phần chính: phần encoder (bên trái) thực hiện trích xuất đặc trưng qua các lớp convolutional và pooling, và phần decoder (bên phải) có chức năng tái tạo lại bản đồ phân đoạn bằng cách sử dụng các lớp upsampling và kết nối skip connection từ encoder.



Hình 1. Kiến trúc U-Net 2D ban đầu (Azad et al., 2024)

Kiến trúc U-Net và các biến thể của nó đã được chứng minh là cực kỳ hiệu quả trong các bài toán phân đoạn ảnh y tế. Chẳng hạn, các nghiên cứu của Tran và cộng sự (2022) đã cải tiến mô hình Attention U-Net để phân đoạn các vùng tổn thương do COVID-19 hay nghiên cứu Wang và cộng sự (2023) để phân đoạn nốt phổi cho thấy khả năng định vị chính xác của mô hình. Tương tự, Truong và Huynh (2023) cũng đã phát triển một phương pháp dựa trên học sâu vừa phân loại vừa phân đoạn tổn thương COVID-19 trên ảnh CT, khẳng

định sức mạnh của các kiến trúc CNN trong phân tích hình ảnh y tế. Tuy nhiên, các nghiên cứu trên chủ yếu tập trung vào bài toán phân đoạn. Hướng tiếp cận của chúng tôi có sự khác biệt: thay vì sử dụng U-Net để định vị pixel, chúng tôi biến đổi kiến trúc này để thực hiện bài toán phân loại toàn bộ hình ảnh. Bằng cách loại bỏ phần giải mã không cần thiết cho nhiệm vụ phân loại, chúng tôi kế thừa khả năng trích xuất đặc trưng mạnh mẽ của U-Net trong khi tạo ra một mô hình gọn nhẹ và hiệu quả hơn cho việc chẩn đoán.

Trong bài toán phân đoạn ảnh y khoa bằng CNN, xác suất phân lớp được tính cho từng pixel bằng softmax. Mỗi pixel có một vector xác suất, và hàm mất mát cross-entropy được áp dụng trên toàn bộ ma trận đầu ra. Trọng số mạng được cập nhật thông qua lan truyền ngược, tương tự như các mô hình CNN phân loại truyền thống.

$$W_{t+1} = W_t - \mu \cdot \Delta L(W_t)$$

trong đó W_t là trọng số tại thời điểm t , μ là learning rate, và ∇L là đạo hàm của hàm mất mát với các trọng số. Mỗi vòng lặp huấn luyện (epoch), mô hình sẽ lan truyền xuôi để tính đầu ra, sau đó lan truyền ngược (backpropagation) để cập nhật trọng số theo hướng giảm sai số.

Sau khi quá trình huấn luyện hoàn tất và mô hình CNN đạt đến trạng thái hội tụ, hiệu suất được đánh giá bằng các chỉ số chính như độ chính xác (accuracy), độ nhạy (recall), độ chính xác (precision) và F1-score.

Đối với bài toán phân loại, độ chính xác được định nghĩa như sau:

$$\text{accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN} \times 100$$

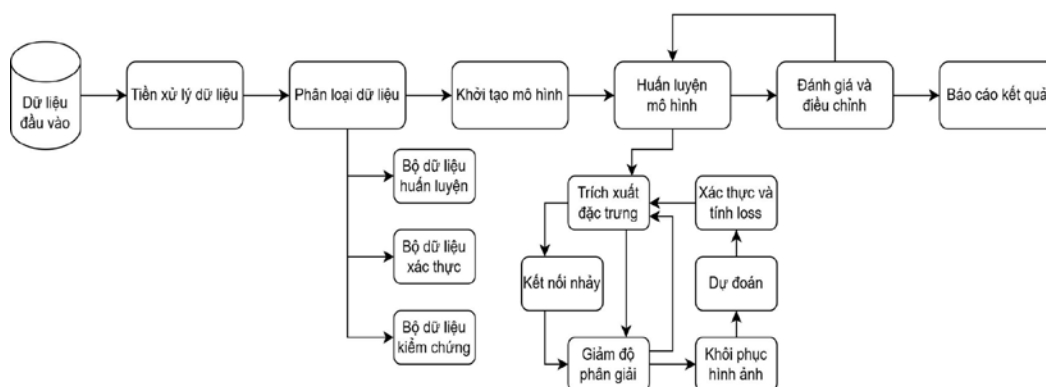
trong đó: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

2.2. Phương pháp

Kiến trúc U-Net gốc được thiết kế cho bài toán phân đoạn ảnh (image segmentation), với mục tiêu là phân loại từng pixel. Do đó, phần giải mã (decoder) có vai trò thiết yếu trong việc khôi phục lại ảnh đầu ra có cùng kích thước với ảnh đầu vào, tạo ra một bản đồ phân đoạn chi tiết. Tuy nhiên, bài toán phân loại ảnh (image classification), mục tiêu là gán một nhãn duy nhất (bình thường, lành tính, ác tính) cho toàn bộ ảnh đầu vào. Trong trường hợp này, việc tái tạo lại ảnh ở mức pixel trở nên không cần thiết và thừa thãi. Phần mã hóa (encoder) của U-Net vốn đã là một mạng nơ-ron tích chập mạnh mẽ, có khả năng trích xuất các đặc trưng hữu ích từ ảnh. Vì vậy, chúng tôi đề xuất loại bỏ phần decoder và chỉ tận dụng phần encoder như một bộ trích xuất đặc trưng hiệu quả, sau đó đưa các đặc trưng này vào một nhánh phân loại đơn giản.

Cách tiếp cận này hoàn toàn có thể áp dụng cho các bộ dữ liệu và bài toán phân loại ảnh y tế khác, không chỉ riêng IQ-OTH/NCCD. Về bản chất, bất kỳ bài toán phân loại ảnh nào cũng có thể hưởng lợi từ một bộ trích xuất đặc trưng mạnh mẽ. Việc sử dụng encoder của U-Net cho phép kế thừa khả năng học các đặc trưng đa cấp độ của nó, trong khi việc loại bỏ decoder giúp mô hình trở nên gọn nhẹ và tập trung hơn vào nhiệm vụ phân loại.

Cấu trúc chi tiết của mô hình U-Net cải tiến:



Hình 2. Quy trình phân loại ung thư phổi của hệ thống xây dựng theo mô hình U-Net.

Hình 2 mô tả quy trình hệ thống sử dụng mô hình U-Net cải tiến của chúng tôi. Dữ liệu đầu vào sau khi qua bước tiền xử lý sẽ được chia thành các tập huấn luyện, xác thực và kiểm chứng. Trong quá trình huấn luyện, mô hình U-Net chỉ sử dụng phần mã hóa (encoder) để trích xuất đặc trưng từ ảnh. Các đặc trưng này sau đó được đưa trực tiếp vào một nhánh phân loại (thay vì phân giải mã) để đưa ra dự đoán. Quá trình huấn luyện được đánh giá và điều chỉnh dựa trên hàm mất mát (loss) và các chỉ số độ chính xác. Quy trình hệ thống của mô hình U-Net cải tiến được sử dụng để phân loại ung thư phổi. Hệ thống bắt đầu với “Dữ liệu đầu vào”. Sau đó, dữ liệu được “Tiền xử lý dữ liệu” và “Phân loại dữ liệu” thành “Bộ dữ liệu huấn luyện”, “Bộ dữ liệu xác thực” và “Bộ dữ liệu kiểm chứng”. Tiếp theo là “Khởi tạo mô hình” và “Huấn luyện mô hình”. Trong quá trình huấn luyện, mô hình thực hiện “Trích xuất đặc trưng” (bằng encoder của U-Net), “Giảm độ phân giải” và “Kết nối nhảy” (trong encoder). Sau đó, “Dự đoán” được đưa ra và “Xác thực và tính loss”. Cuối cùng, có bước “Đánh giá và điều chỉnh” trước khi “Báo cáo kết quả”.

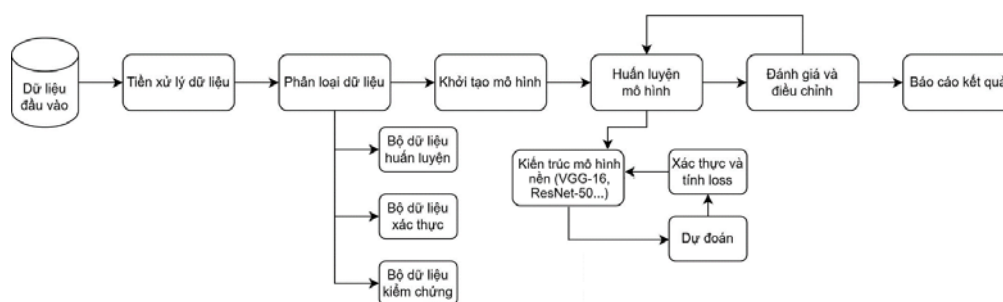
Lưu ý: Trong quy trình hệ thống xây dựng dựa trên mô hình U-Net cải tiến của chúng tôi, bước “Khôi phục hình ảnh” và các kết nối từ “Giảm độ phân giải” và “Kết nối nhảy” đến “Khôi phục hình ảnh” trong kiến trúc U-Net gốc đã được loại bỏ, thay vào đó là một nhánh phân loại trực tiếp từ các đặc trưng được trích xuất bởi encoder. Sau khi ảnh đầu vào được xử lý qua các tầng của encoder và tạo ra một ma trận đặc trưng ở tầng đáy (bottleneck), chúng tôi áp dụng một lớp pooling có tên gọi là GlobalAveragePooling2D để tổng hợp toàn bộ thông tin không gian thành một vector đặc trưng có kích thước cố định. Việc sử dụng lớp pooling này giúp giảm mạnh số lượng tham số so với việc dùng các lớp Dense trực tiếp, đồng thời duy trì được tính khái quát của mô hình. Vector đặc trưng sau đó được đưa qua một hoặc nhiều lớp Dense kết hợp với Dropout để học các biểu diễn trừu tượng hơn và hạn chế hiện tượng overfitting. Cuối cùng, lớp Dense đầu ra với hàm kích hoạt softmax sẽ sinh ra phân phối xác suất tương ứng với ba lớp chẩn đoán (bình thường, u lành tính, u ác tính), cho phép mô hình dự đoán nhãn bệnh lí của ảnh đầu vào.

Việc thay thế phần decoder bằng đầu ra phân loại không chỉ giúp giảm độ phức tạp của mô hình mà còn phù hợp hơn với bản chất của bài toán chẩn đoán. Điều này được minh chứng qua:

Giảm số lượng tham số và tính toán: Phần decoder trong U-Net gốc bao gồm nhiều lớp tích chập, upsampling và các kết nối bỏ qua (skip connections) phức tạp, đòi hỏi một lượng lớn phép tính và tham số. Việc loại bỏ phần này giúp giảm đáng kể chi phí tính toán và yêu cầu bộ nhớ, làm cho mô hình nhẹ hơn và nhanh hơn khi huấn luyện và dự đoán. Mặc dù không có số liệu cụ thể về số lượng tham số được trình bày trong bài báo, đây là một nguyên lý cơ bản trong thiết kế mạng nơ-ron sâu.

Tối ưu cho bài toán phân loại: Mục tiêu của phân loại ảnh là đưa ra một nhãn duy nhất cho toàn bộ ảnh. Encoder của U-Net đã rất hiệu quả trong việc trích xuất các đặc trưng cấp cao và ngữ cảnh từ ảnh y khoa. Bằng cách kết nối trực tiếp các đặc trưng này với một nhánh phân loại đơn giản, mô hình có thể tập trung vào việc học các đặc trưng phân biệt mà không cần phải tái tạo lại ảnh ở độ phân giải gốc, điều này là không cần thiết cho bài toán phân loại. Đồng thời, tận dụng kiến trúc encoder mạnh mẽ của U-Net, đặc biệt là khả năng giữ lại thông tin ngữ cảnh và chi tiết qua các skip connection (được sử dụng trong encoder để truyền thông tin qua các lớp), giúp mô hình vẫn học được những đặc trưng sâu sắc trong ảnh y khoa. Sự điều chỉnh này chứng minh rằng kiến trúc U-Net không chỉ hữu ích trong phân đoạn ảnh mà còn có thể được tùy biến linh hoạt để giải quyết hiệu quả các bài toán phân loại trong lĩnh vực học sâu ứng dụng y khoa.

Ngoài ra, việc loại bỏ phần decoder và thay thế bằng nhánh phân loại là một kỹ thuật tổng quát, có thể áp dụng cho bất kỳ kiến trúc mạng nào có phần encoder mạnh mẽ trong việc trích xuất đặc trưng hình ảnh. Do đó, việc này hoàn toàn có thể áp dụng cho các dataset ảnh y khoa khác, miễn là mục tiêu bài toán là phân loại ảnh tổng thể chứ không phải phân đoạn pixel. Hiệu quả của việc áp dụng có thể khác nhau tùy thuộc vào đặc điểm của từng dataset và cần được kiểm chứng thông qua thực nghiệm.



Hình 3. Quy trình phân loại ung thư phổi của hệ thống xây dựng dựa trên các mô hình ResNet-50, VGG-16, NasNetMobile và ViT

Hình 3 mô tả quy trình hệ thống so sánh, áp dụng cho các mô hình VGG-16, ResNet-50, NasNet Mobile và ViT. Quy trình tổng thể tương tự như Hình 2, tuy nhiên, khối “Trích xuất đặc trưng” trong Hình 2 được thay thế bằng toàn bộ kiến trúc phân loại của các mô hình nền tương ứng. Các mô hình này cũng thực hiện nhiệm vụ phân loại và được đánh giá trên cùng tập dữ liệu để đảm bảo tính công bằng khi so sánh với mô hình U-Net cải tiến. Đặc biệt, sau khi trích xuất đặc trưng bằng các tầng tích chập (encoder của mô hình nền), các đặc

trung này sẽ được đưa trực tiếp qua các lớp phân loại để thực hiện “Dự đoán”, vì vậy các thành phần thuộc phần decoder (Giảm độ phân giải, Khôi phục hình ảnh, Kết nối nhảy) đều không có trong Hình 3. Quy trình hệ thống tổng quát cho các mô hình so sánh bao gồm ResNet-50, VGG-16, NasNet Mobile, và ViT. Tương tự như Hình 2, hệ thống bắt đầu với “Dữ liệu đầu vào”, “Tiền xử lý dữ liệu” và “Phân loại dữ liệu” thành các tập huấn luyện, xác thực và kiểm chứng. Sau đó là “Khởi tạo mô hình” và “Huấn luyện mô hình”. Trong quá trình huấn luyện, các mô hình này cũng thực hiện “Trích xuất đặc trưng”. Tuy nhiên, thay vì phần decoder của U-Net, các mô hình này sẽ trực tiếp đưa các đặc trưng được trích xuất qua các lớp phân loại (ví dụ các lớp Dense) để thực hiện “Dự đoán”. “Xác thực và tính loss” được thực hiện dựa trên dự đoán này. Cuối cùng, có bước “Đánh giá và điều chỉnh” trước khi “Báo cáo kết quả”. Khác biệt cốt lõi so với U-Net cải tiến là các mô hình này không có cấu trúc encoder-decoder hình chữ U và không có các “kết nối nhảy” theo kiểu của U-Net, mà thay vào đó là các kiến trúc riêng biệt đã được tối ưu cho bài toán phân loại hoặc trích xuất đặc trưng.

Chúng tôi cũng tiến hành thực nghiệm lại mô hình VGG-16, ResNet-50, NasNet Mobile, ViT để có sự so sánh với mô hình U-Net cải tiến mà chúng tôi đề xuất.

3. Kết quả và thảo luận

3.1. Bộ dữ liệu

Được giới thiệu bởi Alyasriy và AL-Huseiny (2021), toàn bộ tiêu bản trong bộ dữ liệu đã được chú thích và xác nhận bởi các bác sĩ chuyên ngành ung thư và bác sĩ chẩn đoán hình ảnh tại hai trung tâm khác nhau. Tổng cộng, bộ dữ liệu bao gồm 1190 ảnh CT-Scan từ 110 trường hợp lâm sàng, được phân loại thành ba nhóm chính: u lành tính (Benign), u ác tính (Malignant), bình thường (Normal). Cụ thể, có 40 trường hợp ác tính, 15 trường hợp lành tính, và 55 trường hợp bình thường. Các kỹ thuật tăng cường dữ liệu cũng đã được áp dụng để giải quyết vấn đề mất cân bằng giữa các lớp phân loại. Bộ dữ liệu gốc chứa: Lành tính = 120 mẫu; Ác tính = 561 mẫu; Bình thường = 416 mẫu. Sau đó, Subhajeet Das (2024) đã áp dụng các kỹ thuật tăng cường được sử dụng bao gồm lật ngang, lật dọc, xoay, điều chỉnh độ sáng, cắt đường viền, làm mờ theo chuẩn Gauss, sắc nét, tương phản, cân bằng biểu đồ. Bộ dữ liệu tăng cường có: lành tính = 1200 mẫu, ác tính = 1201 mẫu, bình thường = 1208 mẫu.

Dữ liệu đầu vào là bộ hình ảnh mô bệnh học từ bộ dữ liệu IQ-OTH/NCCD đã qua tăng cường (3609 mẫu). Dữ liệu được chia theo phương pháp Hold-old: 80% dùng cho huấn luyện mô hình, 20% dùng cho kiểm thử.

3.2. Môi trường thực nghiệm

Bảng 1. Môi trường thực nghiệm cho mô hình U-Net cải tiến

Môi trường	Kaggle	Ngôn ngữ	Thư viện
GPU	T4 15 GB	Python 3.10.12	Tensorflow 2.15.0

Trong bài báo gốc, Ronneberger và cộng sự (2015) đã giới thiệu kiến trúc U-Net, và các phân tích chi tiết cho thấy mô hình này có khoảng 31 triệu tham số huấn luyện. Tuy

nhiên, con số này có thể thay đổi tùy thuộc vào cách triển khai cụ thể; các phiên bản hiện đại trên những nền tảng như PyTorch có thể có số lượng tham số cao hơn, dao động trong khoảng 34-39 triệu, do sự khác biệt trong cấu trúc các lớp và việc sử dụng các kỹ thuật như chuẩn hóa theo lô (Batch Normalization).

Việc thay thế phần decoder bằng đầu ra phân loại không chỉ giúp giảm độ phức tạp của mô hình mà còn phù hợp hơn với bản chất của bài toán chẩn đoán. Để chứng minh cho việc giảm độ phức tạp, chúng tôi đã tính toán số lượng tham số huấn luyện của mô hình U-Net gốc so với phiên bản cải tiến của chúng tôi. Mô hình U-Net gốc có khoảng 31 triệu tham số, trong khi mô hình U-Net cải tiến sau khi loại bỏ phần decoder chỉ còn khoảng 7.8 triệu tham số. Việc giảm hơn 75% số lượng tham số không chỉ giúp tăng tốc độ huấn luyện mà còn giảm nguy cơ overfitting, đặc biệt khi làm việc với các bộ dữ liệu y tế thường có kích thước hạn chế.

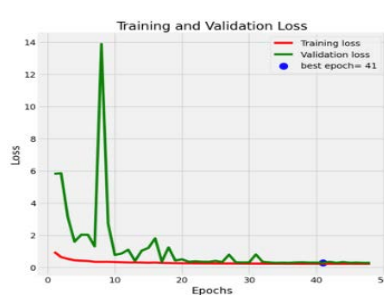
3.3. Phân tích kết quả

Đầu vào là tập dữ liệu hình ảnh mô bệnh học ung thư phổi đã được tiền xử lý để đưa vào huấn luyện và kiểm tra. Các thuật toán phân đoạn và phân loại được áp dụng để xác định bệnh. Hiệu suất của mô hình được đánh giá qua độ chính xác.

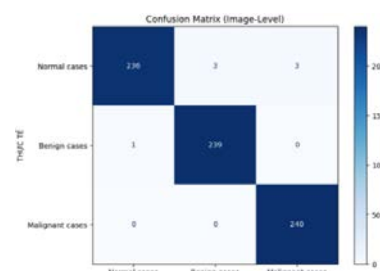
Sau khi huấn luyện mô hình U-Net cải tiến thu được biểu đồ độ chính xác và biểu đồ hàm mất mát:



Hình 4. Biểu đồ độ chính xác



Hình 5. Biểu đồ mất mát



Hình 6. Ma trận nhầm lẫn của mô hình U-Net cải tiến

Ở Hình 4, training và validation accuracy gần bằng nhau ở giai đoạn cuối, suy ra mô hình không bị overfitting. Ở Hình 5, training loss giảm đều đặn và ổn định về gần 0, suy ra mô hình học tốt trên tập huấn luyện. Validation loss giai đoạn đầu không ổn định có vài điểm cao đột ngột, sau đó giảm nhanh và tiệm cận thấp dần, đến ở epoch 41 thì hội tụ gần training loss.

Dựa vào ma trận nhầm lẫn ở Hình 6, có thể thấy mô hình đã dự đoán đúng 715/722 ảnh và chỉ có 7 trường hợp bị sai, lớp “Malignant cases” dự đoán kết quả hoàn toàn chính xác (240/240) không có mẫu nào nhầm lẫn sang lớp khác (precision, recall=100%), lớp “Benign cases” có 239/240 được dự đoán đúng và có một mẫu bị nhầm sang lớp “Normal” nên ảnh hưởng nhẹ đến recall, lớp “Normal” có 3 trường hợp nhầm là “Benign” và 3 trường hợp nhầm là “Malignant”, đây là tỉ lệ lỗi cao nhất trong 3 lớp. Nếu tiếp tục huấn luyện thì kết quả sẽ rất khả quan.

So sánh kết quả thu được với các phương pháp khác được công bố trong các nghiên cứu trước đây, cùng một bộ dữ liệu và phân loại, ta cần xem xét các tiêu chí tương tự được báo cáo bởi các nghiên cứu trước đó.

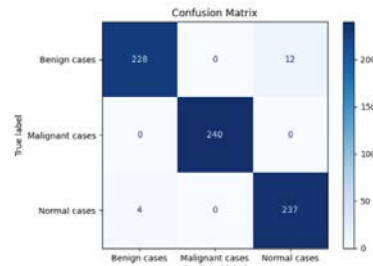
Thực nghiệm mô hình VGG-16:



Hình 7. Biểu đồ chính xác



Hình 8. Biểu đồ mất mát

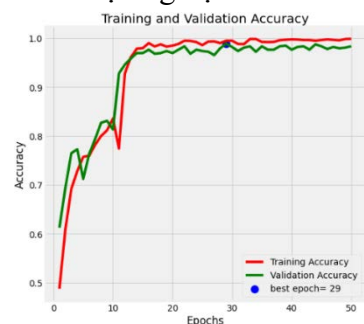


Hình 9. Ma trận nhầm lẫn

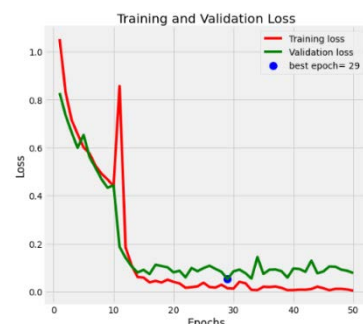
Hình 7 cho thấy độ chính xác (accuracy) trên cả tập huấn luyện và kiểm tra đều tăng dần theo thời gian, cho thấy mô hình học hiệu quả. Độ chính xác trên tập kiểm tra luôn cao hơn hoặc xấp xỉ tập huấn luyện và ổn định ở mức cao ($\sim 0.98-0.99$) sau khoảng 30 epoch, cho thấy mô hình tổng quát tốt và không có dấu hiệu overfitting rõ rệt. Hình 8 cho thấy loss trên cả tập huấn luyện và kiểm tra đều giảm ổn định theo số epoch, chứng tỏ mô hình học tốt và hội tụ. Loss kiểm tra không tăng mạnh, cho thấy không có dấu hiệu overfitting rõ rệt. Mô hình đạt hiệu suất ổn định sau khoảng 40 epoch.

Ma trận nhầm lẫn ở Hình 11 cho thấy mô hình phân loại rất tốt, hầu hết các mẫu được dự đoán đúng, với rất ít sai sót: chỉ 12 mẫu Benign bị nhầm sang Normal và 4 mẫu Normal bị nhầm sang Benign. Mô hình phân biệt chính xác hoàn toàn các trường hợp ác tính. Điều này cho thấy mô hình có độ chính xác cao và khả năng phân loại rất tốt giữa các loại tổn thương phổi.

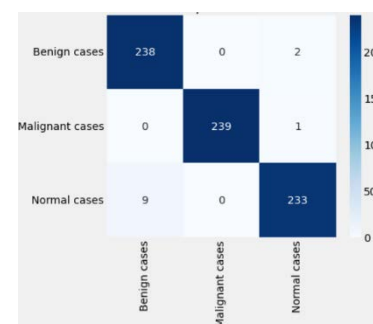
Thực nghiệm mô hình ResNet-50:



Hình 10. Biểu đồ chính xác



Hình 11. Biểu đồ mất mát

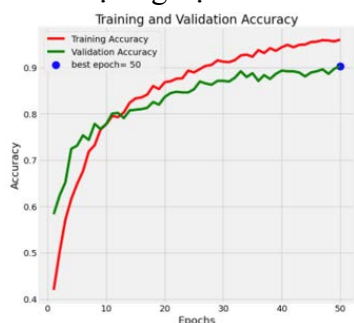


Hình 12. Ma trận nhầm lẫn

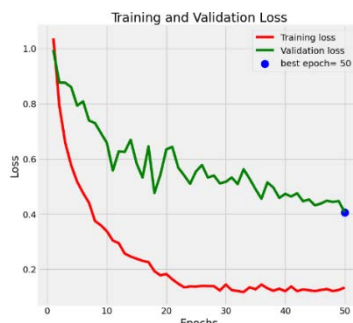
Hình 10 cho thấy độ chính xác trên cả tập huấn luyện và kiểm tra đều tăng nhanh và đạt mức rất cao (trên 0.95) sau khoảng 15 epoch. Mô hình đạt kết quả tốt nhất tại epoch 29, với validation accuracy gần đạt tối đa. Sau đó, độ chính xác ổn định và dao động nhẹ, cho thấy mô hình học hiệu quả và không có dấu hiệu overfitting. Hình 11 cho thấy loss trên cả tập huấn luyện và kiểm tra đều giảm mạnh trong các epoch đầu và dần ổn định ở mức rất thấp. Mô hình đạt hiệu suất tốt nhất tại epoch thứ 29 (được đánh dấu bằng chấm xanh), nơi validation loss thấp nhất. Sau đó, loss dao động nhẹ nhưng không tăng đáng kể, cho thấy mô hình không bị overfitting và đã hội tụ hiệu quả.

Ma trận nhầm lẫn ở Hình 12 cho thấy mô hình phân loại rất tốt ba lớp: Benign, Malignant và Normal. Hầu hết các mẫu đều được dự đoán chính xác, chỉ có một vài trường hợp nhầm lẫn nhỏ, như 2 mẫu Benign bị nhầm sang Normal và 9 mẫu Normal bị nhầm sang Benign. Điều này cho thấy mô hình có độ chính xác cao và khả năng phân biệt giữa các loại tổn thương phổi rất tốt.

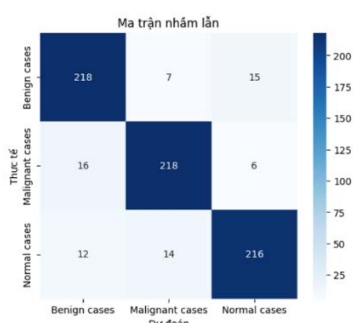
Thực nghiệm mô hình NasNet Mobile:



Hình 13. Biểu đồ chính xác



Hình 14. Biểu đồ mất mát

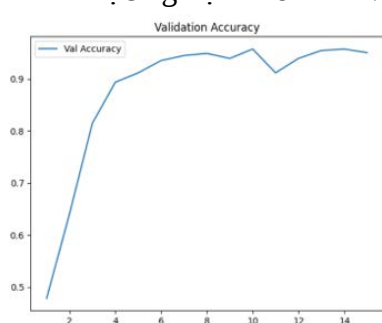


Hình 15. Ma trận nhầm lẫn

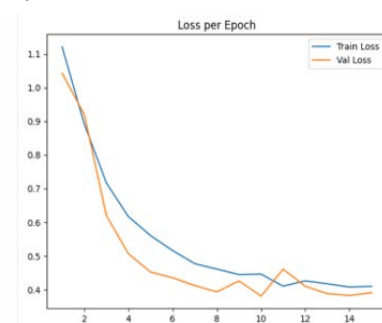
Hình 13 cho thấy độ chính xác trên cả tập huấn luyện và kiểm tra đều tăng ổn định theo số epoch. Mô hình đạt độ chính xác tốt nhất ở epoch thứ 50 với validation accuracy xấp xỉ 0.9. Tuy có khoảng cách nhỏ giữa hai đường, nhưng không đáng kể, cho thấy mô hình học hiệu quả và không bị overfitting rõ rệt. Hình 14 cho thấy mô hình học (training) rất tốt khi loss giảm nhanh chóng và ổn định theo thời gian. Tuy nhiên, validation loss có xu hướng tăng nhẹ sau khoảng 20 epochs, cho thấy có dấu hiệu overfitting. Epoch thứ 50 được đánh dấu là “best epoch”, có lẽ là điểm dừng tối ưu để cân bằng giữa việc học tốt và tránh overfitting.

Ma trận nhầm lẫn ở Hình 15 cho thấy mô hình dự đoán đúng 218 trường hợp Benign, 218 trường hợp Malignant, và 216 trường hợp Normal. Tuy nhiên, vẫn còn một số nhầm lẫn như 16 trường hợp Malignant bị dự đoán là Benign, 15 trường hợp Benign bị nhầm sang Normal, và 14 trường hợp Normal bị nhầm sang Malignant. Nhìn chung, mô hình có hiệu suất tốt nhưng cần cải thiện khả năng phân biệt giữa các lớp bệnh lý có biểu hiện gần nhau.

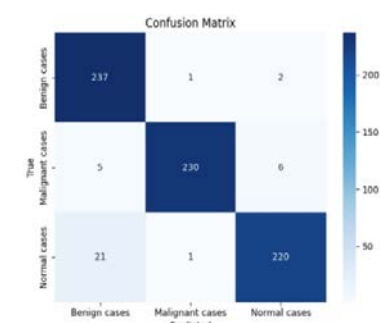
Thực nghiệm mô hình ViT:



Hình 16. Biểu đồ chính xác



Hình 17. Biểu đồ mất mát



Hình 18. Ma trận nhầm lẫn

Hình 16 cho thấy độ chính xác trên tập validation tăng nhanh trong các epoch đầu và duy trì ở mức cao (trên 90%) từ khoảng epoch thứ 6 trở đi. Điều này cho thấy mô hình học tốt, tổng quát hóa tốt và không có dấu hiệu overfitting rõ rệt. Hình 17 cho thấy cả train loss

và val loss đều giảm đều qua các epoch, cho thấy mô hình học tốt và ổn định. Không có dấu hiệu overfitting, vì val loss không tăng trở lại và duy trì ở mức thấp.

Ma trận nhầm lẫn ở hình 18 cho thấy mô hình phân loại rất tốt, với phần lớn các trường hợp được dự đoán đúng: 237 Benign, 230 Malignant và 220 Normal. Số lượng nhầm lẫn rất ít, chỉ có vài trường hợp như 5 Malignant bị nhầm là Benign và 21 Normal bị nhầm là Benign. Nhìn chung, mô hình có độ chính xác cao và khả năng phân biệt tốt giữa các lớp.

Bảng 2. Kết quả huấn luyện mô hình U-Net cải tiến

Classes	Precision (%)	Recall/Sensitivity (%)	F1-Score (%)
Normal	99.58	97.52	98.54
Benign	98.76	99.58	99.17
Malignant	98.77	100	99.38

Bảng 3. Kết quả huấn luyện của các mô hình

Mô hình	Test accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
U-Net cải tiến	99.04	99.04	99.03	99.03
VGG-16	97.78	97.82	97.78	97.78
ResNet-50	98.34	98.36	98.35	98.35
NasNet Mobile	90.31	90.31	90.31	90.30
ViT	94.99	95.25	95.03	95.03

Theo Bảng 1, mô hình U-Net cải tiến đạt hiệu suất rất cao trên cả ba lớp (Normal, Benign, Malignant), với các chỉ số precision, recall và F1-Score đều trên 97.5%, lớp Malignant có recall đạt 100%, cho thấy mô hình có khả năng phát hiện chính xác tất cả các trường hợp ác tính, đồng thời precision cũng rất cao (98.77%), chỉ ra tỉ lệ dự đoán sai thấp, lớp Benign và Normal cũng có F1-Score gần 99%, thể hiện sự cân bằng tốt giữa precision và recall. Điều này cho thấy mô hình phân loại hiệu quả trên cả ba lớp.

Theo Bảng 2, U-Net cải tiến vượt trội so với các mô hình khác (VGG-16, ResNet-50, NasNet Mobile, ViT) với Test accuracy, precision, recall và F1-Score đều đạt khoảng 99%. Điều này khẳng định ưu điểm của kiến trúc U-Net được tối ưu cho bài toán cụ thể này. ResNet-50 và VGG-16 có hiệu suất tốt nhưng thấp hơn U-Net, lần lượt đạt ~98.3% và ~97.8% F1-Score. Đây vẫn là các mô hình mạnh nhưng có thể ít phù hợp hơn với tác vụ phân loại này so với U-Net. NasNet Mobile có hiệu suất thấp nhất (F1-Score ~90.3%), có thể do kiến trúc nhẹ không đủ để nắm bắt đặc trưng phức tạp của dữ liệu. ViT (Vision Transformer) đạt hiệu suất trung bình (~95%), cho thấy mô hình dựa trên Transformer có tiềm năng nhưng cần thêm tinh chỉnh để cạnh tranh với U-Net.

Từ kết quả và so sánh trên, chúng ta có thể kết luận rằng U-Net cải tiến là mô hình tốt nhất trong số các mô hình được so sánh, đặc biệt mạnh trong phát hiện lớp Malignant (recall 100%). Sự chênh lệch hiệu suất giữa U-Net và các mô hình khác (đặc biệt là NasNet Mobile) cho thấy tầm quan trọng của việc lựa chọn kiến trúc phù hợp với bài toán. Các chỉ số cao và cân bằng ở cả hai bảng phản ánh mô hình U-Net cải tiến không chỉ chính xác mà còn ổn định trong dự đoán đa lớp.

4. Kết luận và hướng phát triển

4.1. Kết luận

Các mô hình CNN, đặc biệt là phiên bản cải tiến của U-Net, đã chứng minh hiệu quả cao trong chẩn đoán ung thư nhờ độ chính xác vượt trội và khả năng học tập tối ưu. Những tiến bộ này không chỉ mở ra cơ hội sống lớn hơn cho bệnh nhân mà còn giảm đáng kể áp lực cho đội ngũ y bác sĩ tại các bệnh viện tuyến đầu. Tuy nhiên, nghiên cứu hiện tại vẫn tồn tại một hạn chế quan trọng: dữ liệu huấn luyện và kiểm thử được thu thập trong phạm vi hẹp và đã qua chuẩn hóa kỹ lưỡng, do đó chưa phản ánh đầy đủ tính đa dạng và phức tạp của dữ liệu lâm sàng thực tế. Nói cách khác, mô hình chưa được đánh giá trên các bộ dữ liệu y khoa đa dạng từ nhiều nguồn khác nhau, điều này có thể ảnh hưởng đến khả năng tổng quát hóa trong ứng dụng thực tiễn.

4.2. Hướng phát triển

Để kiểm chứng khả năng tổng quát hóa của phương pháp, một hướng đi quan trọng là áp dụng mô hình U-Net cải tiến trên các bộ dữ liệu y tế cho các bài toán phân loại khác. Ví dụ, trong lĩnh vực chẩn đoán ung thư da, nhiều nghiên cứu đã sử dụng các mạng CNN (Huynh et al., 2022) hoặc kiến trúc U-Net cho việc khoanh vùng tổn thương (Kibriya et al., 2025). Việc thử nghiệm mô hình của chúng tôi trên các bộ dữ liệu tương tự sẽ giúp đánh giá liệu phương pháp loại bỏ decoder có thực sự hiệu quả trên các loại ảnh và bệnh lý khác nhau hay không, từ đó khẳng định tính linh hoạt và tiềm năng ứng dụng rộng rãi của nó.

Để tiếp tục tăng hiệu quả của mô hình, chúng tôi sẽ mở rộng tập dữ liệu lâm sàng thực nhờ hợp tác thu thập từ các cơ sở y khoa trong nước, xây dựng hệ thống có khả năng giải thích các yếu tố ảnh hưởng đến kết quả dự đoán, sử dụng các kiến trúc mạnh và hiện đại hơn để tối ưu hiệu suất và tốc độ triển khai. Ngoài ra, cũng cần tăng cường dữ liệu chuyên biệt cho ảnh y khoa để giảm thiểu mất đặc trưng, đồng thời cải thiện độ chính xác và tốc độ xử lý. Hơn nữa, cần có sự hợp tác đa ngành giữa các nhà nghiên cứu để phát triển thuật toán tối ưu, bác sĩ lâm sàng đóng vai trò cung cấp dữ liệu và đánh giá thực tiễn và các kỹ sư phần mềm để triển khai các hệ thống sẽ giúp phát hiện ung thư sớm, rút ngắn thời gian chẩn đoán và hỗ trợ các phác đồ điều trị cho từng bệnh nhân.

❖ **Tuyên bố về quyền lợi:** Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.

❖ **Lời cảm ơn:** Nghiên cứu này được tài trợ bởi Nguồn ngân sách khoa học và công nghệ Trường Đại học Sư phạm Thành phố Hồ Chí Minh trong đề tài mã số CS.2023.19.21.

TÀI LIỆU THAM KHẢO

- Alyasriy, H., & AL-Huseiny, M. (2021). *The IQ-OTH/NCCD lung cancer dataset* (Version 2) [Data set]. Mendeley Data. <https://doi.org/10.17632/bhmdr45bh2.2>
- Arun, M., Deenadhayalan, M., Deepak, R., Balaraman, S., & Deepak, D. (2024, July). Lung cancer detection using ResNet-50 CNN architecture. In *2024 Second International Conference on Advances in Information Technology (ICAIT)* (Vol. 1, pp. 1-6). IEEE. <https://doi.org/10.1109/ICAIT60841.2024.10540822>
- Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2024). Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2024.3435571>.
- Huynh, A. T., Hoang, V. D., Vu, S., Le, T. T., & Nguyen, H. D. (2022, July). Skin cancer classification using different backbones of convolutional neural networks. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 160-172). Cham: Springer International Publishing.
- Hosseini, S. H., Monsefi, R., & Shadroo, S. (2024). Deep learning applications for lung cancer diagnosis: A systematic review. *Multimedia Tools and Applications*, 83(5), 14305-14335. <https://doi.org/10.1007/s11042-023-17930-1>
- Howlader, N., Forjaz, G., Mooradian, M. J., Meza, R., Kong, C. Y., Cronin, K. A., Mariotto, A. B., Lowy, D. R., & Feuer, E. J. (2020). The effect of advances in lung-cancer treatment on population mortality. *New England Journal of Medicine*, 383(7), 640-649. <https://doi.org/10.1056/NEJMoa1916623>
- IARC. (2022). *Viet Nam Fact Sheet – Globocan 2022*. World Health Organization. <https://gco.iarc.who.int>
- Kibriya, H., Siddiq, A., & Khan, W. Z. (2025). Melanoma lesion localization using U-Net and explainable AI. *Neural Computing and Applications*, 37, 10175-10196
- Kumar, V., Altahan, B. R., Rasheed, T., Singh, P., Soni, D., Alsaab, H. O., & Ahmadi, F. (2023). Improved U-Net deep learning model for automatic detection of lung cancer nodules. *Computational Intelligence and Neuroscience*, 2023(1), 9739264. <https://doi.org/10.1155/2023/9739264>
- Kumar, V., Prabha, C., Sharma, P., Mittal, N., Askar, S. S., & Abouhawwash, M. (2024). Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images. *BMC medical imaging*, 24(1), 63. <https://doi.org/10.1186/s12880-024-01235-2>
- Nasra, P. (2024, August). Lung cancer classification using ResNet50. In *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)* (pp. 1201-1205). IEEE. <https://doi.org/10.1109/ICoICI61130.2024.10540823>
- Prasad, P. H. S., Daswanth, N. M. V. S., Kumar, C. V. S. P., Yeeramally, N., Mohan, V. M., & Satish, T. (2023, February). Detection of lung cancer using VGG-16. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 860-865). IEEE. <https://doi.org/10.1109/ICCMC56507.2023.10084029>

- Subhajeet Das. (2024). *IQ-OTH/NCCD Lung Cancer Dataset (Augmented)* [Dataset]. Kaggle. <https://doi.org/10.34740/kaggle/ds/6582139>
- Tran, N. N., Nguyen, H. D., Huynh, N. T., Tran, N. P., & Nguyen, L. V. (2022). Segmentation on chest CT imaging in COVID-19 based on the improvement attention U-Net model. In *New trends in intelligent software methodologies, tools and techniques* (pp. 596-606). IOS Press.
- Truong, H. M., & Huynh, H. T. (2023). A novel approach for CT-based COVID-19 classification and lesion segmentation based on deep learning. *The Computer Journal*, 66(6), 1366-1375.
- Wang, T., Wu, F., Lu, H., & Xu, S. (2023). CA-U-Net: Convolution and attention fusion for lung nodule segmentation. *International Journal of Imaging Systems and Technology*, 33(5), 1469-1479. <https://doi.org/10.1002/ima.22912>
- Xu, H., Yu, Y., Chang, J., Hu, X., Tian, Z., & Li, O. (2024). Precision lung cancer screening from CT scans using a VGG16-based convolutional neural network. *Frontiers in Oncology*, 14, 1424546. <https://doi.org/10.3389/fonc.2024.1424546>

USE OF DEEP LEARNING MODELS FOR LUNG CANCER DIAGNOSIS

Trinh Huy Hoang^{1*}, Tran Van Tuan¹,

Nguyen Le Minh Ngoc¹, Nguyen Duong Quoc Bao¹, Le Hong Thuy Vu²

¹Ho Chi Minh City University of Education, Vietnam

*²Faculty of Information Technology, Ho Chi Minh City University
of Foreign Languages and Information Technology, Vietnam*

**Corresponding author: Trinh Huy Hoang – Email: hoangth@hcmue.edu.vn*

Received: May 27, 2025; Revised: June 02, 2025; Accepted: October 22, 2025

ABSTRACT

Advances in modern medicine have facilitated the early detection of lung cancer. This study proposes a novel deep learning approach by modifying the traditional U-Net architecture, in which the decoder part is entirely removed and replaced with a specialized classification branch appended to the encoder. This change helps the model focus on extracting global image features relevant to classification, rather than reconstructing pixel-level segmentation maps. This not only significantly reduces the computational complexity but also maintains the ability to learn deep features from medical images for lung cancer detection and diagnosis. Based on the IQ-OTH/NCCD dataset, the improved U-Net model outperforms other deep learning methods, including VGG-16, ResNet-50, NasNet Mobile, and ViT.

Keywords: deep learning method; lung cancer diagnosis; U-Net model