

ĐÁNH GIÁ CHỈ SỐ TĂNG SINH KI-67 TỪ ẢNH MÔ HỌC TẾ BÀO UNG THƯ VÚ

EVALUATION OF THE KI-67 PROLIFERATION INDEX FROM HISTOLOGICAL IMAGES OF BREAST CANCER CELLS

LÊ THỊ MỸ TIÊN^{1,a}, PHAN ANH CANG¹,
TRẦN THÁI BẢO¹, TRẦN THỊ KIM NGÂN¹

¹Khoa Công nghệ Thông tin, Trường Đại học Sư phạm Kỹ thuật Vĩnh Long

^aTác giả liên hệ: tienltm@vlute.edu.vn

Nhận bài (Received): 17/9/2024; Phản biện (Reviewed): 01/10/2024; Chấp nhận (Accepted): 18/10/2024

TÓM TẮT

Chỉ số tăng sinh Ki-67 là một trong những chỉ số đánh giá quan trọng trong chẩn đoán mô bệnh học, đặc biệt là liên quan các khối u. Biểu hiện Ki-67 có liên quan chặt chẽ đến sự tăng sinh của tế bào ung thư đóng một vai trò quan trọng trong việc dự đoán và lên kế hoạch trị liệu trong quá trình theo dõi bệnh ung thư vú. Chỉ số Ki-67 là yếu tố dự đoán thường được sử dụng để đưa ra quyết định điều trị ở bệnh nhân ung thư vú. Trong bài báo này chúng tôi đề xuất phương pháp phát hiện ung thư vú dựa trên chỉ số tăng sinh Ki-67 của tế bào ung thư vú từ ảnh mô học đã được hóa mô miễn dịch protein Ki-67. Chúng tôi đề xuất thực hiện trên tập dữ liệu SHIDCB-Ki-67. Các kỹ thuật học sâu như được sử dụng bao gồm: DeepLabv3-Densenet121, FCRN, Unet bên cạnh đó kết hợp với YoLov8 nhằm cho mục đích phát hiện và phân loại protein hạt nhân tế bào bị ung thư vú trên hình ảnh mô học từ đó đánh giá chỉ số tăng sinh Ki-67. Thông qua chỉ số tăng sinh Ki-67 sẽ làm cơ sở để xác định mức độ của bệnh ung thư vú. Kết quả thực nghiệm cho thấy phương pháp đề xuất với mô hình mạng DeepLabv3-Densenet121 đạt độ chính xác cao với độ đo Accuracy đo được là 98,1%.

Từ khóa: Breast cancer, Ki-67, DeepLabv3+, YOLOv8

ABSTRACT

The Ki-67 proliferation index is one of the important indicators in histopathological diagnosis, especially related to tumors. Ki-67 expression is closely associated with the proliferation of cancer cells, playing a crucial role in predicting and planning treatment during the monitoring of breast cancer. The Ki-67 index is a commonly used prognostic factor for making treatment decisions in breast cancer patients. In this paper, we propose a method for detecting breast cancer based on the Ki-67 proliferation index of breast cancer cells from histological images that have undergone immunohistochemical staining for the Ki-67 protein. We propose implementing on the SHIDCB-Ki-67 dataset. The deep learning techniques used include DeepLabv3-Densenet121, FCRN, and Unet, in combination with YoLov8 for the purpose of detecting and classifying nuclear proteins in breast cancer cells on histological images, thereby assessing the Ki-67 proliferation index. Through the Ki-67 proliferation index, a foundation is established for determining the severity of breast cancer. Experimental results show that the proposed method using the DeepLabv3-Densenet121 network model achieves high accuracy, with a measured Accuracy score of 98.1%.

Keywords: Breast cancer, Ki-67, DeepLabv3+, YoLov8

1. GIỚI THIỆU

1.1. Giới thiệu bài toán

Bệnh ung thư vú được biết đến là một trong những nguyên nhân hàng đầu gây tử vong ở phụ nữ. Theo báo cáo của cơ quan nghiên cứu Ung thư quốc tế (IARC) năm 2022 về gánh nặng ung thư toàn cầu dựa trên cuộc khảo sát ở 115 quốc gia thì ung thư vú ở nữ giới (chiếm 11,6% tổng số ca mắc mới) là một trong những bệnh ung thư xảy ra phổ biến nhất trên toàn thế giới sau ung thư phổi (chiếm 12,4%) [13]. Khi các mô bên trong cơ thể bị phá hủy, các tế bào cơ thể phát triển bất thường sẽ dẫn đến các khối u ác tính, thường được gọi là ung thư. Các khối u ác tính có nguy cơ lan sang các bộ phận khác của cơ thể và dẫn đến những hậu quả nghiêm trọng. Phát hiện sớm giúp theo dõi và kịp thời chữa trị ung thư vú. Vì thế, các kỹ thuật để sàng lọc vú bao gồm chụp nhũ ảnh, khám lâm sàng vú, siêu âm hoặc chụp MRI là những phương pháp chẩn đoán sớm nhằm phát hiện bệnh. Tuy nhiên, trong quá trình phát triển của ung thư vú phải xem xét nhiều yếu tố: hình dạng, kích thước, tỷ lệ tế bào, khả năng di căn sang các vùng khác trên cơ thể,... khi đó sẽ sử dụng phương pháp giải phẫu mô học nhằm để xác định tình trạng và các đặc điểm của ung thư cũng như đánh giá về khả năng xâm lấn và mức độ phát triển của bệnh, theo dõi hiệu quả của các phương pháp điều trị và theo dõi các dấu hiệu tái phát. Chúng tôi nghiên cứu đề xuất phương pháp có thể phát hiện ung thư vú dựa trên sự tăng sinh của tế bào từ ảnh mô học tế bào nhằm có thể đưa ra những giải pháp cũng như đánh giá tình trạng bệnh và đưa ra những phương pháp cứu chữa kịp thời và theo dõi điều trị bệnh.

1.2. Các nghiên cứu liên quan

Những năm gần đây, các công trình

nghiên cứu sử dụng các mô hình học sâu để phát hiện các vấn đề liên quan đến bệnh ung thư vú và theo dõi sự phát triển của bệnh ung thư vú được phát triển mạnh mẽ. Trong một nghiên cứu của Monjoy Saha, Chandan Chakraborty và nhóm cộng sự [3] đã sử dụng một đề xuất học sâu nâng cao để phát hiện các điểm nhuộm màu Ki-67 để đánh giá bệnh ung thư vú trên hình ảnh tế bào học. Nghiên cứu này là phân chia nhân tế bào dấu vết vú dựa trên hình ảnh từ các slide tế bào học được chụp bằng kính hiển vi quang học. Kỹ thuật ngưỡng dựa trên biểu đồ đã được sử dụng để phân đoạn hạt nhân tế bào. Kỹ thuật được đề xuất hoạt động tốt trong phân đoạn nhân tế bào học và có khả năng phân biệt các đối tượng nhân và không phải nhân. Trong một nghiên cứu của Yubao Hou và nhóm cộng sự [6] đã trình bày một nghiên cứu về phân loại hình ảnh bệnh ung thư vú dựa trên kỹ thuật Deep learning. Nhóm nghiên cứu sử dụng các phương pháp học sâu và tăng cường dữ liệu nhằm cải thiện độ chính xác mô hình. Kết quả thử nghiệm cho thấy độ chính xác là 91% khi áp dụng mô hình học sâu cho bộ dữ liệu có sẵn công khai của BreakHis. Trong một nghiên cứu về phân tích hình ảnh mô bệnh ung thư vú để phát hiện bệnh bằng SVM của nhóm tác giả Shiksha Singh [7] đã trình bày một mô hình phân tích thử nghiệm trên bộ dữ liệu BreakHis. Với mục đích thử nghiệm, nhóm nghiên cứu đã thử nghiệm K-Nearest Neighbor (KNN), Random forest và khoảng sáu loại thuật toán phân loại Support Vector Machine (SVM). Kết quả thực nghiệm cho thấy phương pháp đề xuất nhằm phát hiện và phân loại ung thư vú đạt độ chính xác tối đa 92,3% với bộ phân loại SVM. Một đề xuất phân loại bệnh ung thư vú trên hình ảnh mô học bằng Deep Neural Network của nhóm nghiên cứu Yan R, Ren F, Wang Z và nhóm cộng sự [8]. Nhóm nghiên cứu đã đề xuất

một mạng lưới thần kinh sâu tích chập và hồi quy để phân loại hình ảnh mô bệnh học ung thư vú. Phương pháp của nhóm nghiên cứu đạt kết quả với độ chính xác trung bình đạt được là 91,3% đối với việc phân loại 4 lớp (ung thư biểu mô bình thường, ung thư lành tính, ung thư tại chỗ và ung thư xâm lấn). Các nghiên cứu tận dụng những ưu điểm của các mạng học sâu do cấu trúc linh hoạt, dễ dàng thay đổi để phù hợp với nhiều thuật toán khác nhau tuy nhiên các nghiên cứu chỉ tập trung vào việc phát hiện và phân loại bệnh ung thư vú chưa đề cập nhiều đến sự phát triển của bệnh.

2. CÁC CÔNG VIỆC LIÊN QUAN

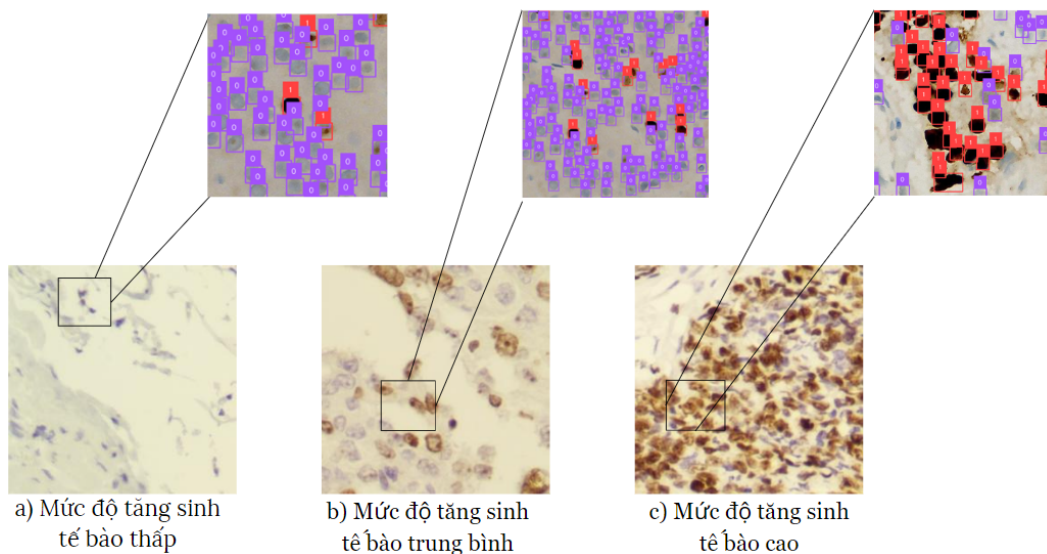
2.1. Protein Ki-67 và chỉ số đánh giá sự tăng sinh của tế bào ung thư vú

Mô bệnh học đóng một vai trò quan trọng trong sử dụng để chẩn đoán các vấn đề liên quan đến bệnh ung thư vú, dự đoán tiên lượng và theo dõi bệnh. Tỷ lệ tăng sinh khối u Ki-67 đã được chứng minh lâm sàng là có liên quan đến mức độ xâm lấn của khối u, tỷ lệ sống sót được dự đoán, tái phát và cũng có tiềm năng lớn trong việc cải thiện các quyết định về các lựa chọn

điều trị. Kháng nguyên Ki-67 [1], còn được gọi là Ki-67 hoặc Marker of Proliferation Ki-67 (MKI67) là một loại protein ở người được mã hóa bởi gen MKI67. Ki-67 vẫn hoạt động trong các giai đoạn G1, S, G2 và M của chu kỳ tế bào, khiến nó trở thành một dấu hiệu về sự tăng sinh tế bào [2] và là dấu hiệu của sự tăng sinh ung thư. Trong kỳ trung gian, kháng nguyên Ki-67 có thể được phát hiện riêng trong nhân tế bào, trong khi ở nguyên phân, hầu hết protein được di chuyển đến bề mặt của nhiễm sắc thể tế bào. Chỉ số Ki-67 được phát hiện ở tất cả các bệnh nhân ung thư vú và được đánh giá, xác định bởi các nhà giải phẫu bệnh có kinh nghiệm thông qua phương pháp hóa mô miễn dịch (IHC) trên ảnh mô học tế bào. Sự lắng đọng màu vàng trong nhân biểu hiện tế bào dương tính với Ki-67. Hoạt động tăng sinh được xác định bằng tỷ lệ tế bào dương tính với Ki-67 (tính theo phần trăm), thường được gọi là chỉ số tăng sinh Ki-67 [9] trình bày theo công thức (1):

$$PI = \frac{N_{Ki-67(+)}}{N_{Ki-67(+)} + N_{Ki-67(-)}} \times 100 \quad (1)$$

Chỉ số tăng sinh Ki-67 (PI), trong đó: $N_{Ki-67(+)}$ là số lượng tế bào dương tính với Ki-67; $N_{Ki-67(-)}$ là số lượng tế bào âm tính Ki-67.



Hình 1. Protein Ki-67 và quá trình tăng sinh trong ảnh mô học ung thư vú

Tỷ lệ biểu hiện của Ki-67 là tỷ lệ tế bào dương tính với Ki-67 (được đánh nhãn màu đỏ) trong tổng số tế bào khối u được đánh giá ở dạng phần trăm như công thức (1). Trong một nghiên cứu của Rokshana Stephny Geread đã chỉ ra rằng phạm vi được xác định với chỉ số tăng sinh ở mức thấp là $< 25\%$, trung bình $25-75\%$ và cao là $>75\%$. Hình 1.c cho thấy chỉ số tăng sinh ở mức cao, số lượng tế bào dương tính với Ki-67 nhiều hơn số lượng tế bào âm tính với Ki-67 (Trong Hình 1 được đánh nhãn màu tím) đồng nghĩa với tế bào đang tăng sinh cao dẫn đến có nguy cơ lan rộng và dẫn đến việc khó điều trị bệnh. Ở trường hợp Hình 1.a chỉ số tăng sinh thấp có nghĩa sự tăng sinh của tế bào chậm có khả năng điều trị bệnh. Hình 1.b mô tả sự tăng sinh của tế bào ở mức trung bình.

2.2. Mô hình máy học và phương pháp được dùng trong huấn luyện

Trong những năm gần đây, nhiều công trình nghiên cứu đã sử dụng các mạng nơ-ron tích chập cho nhiệm vụ phân tích hình ảnh y tế. Trong nghiên cứu này để giải quyết việc phát hiện ung thư vú dựa trên chỉ số tăng sinh Ki-67 từ hình ảnh mô học tế bào chúng tôi sử dụng một số mạng học sâu cải tiến cho nghiên cứu này như: DeepLabv3-Densenet121, U-net, FCRN bên cạnh đó kết hợp với YoLov8 để phát hiện ung thư dựa trên chỉ số tăng sinh Ki-67 được đánh giá dựa trên công thức (1).

2.2.1 Mạng DenseNet121

DenseNet [10] là kiến trúc mạng thần kinh tích chập (CNN) được giới thiệu vào năm 2016. Đây là một kiến trúc tiên tiến đã đạt được hiệu suất vượt trội trong các nhiệm vụ phân loại hình ảnh. DenseNet có

hai tính năng chính giúp nó nổi bật so với các kiến trúc CNN khác. Đầu tiên, nó có cấu trúc khối dày đặc, trong đó mỗi lớp được kết nối với mọi lớp khác theo kiểu tiếp nối. Thứ hai, nó sử dụng các lớp thắt cổ chai giúp giảm số lượng tham số mà không làm giảm số lượng tính năng mà mạng đã học được. DenseNet121 bao gồm tổng cộng 121 lớp bao gồm cả Convolutional layers, Batch normalization layers, Pooling layers và Fully connected layers.

2.2.2 DeepLabv3+

DeepLabv3+ [11] được đề xuất mở rộng DeepLabv3 bằng cách sử dụng cấu trúc bộ mã hóa. Mô-đun bộ mã hóa mã hóa thông tin theo ngữ cảnh đa quy mô bằng cách áp dụng tích chập atrous ở nhiều thang đo. Bộ mã hóa-giải mã: mạng mã hóa-giải mã đã được triển khai thành công được áp dụng cho nhiều tác vụ thị giác máy tính như ước tính tư thế của con người, phân đoạn và phát hiện các đối tượng. Mô hình bộ mã hóa-giải mã chứa mô-đun encoder giảm dần các bản đồ đặc trưng và thu được thông tin ngữ nghĩa cao hơn và mô-đun decoder phục hồi dần thông tin không gian. Dựa trên ý tưởng này, chúng tôi đề xuất sử dụng DeepLabv3+ làm mô-đun bộ mã hóa và thêm mô-đun giải mã đơn giản và hiệu quả để thu được các phân đoạn sắc nét hơn.

2.2.3 U-net

U-Net [5] là một trong những kiến trúc được sử dụng phổ biến nhất trong phân đoạn ảnh y sinh bao gồm kiến trúc hình chữ U đối xứng với hai đường dẫn có tên là bộ giải mã và bộ mã hóa. Trong mỗi lớp của bộ giải mã, một lớp lấy mẫu lên sẽ tăng kích thước bản đồ đặc trưng cho đến khi đạt đến kích thước hình ảnh đầu vào.

U-Net là một mô hình tích chập hoàn chỉnh được tạo từ 19 lớp. Điểm mới của phương pháp này là sử dụng bỏ qua kết nối giữa các lớp mã hóa và giải mã tương ứng, truyền các đặc tính có độ chi tiết cao từ các lớp mã hóa đến các lớp có cùng kích thước của bộ giải mã. Cách tiếp cận này dẫn đến việc đạt được kết quả vị trí chính xác.

2.2.4 Fully Convolutional Regression Networks

Fully Convolutional Regression Networks (FCRN) [4] để giải quyết các vấn đề hồi quy, tập trung vào việc đếm tế bào. Cách tiếp cận có thể thực hiện suy luận nhanh và đếm tế bào chính xác cho hình ảnh kính hiển vi thực. FCRN để hồi quy mật độ không gian tế bào trên hình ảnh. Điều này có thể áp dụng cho các tình huống trong đó việc phân đoạn tế bào truyền thống dựa trên các phương pháp không hoạt động tốt do các tế bào bị chồng chéo lên nhau. Cách tiếp cận này dẫn đến việc đạt được kết quả

vị trí chính xác.

2.2.5 YOLOv8

YOLOv8 [12] là phiên bản mới nhất của một dòng mô hình phát hiện đối tượng nổi tiếng có tên là YOLO (You Only Look Once). YOLO được thiết kế để thực hiện nhiệm vụ nhận diện và phân loại đối tượng trong hình ảnh một cách nhanh chóng và hiệu quả. YOLOv8 được cải thiện về độ chính xác và tốc độ so với các phiên bản trước. Nó có thể phát hiện các tế bào nhỏ trong hình ảnh y tế một cách nhanh chóng và chính xác.

2.3. Các chỉ số đánh giá mô hình

2.3.1 Độ đo Loss

Độ đo Loss là một thành phần quan trọng giúp định lượng sự khác biệt giữa kết quả đầu ra dự đoán của thuật toán máy học và giá trị mục tiêu thực tế. Công thức xác định độ đo loss (2):

$$loss = -(y \log(p) + (1 - y) \log(1 - p)) \quad (2)$$

Trong đó: y là giá trị thực tế và p là giá trị dự đoán.

2.3.2 Accuracy

Độ chính xác (Accuracy) là thước

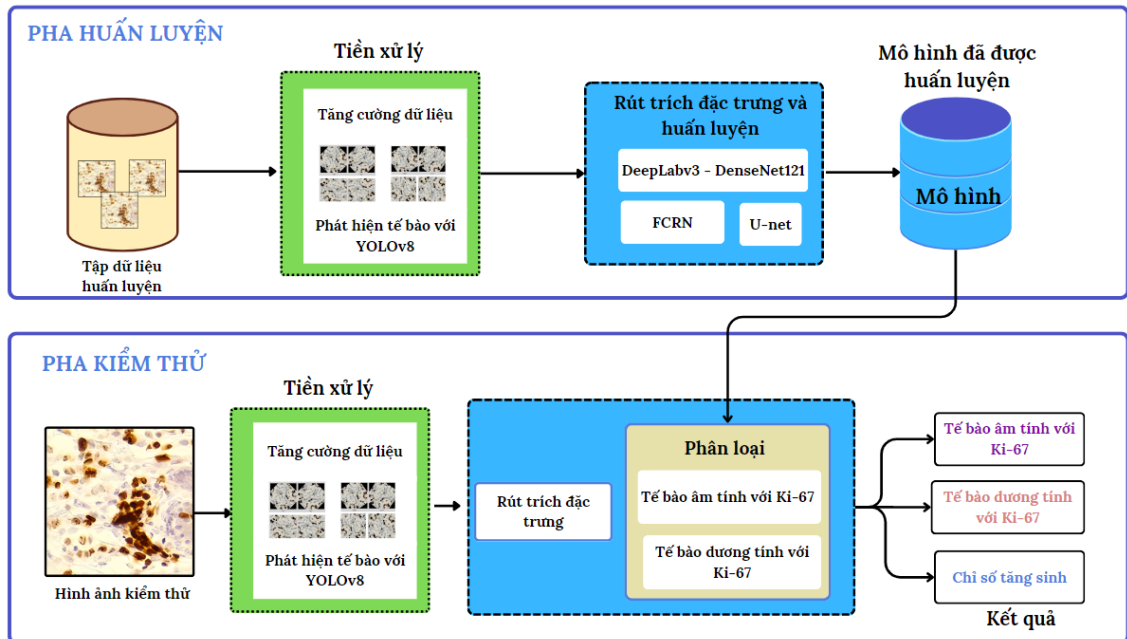
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP+TN}{\text{Số lượng mẫu}} \quad (3)$$

3. PHƯƠNG PHÁP ĐỀ XUẤT

Để giải quyết việc đánh giá chỉ số tăng sinh Ki-67 từ ảnh mô học tế bào ung thư vú trong nghiên cứu này chúng tôi áp dụng mô hình mạng học sâu và tiến hành phân đoạn hình ảnh của các tế bào khối u. Tận dụng những ưu điểm của các kiến trúc mạng học sâu: DeepLabv3-Densenet121, U-net,

đo mô tả chung cách thức hoạt động của mô hình trên tất cả các lớp. Độ chính xác (Accuracy) được tính bằng tỷ lệ giữa số lần dự đoán đúng trên tổng số lần dự đoán.

Fully Convolutional Regression Networks (FCRN) và kết hợp YoLov8 chúng tôi nghiên cứu đề xuất hệ thống nhằm phát hiện ung thư vú dựa trên chỉ số tăng sinh Ki-67 từ ảnh mô học tế bào. Quá trình thực hiện dựa trên các mô hình đề xuất bao gồm 2 pha được trình bày như ở Hình 2 gồm pha huấn luyện và pha kiểm thử.



Hình 2. Mô hình đề xuất đánh giá chỉ số tăng sinh Ki-67 từ ảnh mô học tế bào ung thư vú

3.1 Pha huấn luyện

Giai đoạn tiền xử lý dữ liệu: Hình ảnh mô bệnh học được chuẩn bị từ quá trình lấy mẫu sinh thiết cắt trụ của các khối u ở mô ung thư vú. Tiếp theo, qua quá trình hóa mô miễn dịch (IHC) các mẫu được nhuộm với protein Ki-67 và nhóm nhà nghiên cứu mô bệnh học xác định là vùng bệnh lý (pROI) để tiến hành kiểm tra mức độ bệnh. Tiếp đến quá trình chú thích được thực hiện bởi chuyên gia y tế để đảm bảo tính chính xác. Để nâng cao độ chính xác của mô hình và tăng chất lượng hình ảnh chúng tôi sử dụng một số phương pháp khử nhiễu và làm nét hình ảnh. Tuy nhiên do đặc trưng về màu sắc tế bào chúng tôi vẫn đảm bảo về độ tương phản và màu sắc để mô hình có thể nhận dạng chính xác và đảm bảo được độ chính xác khi phân loại các tế bào do sự thay đổi về màu sắc và độ tương phản sẽ dẫn đến sai các đặc trưng vốn có của hình ảnh tế bào mô học. Chúng tôi sử dụng ma trận biến đổi để tăng số lượng hình ảnh mẫu bằng cách áp dụng ma trận xoay, lật ảnh nhằm tăng sự đa dạng từ nhiều góc độ cho

bộ dữ liệu. Quá trình này được thực hiện dựa trên mô hình YoLov8 một trong những mô hình cải tiến giúp tối ưu hóa dữ liệu trước khi đưa vào mô hình để huấn luyện và phát hiện hạt nhân tế bào ung thư vú. YOLOv8 được sử dụng chủ yếu cho nhiệm vụ dò tìm nhanh các tế bào trong ảnh mô học. Đây là một mô hình mạnh mẽ trong việc xác định các tế bào.

Rút trích đặc trưng và huấn luyện:

Sau giai đoạn tiền xử lý dữ liệu, ở giai đoạn rút trích đặc trưng dựa trên những mô hình đề xuất chúng tôi tiến hành rút trích đặc trưng và huấn luyện với các mô hình mạng: FCRN, DeepLabv3-Densenet121, U-net. Dựa trên các đặc trưng cơ bản về màu sắc, kích thước, hình dạng và đặc điểm hình thái của mô bệnh học tế bào ung thư vú. Bằng cách sử dụng kết hợp các kỹ thuật rút trích đặc trưng, cùng với việc tối ưu hóa các siêu tham số qua quá trình tinh chỉnh và điều chỉnh cho phù hợp với các đặc điểm cụ thể của bộ dữ liệu. Quá trình này nhằm mục đích nâng cao hiệu suất và độ chính xác của mô hình cho các nhiệm vụ

vụ nhận dạng và phân loại hình ảnh liên quan đến hình ảnh y tế. Quá trình rút trích đặc trưng và huấn luyện dựa trên các mô hình đề xuất được tiến hành trên cùng tập dữ liệu và cùng môi trường để có thể đưa ra kết quả đánh giá so sánh một cách khách quan nhất. Giai đoạn tiếp theo chúng tôi thực hiện đánh giá chỉ số Ki-67 bằng cách so sánh tổng diện tích bề mặt mà tế bào dương tính Ki-67 chiếm giữ với tổng diện tích bề mặt của các tế bào có liên quan dựa trên công thức (1).

Bảng 1. Các kịch bản thử nghiệm dựa trên mô hình đề xuất

| Kịch bản | Mạng huấn luyện & rút trích đặc trưng | Learning Rate | Epochs | Num classes |
|------------|---------------------------------------|---------------|--------|-------------|
| Kịch bản 1 | FCRN | 1e-3 | 250 | 2 |
| Kịch bản 2 | DeepLabv3- DenseNet121 | 1e-3 | 250 | 2 |
| Kịch bản 3 | U-net | 1e-3 | 250 | 2 |

4.2 Môi trường cài đặt và tập dữ liệu thực nghiệm

4.2.1 Môi trường cài đặt

Hệ thống này được cài đặt bằng Python và hoạt động trong môi trường Google Colab Pro, với cấu hình RAM 25.4GB và GPU NVIDIA Tesla P100. Thư viện Tensorflow 2.16.1 được sử dụng để hỗ trợ quá trình huấn luyện mô hình mạng đề xuất.

4.2.2 Tập dữ liệu thực nghiệm

Bộ dữ liệu SHIDC-B-Ki-67 [9] gồm 2357 hình ảnh sinh thiết dưới kính hiển vi của loại ung thư biểu mô ống xâm lấn của ung thư vú. Các hình ảnh mô bệnh học qua quá trình nhuộm màu để làm nổi bật các tế bào hạt nhân, bộ dữ liệu gồm có tổng cộng 162.998 tế bào và trung bình mỗi hình ảnh chứa 69 tế bào với các loại tế bào gồm:

3.2 Pha kiểm thử

Ở pha kiểm thử, chúng tôi tiến hành thực nghiệm với dữ liệu đầu vào được lấy từ bộ dữ liệu thử nghiệm để phát hiện các tế bào hạt nhân dương tính với Ki-67 và các tế bào hạt nhân âm tính với Ki-67 sau đó thực hiện tính toán để đánh giá chỉ số Ki-67.

4. KẾT QUẢ THỰC NGHIỆM

4.1 Kịch bản và tham số thực nghiệm

tế bào khối u dương tính với Ki-67, tế bào khối u âm tính với Ki-67, tế bào khối u và tế bào lympho xâm nhập khối u.

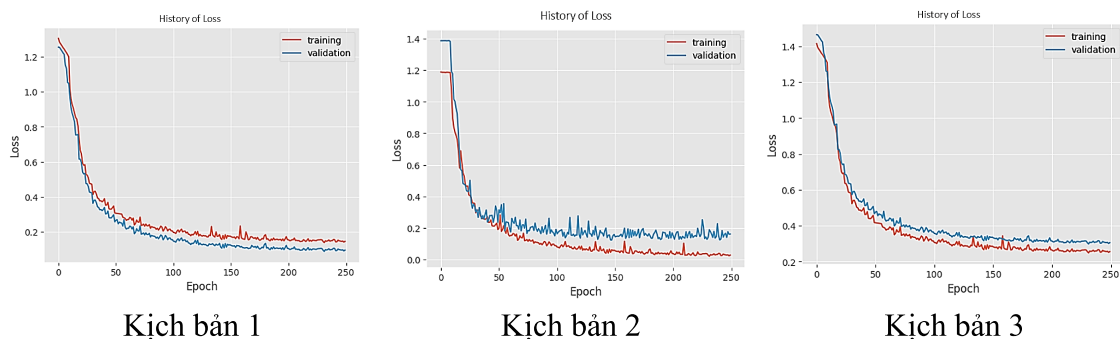
4.3. Kết quả quá trình huấn luyện

Kết quả của quá trình thực nghiệm bao gồm: Đánh giá mô hình qua các chỉ số như loss, accuracy, thời gian huấn luyện và kết quả phát hiện và phân loại âm tính hoặc dương tính trong tế bào ung thư vú dựa trên các kịch bản đã đề xuất và từ đó đánh giá chỉ số Ki-67 (PI).

4.3.1 Đánh giá mô hình

Các chỉ số Accuracy, Loss và thời gian huấn luyện được dùng để đánh giá mô hình. Kết quả thu được dựa trên 3 kịch bản với các mô hình huấn luyện được trình bày bên dưới.

4.3.1.1 Giá trị Loss

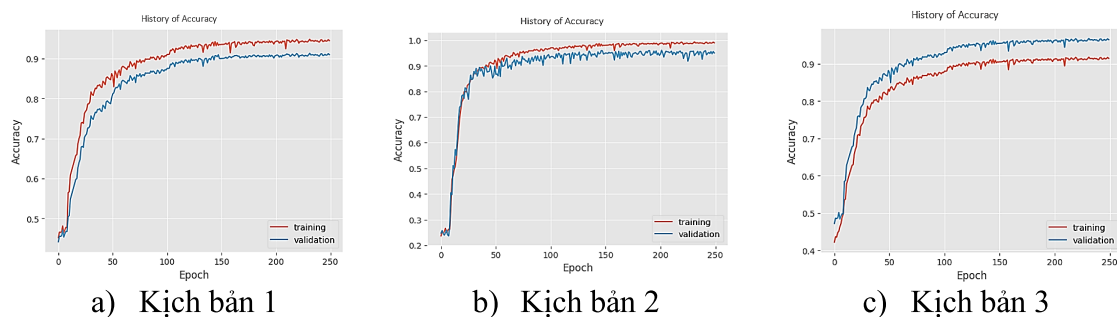


Hình 3. Giá trị Loss của 3 kịch bản

Độ đo Loss được trình bày trong hình 3.a, 3.b, 3.c dưới dạng biểu đồ. Qua quá trình huấn luyện độ đo loss trên 3 kịch bản 1, 2, và 3 tương ứng là 0.14, 0.02, và 0.25. Trong quá trình huấn luyện ở giai đoạn đầu với kịch bản 1 cho thấy độ đo Loss giảm nhanh từ bước học thứ 25 và sau đó dần ổn định ở bước học thứ 250. Mặc dù giảm nhanh, giá trị Loss vẫn khá cao, cụ thể tại bước học thứ 250 là 0.3. Với kịch bản 3, giá trị Loss có xu hướng giảm đáng kể từ bước

học thứ 50 trở đi. Đến bước học thứ 250, khi giá trị Loss không thay đổi. Đối với kịch bản 2, trong các bước học dưới 50, giá trị Loss giảm và dần ổn định từ các bước học thứ 50. Trong quá trình huấn luyện có thể thấy được mô hình đã trích xuất và học được các đặc trưng hiệu quả hơn. Khi giá trị Loss không còn giảm ở bước học thứ 250, chúng tôi quyết định dừng huấn luyện trên cả 3 kịch bản.

4.3.1.2 Giá trị Accuracy



Hình 4. Giá trị Accuracy của 3 kịch bản

Hình 4.a, 4.b, 4.c hiển thị giá trị accuracy của các kịch bản 1, 2, và 3 trên tập dữ liệu thử nghiệm dưới dạng biểu đồ. Độ đo accuracy của các kịch bản dựa trên mô hình đề xuất lần lượt là 0.945, 0.981, và 0.917. Ở kịch bản 1, từ bước học thứ 100 trở đi độ đo accuracy có thể thấy accuracy đã không tăng thêm. So với kịch bản 2, độ đo accuracy của kịch bản 1 thấp hơn 0.036. Kịch bản 2 có độ đo Accuracy cao nhất trong ba kịch bản, cho thấy khả năng trích xuất tính năng tốt hơn, với độ đo Accuracy

đạt 0.981 vào bước học thứ 250. Trong quá trình huấn luyện, ở giai đoạn đầu tại bước học dưới 25 giá trị Accuracy tăng nhanh và ổn định hơn. Với kịch bản 3, mặc dù ban đầu độ đo Accuracy khá thấp, nhưng sau quá trình huấn luyện, giá trị này đạt 0.917 vào bước học thứ 250. Chúng tôi tiến hành dừng huấn luyện trên cả ba kịch bản tại bước học thứ 250.

4.3.1.3 Thời gian huấn luyện

Tổng thời gian huấn luyện của các kịch

bản đề xuất được minh họa ở Hình 5. Cụ thể, mô hình kịch bản 1 mất tổng cộng 522 phút, kịch bản 2 là 495 phút và kịch bản 3

là 632 phút. Kết quả cho thấy kịch bản 1 có thời gian huấn luyện ngắn nhất, trong khi kịch bản 3 có thời gian huấn luyện dài nhất.



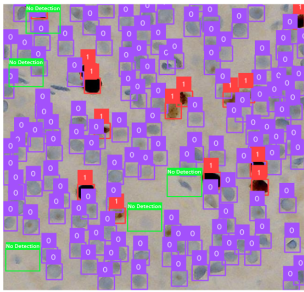
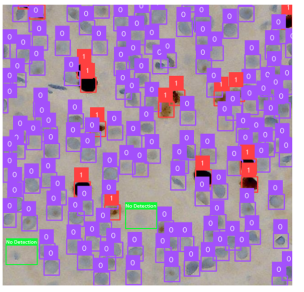
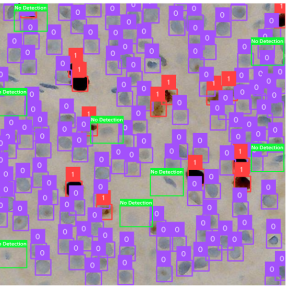
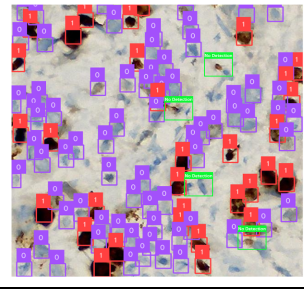
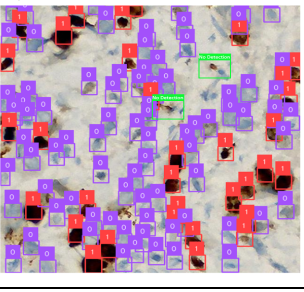
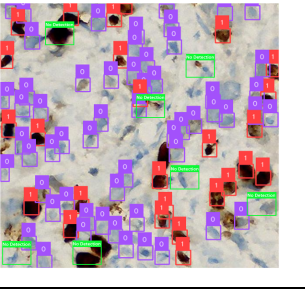
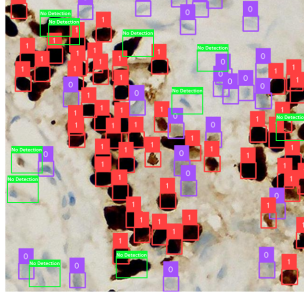
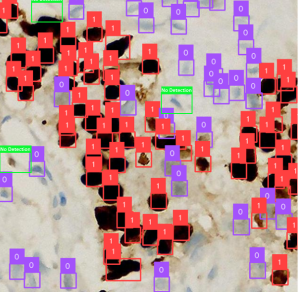
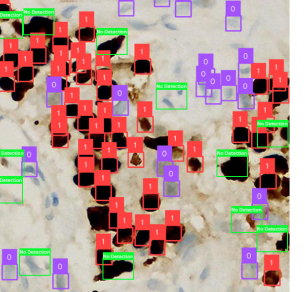
Hình 5. Thời gian huấn luyện dựa trên các kịch bản được đề xuất

4.3.1.4 Kết quả thực nghiệm đánh giá chỉ số tăng sinh Ki-67 (PI)

Kết quả thử nghiệm với hình ảnh mô ung thư vú về việc phát hiện các tế bào âm tính, dương tính với Ki-67 và đánh giá chỉ số tăng sinh Ki-67 được trình bày ở Bảng 2. Chúng tôi thực hiện đánh giá các kịch bản dựa trên các mô hình đề xuất trong 3 trường hợp chỉ số tăng sinh ở mức thấp, mức trung bình và mức cao tương ứng với trường hợp A, B và C với các loại tế bào khối u dương tính với Ki-67 (được đánh dấu nhãn màu đỏ), tế bào khối u âm tính với Ki-67 (được đánh dấu nhãn màu tím). Trong trường hợp A, ở kịch bản 1 số lượng tế bào khối u dương tính với Ki-67 được phát hiện là 14 trên tổng số 139 tế bào phát hiện được. Từ (1) kết quả chỉ số tăng sinh Ki-67 cho thấy mật độ Ki-67 đánh giá được là 10% ở mức thấp. Từ đó cho thấy được sự tăng sinh của tế bào không cao, mức độ ung thư vú nhẹ. Tương tự với các kịch bản

2 và 3. So sánh 3 kịch bản số tế bào phát hiện được ở kịch bản 2 nhiều hơn so với 2 kịch bản còn lại. Việc một số tế bào không được phát hiện (được đánh nhãn màu xanh lá) cũng cho thấy hiệu suất của mô hình đề xuất ở kịch bản 2 vượt trội hơn so với kịch bản 1 và 3. Trong trường hợp B, do tổng số tế bào phát hiện được ở kịch bản 3 ít hơn so với 2 kịch bản còn lại dẫn đến việc chênh lệch khi đánh giá chỉ số Ki-67. Trường hợp C, tương tự ở kịch bản 2 số tế bào phát hiện được cũng nhiều hơn so với kịch bản 1 và kịch bản 2. Ở trường hợp C, mật độ Ki-67 đạt mức cao do số lượng tế bào dương tính với Ki-67 nhiều hơn so với tế bào âm tính. Trường hợp này cho thấy tế bào đang tăng sinh cao ung thư có nguy cơ lan rộng và nguy hiểm dẫn đến việc rất khó điều trị. Dựa trên những kết quả đánh giá trên cho thấy ở kịch bản 2 đạt hiệu suất tốt hơn so với các kịch bản còn lại trên cả 3 trường hợp A, B, C với độ chính xác đo được là 98%.

Bảng 2. Kết quả phân loại và đánh giá chỉ số Ki-67

| | Kịch bản 1 | Kịch bản 2 | Kịch bản 3 |
|----------|--|--|---|
| A | Total cell:139 Ki-67 positive cells: 54 Ki-67 proliferation Index: 38%  | Total cell:140 Ki-67 positive cells: 55 Ki-67 proliferation Index: 39%  | Total cell:132 Ki-67 positive cells: 54 Ki-67 proliferation Index: 41%  |
| B | Total cell:104 Ki-67 positive cells: 32 Ki-67 proliferation Index: 31%  | Total cell:115 Ki-67 positive cells: 33 Ki-67 proliferation Index: 29%  | Total cell:87 Ki-67 positive cells: 30 Ki-67 proliferation Index: 35%  |
| C | Total cell:175 Ki-67 positive cells: 50 Ki-67 proliferation Index: 29%  | Total cell:190 Ki-67 positive cells: 57 Ki-67 proliferation Index: 30%  | Total cell:163 Ki-67 positive cells: 44 Ki-67 proliferation Index: 27%  |

4.3.2 So sánh đánh giá các mô hình đề xuất

Qua quá trình thực nghiệm, dựa trên

các kết quả đánh giá trong quá trình huấn luyện và kiểm thử chúng tôi tiến hành so sánh các kết quả như Bảng 3.

Bảng 3. Bảng so sánh đánh giá các mô hình đề xuất

| Kịch bản | Huấn luyện | | | Kiểm thử | | |
|------------|--------------|------------|----------------------|--------------------|--------------|--------------------|
| | Độ chính xác | Độ đo Loss | Thời gian huấn luyện | Thời gian kiểm thử | Độ chính xác | Thời gian kiểm thử |
| Kịch bản 1 | 0.945 | 0.14 | 522.5 phút | 15 giây | 0.938 | 15 giây |
| Kịch bản 2 | 0.981 | 0.02 | 495 phút | 8 giây | 0.978 | 8 giây |
| Kịch bản 3 | 0.917 | 0.25 | 632.5 phút | 18 giây | 0.901 | 18 giây |

Kết quả so sánh đánh giá có thể thấy ở kịch bản 2 về độ đo Loss đạt được là 0.02 thấp nhất trong 3 kịch bản. Với độ đo Accuracy độ đo Accuracy đo được là 0.981 cao nhất trong 3 kịch bản. Ở kịch bản 2 có thể thấy được khả năng rút trích các đặc trưng hiệu quả hơn nên độ chính xác cao hơn so với các kịch bản còn lại. Ngoài ra về thời gian kiểm thử và thời gian huấn luyện qua so sánh đánh giá có thể thấy ở kịch bản 1 và kịch bản 3 chưa tốt so với kịch bản 2.

6. KẾT LUẬN

Nghiên cứu đề xuất để đánh giá chỉ số tăng sinh Ki-67 từ ảnh mô học tế bào ung thư vú được sử dụng để phát hiện và đánh giá các loại tế bào ung thư vú từ đó làm cơ sở để đánh giá mức độ bệnh ung thư vú. Nghiên cứu sử dụng các mô hình học sâu bên cạnh đó kết hợp với YoLov8 một mô hình tiên tiến để phát hiện các tế bào âm tính hoặc dương tính với Ki-67 từ đó dựa trên các tế bào để đánh giá mật độ Ki-67 xuất hiện trong mô ung thư vú. Kết quả thực nghiệm cho thấy mô hình đề xuất ở kịch bản 2 đạt kết quả tốt nhất với độ chính

xác là 0.981 trong việc phát hiện các tế bào và đánh giá mật độ tế bào chính xác hơn. Trong nghiên cứu này nhóm nghiên cứu sử dụng tập dữ liệu về ảnh mô bệnh học ung thư vú được công khai sau đó tiến hành tăng cường dữ liệu và cải thiện chất lượng hình ảnh để tập dữ liệu được tốt hơn và tăng sự đa dạng về tế bào mô bệnh học. Đề xuất các mô hình như FCRN, DeepLabv3-Densenet121, và U-net, bên cạnh đó trong quá trình huấn luyện thực hiện kết hợp với mô hình YoLov8 và tiến hành cải tiến tinh chỉnh các tham số cũng như thay đổi các lớp mạng để hiệu suất mô hình được hiệu quả một cách tốt nhất. Đưa ra được các kết quả so sánh đánh giá để giúp việc phát hiện, theo dõi sự hình thành bệnh ung thư vú và làm cơ sở cho việc phát triển các mô hình đề xuất tiếp theo. Trong tương lai, chúng tôi sẽ tiếp tục nghiên cứu và phát triển các phương pháp cải tiến để rút ngắn thời gian huấn luyện đồng thời tập trung vào việc thu thập và xử lý thêm dữ liệu hình ảnh mô học từ nhiều nguồn khác nhau để nâng cao tính đa dạng và đồng nhất của dữ liệu.

TÀI LIỆU THAM KHẢO

- [1]. D. M. Schonk et al., (1989), "Assignment of the gene(s) involved in the expression of the proliferation-related Ki-67 antigen to human chromosome 10," *Hum Genet*, vol. 83, no. 3.
- [2]. R. Rahmanzadeh, G. Hüttmann, J. Gerdes, and T. Scholzen, (2007), "Chromosome-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis," *Cell Prolif*, vol. 40, no. 3.
- [3]. M. Saha et al., (2015), "Histogram Based Thresholding for Automated Nucleus Segmentation Using Breast Imprint Cytology".
- [4]. W. Xie, J. A. Noble, and A. Zisserman, (2018) "Microscopy cell counting and detection with fully convolutional regression networks," *Comput Methods Biomech Biomed Eng Imaging Vis*, vol. 6, no. 3.
- [5]. J. Dolz, C. Desrosiers, and I. Ben Ayed, (2019) "IVD-net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet," in *Lecture Notes in Computer Science*.

- [6]. Y. Hou et al., (2020), “Breast cancer pathological image classification based on deep learning,” *J Xray Sci Technol*, vol. 28, no. 4.
- [7]. S. Singh and R. Kumar, (2020), “Histopathological image analysis for breast cancer detection using cubic SVM,” in *International Conference on Signal Processing and Integrated Networks*.
- [8]. R. Yan et al., (2020), “Breast cancer histopathological image classification using a hybrid deep neural network,” *Methods*, vol. 173.
- [9]. F. Negahbani et al., (2021), “PathoNet introduced as a deep neural network backend for evaluation of Ki-67 and tumor-infiltrating lymphocytes in breast cancer,” *Sci Rep*, vol. 11.
- [10]. S. B. Thunuguntla, S. Murugaanandam, and R. Pitchai, (2023), “Densenet121-DNN-Based Hybrid Approach for Advertisement Classification and User Identification,” *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 3.
- [11]. H. Zeng, S. Peng, and D. Li, (2020), “Deeplabv3+ semantic segmentation model based on feature cross attention mechanism,” in *Journal of Physics: Conference Series*.
- [12]. A. Khatua, A. Bhattacharya, A. K. Goswami, and B. H. Aithal, (2024). “Developing approaches in building classification and extraction with synergy of YOLOV8 and SAM models,” *Spatial Information Research*.
- [13]. N. Jokhadze, A. Das, and D. S. Dizon, (2024), “Global cancer statistics: A healthy population relies on population health,” *CA Cancer J Clin*.