

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP.HCM
KHOA CÔNG NGHỆ THÔNG TIN



NGUYỄN THỊ THANH HÒA
TRƯỜNG THỊ CẨM LY

KHÓA LUẬN TỐT NGHIỆP

**TIẾP CẬN HỖ TRỢ CHẨN ĐOÁN UNG THƯ TỪ
NHỮNG ẢNH SỬ DỤNG PHƯƠNG PHÁP HỌC SÂU**

Chuyên ngành: Khoa học dữ liệu

Giảng viên hướng dẫn: PGS.TS Huỳnh Trung Hiếu

THÀNH PHỐ HỒ CHÍ MINH, THÁNG 5, NĂM 2023

**BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP.HCM
KHOA CÔNG NGHỆ THÔNG TIN**



**NGUYỄN THỊ THANH HÒA
TRƯỜNG THỊ CẨM LY**

KHÓA LUẬN TỐT NGHIỆP
**TIẾP CẬN HỖ TRỢ CHẨN ĐOÁN UNG THƯ TỪ
NHỮNG ẢNH SỬ DỤNG PHƯƠNG PHÁP HỌC SÂU**

Chuyên ngành: Khoa học dữ liệu

Giảng viên hướng dẫn: PGS.TS Huỳnh Trung Hiếu

THÀNH PHỐ HỒ CHÍ MINH, THÁNG 5, NĂM 2023

MINISTRY OF INDUSTRY AND TRADE
INDUSTRIAL UNIVERSITY OF HO CHI MINH CITY
FACULTY OF INFORMATION TECHNOLOGY



NGUYEN THI THANH HOA

TRUONG THI CAM LY

GRADUATION THESIS
DIAGNOSIS OF BREAST CANCER
USING DEEP LEARNING

Major: Data Science

Instructor: Assoc.Prof. Huynh Trung Hieu

HO CHI MINH CITY, MAY, 2023

TÓM TẮT

Ung thư vú đứng đầu trong các bệnh ung thư thường gặp ở nữ giới và là nguyên nhân gây tử vong hàng đầu do ung thư trên toàn thế giới. Việc xác định và phân loại chính xác khối u là rất quan trọng trong việc phát hiện ung thư vú ở giai đoạn đầu nhằm kịp thời chữa trị cho bệnh nhân. Trong các phương pháp sàng lọc phát hiện sớm tế bào ung thư vú, từ lâu người ta đã công nhận rằng chụp nhũ ảnh là một kỹ thuật được dùng để sàng lọc và phát hiện các bệnh lý về tuyến vú và ung thư vú để phát hiện những thay đổi ở vú đối với những phụ nữ không có dấu hiệu hoặc triệu chứng của ung thư vú. Do đó, kỹ thuật phân đoạn hình ảnh và trí tuệ nhân tạo có thể được áp dụng để chẩn đoán ung thư vú giai đoạn đầu. Hiện nay, để rút ngắn lại thời gian chẩn đoán bệnh cho đội ngũ y tế, người ta đã áp dụng các mô hình học sâu trong việc phát hiện ung thư dựa trên nhũ ảnh. Trong khóa luận này, chúng tôi sử dụng mô hình được xây dựng dựa trên mạng nơ ron tích chập, mô hình ResNet50, mô hình ConvNeXt và mô hình HRNet để phân loại ung thư. Các mô hình được huấn luyện trên các bộ dữ liệu nhũ ảnh như: RSNA (54706 ảnh), DDSM (7808 ảnh), CMMD (5202 ảnh), INbreast (410 ảnh), MIAS (322 ảnh). Trước khi đào tạo mô hình, chúng tôi đã áp dụng một số thuật toán tiền xử lý như chuẩn hóa dữ liệu, tăng cường với CLAHE, xoay và lật ảnh,...sau đó áp dụng mô hình và đạt được độ chính xác đáng kể. Bằng các phương pháp được đề xuất như trên, khóa luận của chúng tôi đạt được kết quả phân loại với độ chính xác khá tốt trên từng tập dữ liệu: Mias (97.52%), INbreast (96.58%), DDSM (96.8%), CMMD (96%).

Từ khóa: Ung thư vú; Chụp quang tuyến vú; Mạng thần kinh tích chập; HRNet; ConvNeXt.

ABSTRACT

Breast cancer ranks first among the most common cancers in women and is the leading cause of cancer mortality in the world. Accurate identification and classification of tumors are very important to detect early-stage breast cancer to treat patients quickly. Among the screening methods for early detection of breast cancer cells, it has long been recognized that mammography is a technique used to screen and detect diseases of the mammary gland and breast cancer. Breast changes in women who do not show signs and symptoms of breast cancer. As a result, image segmentation techniques and artificial intelligence can be applied to the diagnosis of early breast cancer. Currently, to shorten the diagnosis time for the medical team, people have applied deep learning models in cancer detection based on mammography. In this thesis, we use the models built on the convolutional neural network, the ResNet50 model, the ConvNeXt model, and the HRNet model to classify cancer. The models were trained on mammography datasets such as RSNA (54706 images), DDSM (7808 images), CMMD (5202 images), INbreast (410 images), and MIAS (322 images). Prior to forming the model, we applied some pre-processing algorithms such as data normalization, CLAHE enhancement, image rotation and flip, etc., then applied the model and achieved remarkable accuracy. By using the proposed methods as above, our thesis achieves classification results with good accuracy on each dataset: Mias (97.52%), INbreast (96.58%), DDSM (96.8%), and CMMD (96%).

Keywords: Breast cancer; Mammogram; Convolutional neural network; HRNet; ConvNeXt

LỜI CẢM ƠN

Đề khóa luận tốt nghiệp này đạt được kết quả tốt nhất, chúng tôi đã nhận được sự hỗ trợ, giúp đỡ của quý thầy cô Khoa Công Nghệ Thông Tin và nhà trường. Với tình cảm sâu sắc, chân thành, chúng tôi xin được bày tỏ tấm lòng biết ơn đến quý thầy cô, khoa Công Nghệ Thông Tin và nhà trường đã tạo điều kiện, giúp đỡ chúng tôi trong quá trình học tập, nghiên cứu cũng như các hoạt động học thuật.

Lời đầu tiên chúng tôi xin phép gửi lời cảm ơn chân thành đến PGS. TS Huỳnh Trung Hiếu. Thầy là người đã trực tiếp giảng dạy, chỉ bảo, dẫn dắt, góp ý cho chúng tôi trong việc học vấn lẫn kinh nghiệm làm việc, nhờ thầy mà chúng tôi có thể có cơ hội thử sức với đề tài này. Và cũng nhờ thầy mà chúng tôi có thể có cơ hội hoàn thành tốt hơn bài báo cáo này. Chúng tôi cũng xin gửi lời cảm ơn chân thành đến thầy Nguyễn Hữu Tình, giáo viên chủ nhiệm lớp DHKHD15A, cũng là người đã giúp đỡ chúng tôi rất nhiều trong quá trình định hình bản thân, thầy đã truyền lửa cho chúng tôi để chúng tôi biết được rằng, chỉ cần cố gắng thì bất kỳ điều gì mình cũng có thể làm được, mặc kệ xuất phát điểm của bản thân ở đâu.

Chúng tôi cũng xin gửi lời cảm ơn chân thành đến quý thầy, cô ở Khoa Công Nghệ Thông Tin – Trường Đại học Công nghiệp Thành phố Hồ Chí Minh đã giảng dạy chỉ bảo chúng tôi trong suốt quãng thời gian chúng tôi học tập tại trường. Chúng tôi cũng xin bày tỏ lòng biết ơn đến ban lãnh đạo của Trường Đại học Công nghiệp Thành phố Hồ Chí Minh và các khoa, phòng ban chức năng đã trực tiếp hoặc gián tiếp giúp đỡ chúng tôi trong suốt quá trình học tập và thực hiện báo cáo này.

Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế, khóa luận này không thể tránh được những thiếu sót. Chúng tôi rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các thầy cô để chúng tôi có điều kiện bổ sung, nâng cao kiến thức của mình, phục vụ tốt hơn công việc thực tế sau này.

NHẬN XÉT VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Phân đánh giá: (thang điểm 10)

Thái độ thực hiện:

Nội dung thực hiện:.....

Kỹ năng trình bày:.....

Tổng hợp kết quả:.....

Điểm bằng số:.....Điểm bằng chữ:.....

Tp. Hồ Chí Minh, ngày....tháng năm 2023

Trưởng bộ môn chuyên ngành
(Ký ghi rõ họ tên)

Giảng viên hướng dẫn
(Ký ghi rõ họ tên)

NHẬN XÉT VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN PHẢN BIỆN 1

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Tp. Hồ Chí Minh, ngày tháng năm 2023

CHỮ KÝ CỦA GIẢNG VIÊN

(Ký ghi rõ họ tên)

NHẬN XÉT VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN PHẢN BIỆN 2

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Tp. Hồ Chí Minh, ngày..... tháng năm 2023

CHỮ KÝ CỦA GIẢNG VIÊN

(Ký ghi rõ họ tên)

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU	1
1.1 Tổng quan	1
1.2 Các kĩ thuật chụp nhũ ảnh	5
1.3 Mục tiêu nghiên cứu	8
1.4 Lí do chọn đề tài	9
1.5 Phạm vi nghiên cứu	9
1.6 Ý nghĩa khoa học và thực tiễn	10
1.6.1 Ý nghĩa khoa học	10
1.6.2 Ý nghĩa thực tiễn.....	10
1.7 Bố cục luận văn	11
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN	12
2.1 Học sâu	12
2.2 Các nghiên cứu đã có.....	13
2.3 Mô hình Convolution Neural Networks (CNNs)	15
2.4 Mô hình Residual Network (ResNet)	18
2.5 Mô hình ConvNeXt	20
2.6 Mô hình High - Resolution Networks (HRNets).....	26
2.6.1 Giới thiệu	26

2.6.2	Các tích chập song song đa độ phân giải.....	27
2.6.3	Các phép kết hợp đặc trưng	28
2.6.4	Các dạng biểu diễn.....	28
2.7	Phương pháp đánh giá kết quả.....	29
2.7.1	Confusion Matrix	29
2.7.2	Độ chính xác (Accuracy)	30
2.7.3	Recall	30
2.7.4	Precision.....	30
2.7.5	F1 score.....	31
2.8	Các phương pháp fine-tuning model	31
CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT.....		32
3.1	Tổng quan hướng tiếp cận	32
3.2	Đặc trưng mô hình đề xuất	32
3.2.1	Mô hình CNNs.....	32
3.2.2	Mô hình ResNet50	33
3.2.3	Mô hình ConvNeXt	33
3.2.4	Mô hình HRNet	34
3.3	Hàm mất mát Cross Entropy	34
3.4	Hàm kích hoạt.....	35

3.4.1	ReLU.....	35
3.4.2	Softmax.....	36
3.5	Các thuật toán tối ưu.....	37
3.5.1	Stochastic Gradient Descent (SGD)	37
3.5.2	Adam.....	38
3.5.3	AdamW	38
3.6	Các phương pháp tiền xử lí dữ liệu	39
CHƯƠNG 4. THỰC NGHIỆM VÀ KẾT QUẢ		45
4.1	Dữ liệu	45
4.1.1	Tập dữ liệu INbreast	45
4.1.2	Tập dữ liệu Mammographic Image Analysis Society (MIAS).....	46
4.1.3	Tập dữ liệu RSNA Screening Mammography Breast Cancer Detection...	46
4.1.4	Tập dữ liệu Digital Database for Screening Mammography (DDSM)	47
4.1.5	Tập dữ liệu The Chinese Mammography Database (CMMD)	48
4.1.6	Các kĩ thuật xử lí ảnh.....	49
4.1.7	Các phương pháp xử lí dữ liệu mất cân bằng	55
4.1.8	Resampling	55
4.1.9	Trọng số phân lớp (Class weighting).....	56
4.1.10	Làm mịn nhãn (Labels smoothing)	57

4.2	Dữ liệu thực nghiệm với mô hình ConvNeXt và CNNs:	57
4.3	Dữ liệu thực nghiệm với mô hình HRNet	59
4.4	Dữ liệu thực nghiệm với mô hình ResNet50.....	59
4.5	Công nghệ sử dụng	60
4.6	Kết quả.....	60
4.6.1	Kết quả sau khi thực nghiệm	60
4.6.2	Nhận xét về kết quả đạt được	72
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN		76
5.1	Kết luận.....	76
5.2	Kiến thức đạt được	76
5.3	Kĩ năng đạt được	76
5.4	Hướng phát triển trong tương lai.....	76
TÀI LIỆU THAM KHẢO.....		78
NHẬT KÍ LÀM VIỆC		86

MỤC LỤC HÌNH ẢNH

Hình 1.1. Quá trình sinh thiết.....	2
Hình 1.2. Minh họa về tổn thương lành tính (trái) và tổn thương ác tính (phải) [4]	3
Hình 1.3. Minh họa về vi vôi hóa ở nhũ ảnh.....	3
Hình 1.4 Mô phỏng quá trình chụp nhũ ảnh	5
Hình 1.5. Hình ảnh chụp cộng hưởng từ (Magnetic Resonance Imaging (MRI))	5
Hình 1.6. Hình ảnh mô phỏng 2 góc chụp CC và MLO	6
Hình 1.7. Hình mô tuyến vú được chụp từ nhũ ảnh (Loại 1: Hình A; Loại 2: Hình B; Loại 3: Hình C; Loại 4: Hình D)	7
Hình 2.1. Kiến trúc học sâu.....	12
Hình 2.2. Minh họa kiến trúc CNNs giải quyết bài toán phân loại chữ viết tay [18] ...	15
Hình 2.3. Minh họa về stride $S=2$ [18]	16
Hình 2.4. Minh họa về tầng kết nối đầy đủ.....	18
Hình 2.5. Kiến trúc mạng ResNet-50 gồm 2 khối đặc trưng là khối tích chập (Conv Block) và khối Identity [20].....	19
Hình 2.6 Khối Residual [20]	20
Hình 2.7. Kiến trúc ConvNeXt (trong đó bỏ qua không miêu tả các skip connection) ..	21
Hình 2.8. Kiến trúc Swin-Transformer tiêu chuẩn.....	22
Hình 2.9. Minh họa về phương pháp 'stem' (Resnet Stem) và patchify (ConvNeXt stem)	23

Hình 2.10. Minh họa về group convolution	24
Hình 2.11. Minh họa một khối inverted bottleneck.	24
Hình 2.12. Minh họa hàm ReLU và GELU	25
Hình 2.13. Cấu trúc phục hồi độ phân giải cao từ độ phân giải thấp.....	27
Hình 2.14. Minh họa mạng HRNet	27
Hình 2.15. Minh họa cách kết hợp thông tin ngữ nghĩa các đặc trưng độ phân giải cao, trung bình và thấp tương ứng từ trái sang phải.	28
Hình 2.16. Confusion Matrix	29
Hình 3.1. Sơ đồ quá trình giải quyết bài toán	32
Hình 3.2. Kiến trúc CNNs sử dụng cho bài toán chẩn đoán nhũ ảnh.	33
Hình 3.3. Hàm kích hoạt ReLU	35
Hình 3.4. Nhũ ảnh trước và sau khi cắt.....	40
Hình 3.5. Sơ đồ cắt ảnh sử dụng phương pháp tìm liên thông lớn nhất	41
Hình 3.6. Sơ đồ quá trình cắt ảnh sử dụng YOLOv8.....	42
Hình 3.7.Minh họa về tính chính xác của dự đoán với sự trợ giúp của ngưỡng đánh giá IoU.....	43
Hình 3.8. Precision Recall Curve	43
Hình 4.1. Nhũ ảnh trong bộ dữ liệu INbreast: a) Góc chụp CC vú phải, b) Góc chụp CC vú trái, c) Góc chụp MLO vú phải, d) Góc chụp MLO vú trái	46
Hình 4.2. Hình ảnh minh họa dữ liệu trong tập RSNA.....	47

Hình 4.3. Hình ảnh mô phỏng của tập dữ liệu DDSM.....	48
Hình 4.4. Hình ảnh minh họa sơ đồ quá trình CLAHE [42].....	49
Hình 4.5. Ảnh minh họa ảnh nguyên bản (trước) và sau khi CLAHE (phải).....	50
Hình 4.6. Hình ảnh minh họa sử dụng rotate với góc xoay 45 độ	52
Hình 4.7. Hình ảnh minh họa sử dụng kỹ thuật HorizontalFlip.....	53
Hình 4.8. Hình ảnh mô tả sử dụng kỹ thuật VerticalFlip	53
Hình 4.9. Hình ảnh mô tả sử dụng kỹ thuật GaussianBlur	54
Hình 4.10. Minh họa kỹ thuật undersampling	56
Hình 4.11. Minh họa kỹ thuật oversampling ¹¹	56
Hình 4.12. Loss và accuracy của mô hình CNNs train ở giao đoạn đầu	73

DANH MỤC BẢNG BIỂU

Bảng 3.1. Thông số mô hình convnext_small.fb_in22k_ft_in1k_384.....	33
Bảng 3.2. Bảng tổng quan số lượng mẫu ảnh đã sử dụng trong bài toán này.....	39
Bảng 3.3. Dữ liệu huấn luyện mô hình YOLOv8	42
Bảng 4.1. Chia dữ liệu TrainSet1	58
Bảng 4.2. Chia dữ liệu TrainSet2	58
Bảng 4.3. Dữ liệu huấn luyện mô hình HRNet.....	59
Bảng 4.4. Dữ liệu đánh giá mô hình HRNet.....	59
Bảng 4.5. Dữ liệu huấn luyện mô hình ResNet50	59
Bảng 4.6. Dữ liệu đánh giá mô hình ResNet50	59
Bảng 4.7. Mô tả phần cứng được sử dụng trong quá trình thực nghiệm	60

DANH MỤC CÁC THUẬT NGỮ VIẾT TẮT

TỪ NGỮ	Ý NGHĨA
Batch	Một lô các mẫu dữ liệu
Batch size	Kích thước một lô mẫu dữ liệu
Confusion matrix	Ma trận nhầm lẫn
ReLU	Hàm kích hoạt ReLU
MRI	Magnetic Resonance Imaging – chụp cộng hưởng từ
MCs	Micro-Calcifications – vi vôi hóa
AI	Artificial intelligence – trí tuệ nhân tạo
CC	Cranio-Caudal – hình ảnh được chụp từ phía trên
MLO	Mediolateral Oblique - hình ảnh được chụp từ một bên theo một góc sao cho nó có thể nhìn thấy cơ ngực
Logits	Giá trị xác suất lớp cuối cùng của mạng nơ-ron
Softmax	Hàm kích hoạt Softmax
BatchNorm	Batch Normalization
Conv	Lớp tích chập (Convolutional)
CT	Chụp cắt lớp vi tính
Malignant	Các tổn thương ác tính
CNN	Convolution Neural Network

HRNet	Hight Residual Network
ResNet	Residual Network
DDSM	Digital Database for Screening Mamography
Mias	Mamographic Image Analysis Society
RSNA	Radiological Society of North America
CMMD	The Chinese Mammography Database
BI-RADS	Breast Imaging Reporting and Data System – thang đo thể hiện mức độ dày đặc của tuyến vú
CAD	Computer – Aided Detection – Phát hiện hỗ trợ bởi máy tính
ViTs	Vision Transformer
Acc	accuracy

CHƯƠNG 1. GIỚI THIỆU

1.1 Tổng quan

Ung thư là căn bệnh vô cùng nguy hiểm và đáng sợ, đặc trưng bởi khả năng xâm lấn, lây lan đến các bộ phận của cơ thể. Nếu không phát hiện và điều trị kịp thời, người bệnh sẽ tử vong nhanh chóng. Tế bào ung thư phát triển và phân chia rất nhanh trong cơ thể bệnh nhân. Hiện nay, tỷ lệ mắc ung thư và tử vong do ung thư tiếp tục tăng trên thế giới. Số liệu được Globocan (dự án của Cơ quan nghiên cứu ung thư quốc tế) công bố năm 2020, trên thế giới ước tính có khoảng 19,3 triệu ca mới và 9,9 triệu ca tử vong, cao hơn nhiều so với khoảng 14 triệu ca mới và 8,2 triệu ca tử vong vào năm 2012 [1].

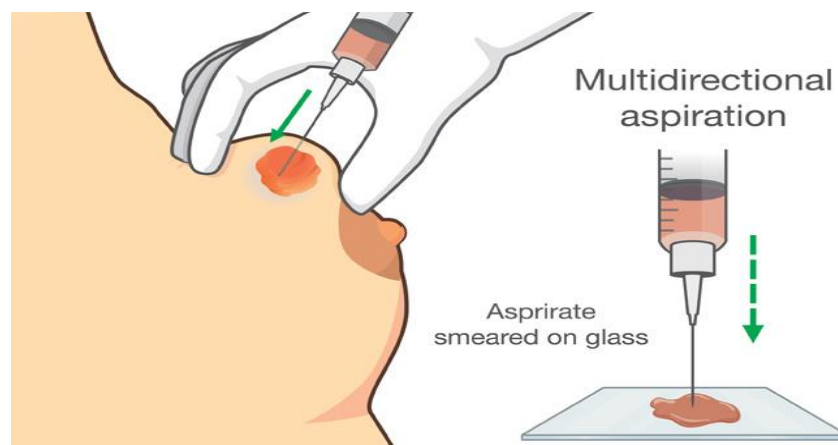
Các loại ung thư phổ biến nhất bao gồm ung thư phổi, ung thư vú, ung thư đại trực tràng, ung thư gan và ung thư dạ dày. Tuy nhiên, tần suất mắc và tỷ lệ tử vong vì ung thư cũng phụ thuộc vào độ tuổi, giới tính, chế độ ăn uống, môi trường sống, di truyền và những yếu tố khác.

Trên thế giới, có rất nhiều nỗ lực để giảm thiểu tác động của bệnh ung thư, từ việc phát triển phương pháp chẩn đoán và điều trị mới, đến việc tăng cường tư vấn và giáo dục cho người dân về cách phòng ngừa bệnh ung thư. Tuy nhiên, bệnh ung thư vẫn đang là một thách thức lớn đối với ngành y tế và toàn xã hội trên toàn thế giới.

Chẩn đoán ung thư vú là một trong những bài toán quan trọng trong lĩnh vực y tế. Theo thống kê toàn cầu về ung thư, ung thư vú chiếm tỷ lệ cao nhất trong số các loại ung thư phát hiện ở phụ nữ trên toàn thế giới. Mặc dù đã có sự tiến bộ về công nghệ và phương pháp chẩn đoán ung thư vú, nhưng việc phát hiện sớm và chẩn đoán chính xác vẫn là thách thức đối với các chuyên gia y tế. Theo Hiệp hội ung thư Hoa Kỳ thì năm 2022 có tới 1,919,030 ca ung thư mới và 609,360 ca tử vong tính riêng ở Hoa Kỳ. Ung thư vú chiếm tới 19% trong tổng số ca ung thư và chiếm 30% trong số ca ung thư ở phụ nữ.

Hơn nữa, ung thư vú còn tăng tỷ lệ mỗi năm trong khoảng 0.5% tính từ giữa những năm 2000. Chụp nhũ ảnh là một phương pháp kiểm tra ung thư ở cả 2 phía, kết hợp với khám sức khỏe giúp sàng lọc ung thư hiệu quả. Kết quả cho thấy rằng tỷ lệ tử vong do ung thư vú đã giảm đi 42% kể từ 1989 [2].

Có hai dạng tổn thương nhũ đó là tổn thương lành tính (benign) và ác tính (malignant). Các tổn thương lành tính thường ít được quan tâm trừ khi nếu chúng xâm lấn qua các mô xung quanh hoặc gây chèn ép tới các cơ quan quan trọng gần đó như mạch máu hoặc dây thần kinh [3]. Các tổn thương ác tính thì không ổn định và thường liên quan đến ung thư vú, trong trường hợp di căn, tế bào ung thư trong vú có thể tách khỏi khối u ban đầu và lưu thông qua hệ thống mạch máu hoặc mạch lympho đến các bộ phận khác của cơ thể. Khi một tổn thương đáng ngờ được xác định khi chụp quang tuyến vú, mức độ nghiêm trọng của tổn thương đó luôn được xác nhận lại bằng sinh thiết.

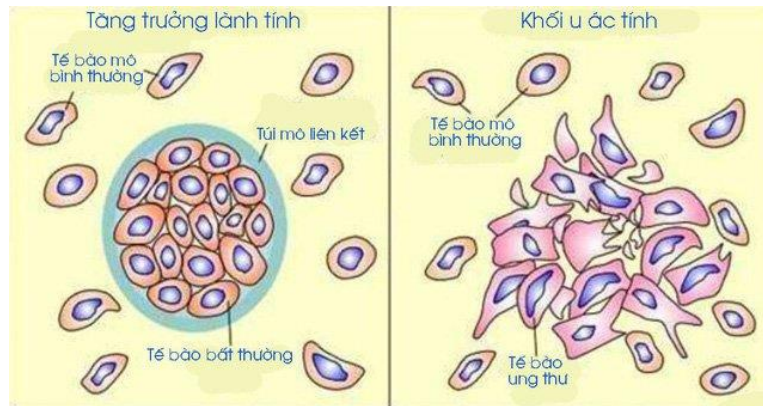


Hình 1.1. Quá trình sinh thiết¹

Ung thư vú phát sinh do sự phát triển bất thường của các tế bào trong mô vú, tạo nên các tổn thương. Chúng gồm sự bất đối xứng giữa hai ngực phải và trái, biến dạng cấu trúc

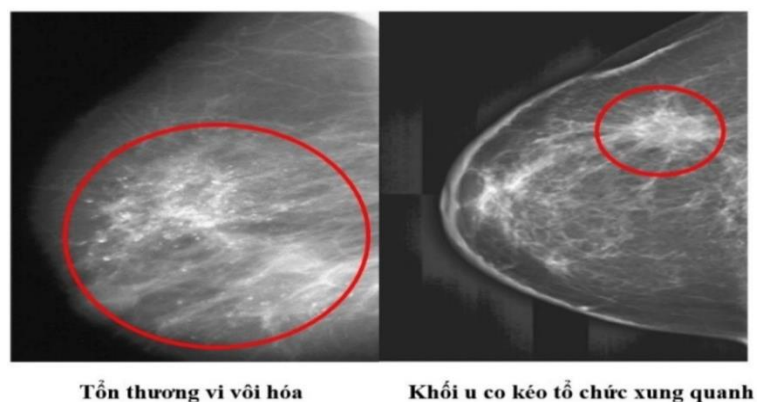
¹ <https://benhvienungbuouhanoi.vn/kien-thuc-ung-thu-1/cac-phuong-phap-sinh-thiet-va-xet-nghiem-te-bao-hoc-de-phat-hien-ung-thu.html>

bình thường của mô, sự xuất hiện của vi vôi hóa “Micro-Calcification” (MCs) và các khối có kích thước, hình dạng khác nhau.



Hình 1.2. Minh họa về tổn thương lành tính (trái) và tổn thương ác tính (phải) [4]

Hình ảnh chụp quang tuyến vú được các chuyên gia xem xét để xác định được xem là có bất kỳ dấu hiệu ác tính nào không. Vi vôi hóa là những chất canxi cực nhỏ thường xuất hiện ở vú, chúng thường là những điểm sáng trên phim chụp nhũ ảnh do hệ số suy giảm tia X của canxi cao hơn so với các mô bình thường. Hầu hết các vi vôi hóa là lành tính. Tuy nhiên, cụm vi vôi hóa (Micro-Calcification Clusters) được coi là tiền thân của bệnh ung thư.



Hình 1.3. Minh họa về vi vôi hóa ở nhũ ảnh²

² <https://thanhvien.vn/kiem-soat-nguy-co-ung-thu-gia-dinh-do-dot-bien-gien-1851488929.htm>

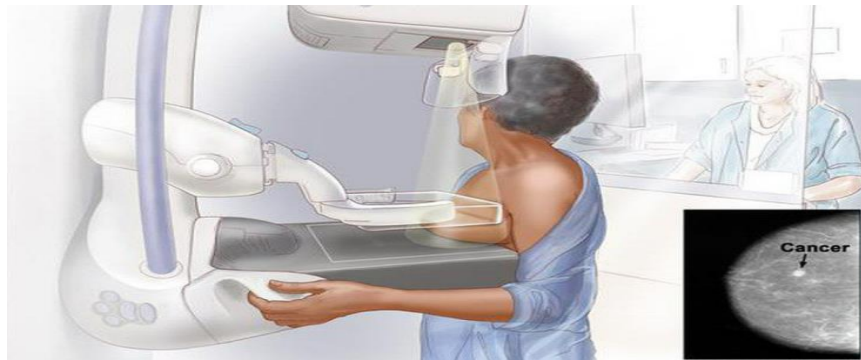
Vôi hóa lành tính thường to hơn, tròn hơn, số lượng ít hơn và thường có hình dạng đồng nhất. Vôi hóa của giai đoạn tiền ung thư thường là theo cụm, nhỏ, có hình dáng và kích thước không đều và phân nhánh theo hướng. Một phát hiện nữa từ nhũ ảnh đó là các khối u. Các khối u trong vú thường xuất hiện trong nhũ ảnh dưới dạng là các vùng khá dày đặc. Nó có thể được phân loại là lành tính hoặc nghi ngờ bị ác tính (suspicious), tùy thuộc vào thông số như kích thước, chu vi, mật độ,... Thông thường thì khối u lành tính có hình tròn, nhẵn và có ranh giới rõ ràng, ngược lại một khối u đang nghi ngờ thì có ranh giới gồ ghề và mờ.

Để đưa ra kết quả chẩn đoán chính xác, các bác sĩ cần phải thực hiện nhiều bước kiểm tra và xét nghiệm, từ kiểm tra cảm giác và xem độ mềm của tuyến vú đến siêu âm và chụp cộng hưởng từ (MRI) [4]. Điều này đòi hỏi sự tập trung cao độ và mất thời gian, đồng thời có thể xảy ra những sai sót trong quá trình kiểm tra.

Hiện nay yêu cầu hai bác sĩ X-quang đánh giá kết quả chụp (và cả người thứ ba, nếu không có sự đồng thuận), đây là dấu hiệu cho thấy những thách thức ngay cả các chuyên gia trong lĩnh vực này cũng phải gặp trong chụp quang tuyến vú. Nhũ bình thường hoặc các tổn thương lành tính có thể bị nhầm thành ung thư vú vì nó có các đặc điểm tương tự như vết thương ác tính. Phân loại khối u là một trong những thách thức, không phải vì độ tương phản thấp mà còn về hình dạng, kích thước. Hơn nữa là chúng thường được vây quanh bởi các cơ quan, cấu trúc quan trọng khác như mạch máu, cơ và các mô bình thường. Trong vài trường hợp thì ung thư có thể có trong nhũ ảnh nhưng bị các bác sĩ bỏ sót do các đặc điểm nhỏ hoặc do cường độ cao của vú dẫn đến giảm độ nhạy của nhũ ảnh. Tỷ lệ dương tính giả khi chụp quang tuyến vú, được thực hiện tại Mỹ trong năm 2013 là khoảng 61%.

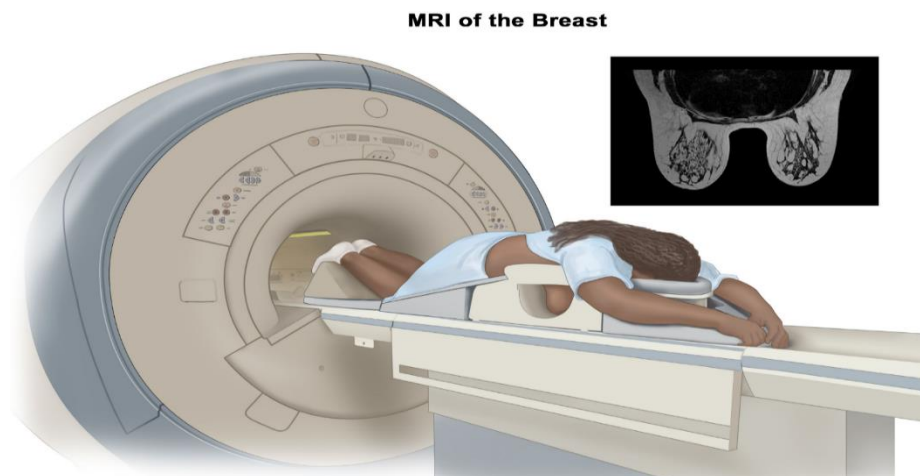
1.2 Các kĩ thuật chụp nhũ ảnh

Công cụ chính để xác định ung thư vú thường là chụp nhũ ảnh, sử dụng tia X năng lượng thấp để xác định tổn thương của vú. Chụp nhũ ảnh là phương pháp hiệu quả nhất được sử dụng để xác định xem khối u ở vú là lành tính hay ác tính.



Hình 1.4 Mô phỏng quá trình chụp nhũ ảnh³

Ngoài ra, hình ảnh siêu âm, chụp ảnh nhiệt độ, mô bệnh học và chụp cộng hưởng từ (MRI) thường được sử dụng để chẩn đoán ung thư vú.



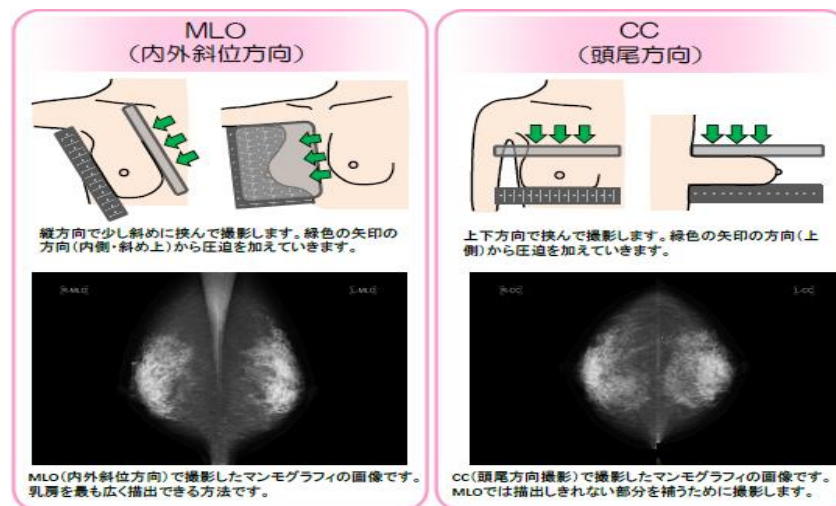
Hình 1.5. Hình ảnh chụp cộng hưởng từ (Magnetic Resonance Imaging (MRI))⁴

³ <https://ihope.vn/chup-nhu-anh/>

⁴ <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/magnetic-resonance-imaging>

Các chuyên gia nghiên cứu các phương pháp này và phát hiện sớm là mục tiêu chính. Chuyên gia kiểm tra cẩn thận các hình ảnh, phát hiện ung thư, nếu có, sau đó phân biệt các tế bào ung thư này là lành tính hay ác tính. Tuy nhiên, chuyên gia không thể dễ dàng phân biệt sự hiện diện của khối u hoặc sự khác biệt giữa hai khối u. Hình ảnh chụp nhũ ảnh có nhược điểm sẽ gây khó khăn cho bác sĩ X-quang khi phân tích hình ảnh. Nhược điểm chính của hình ảnh chụp nhũ ảnh là khó phân biệt được sự khác biệt giữa mô đặc và mô ung thư. Phân tích ung thư vú phụ thuộc vào kinh nghiệm của các nhà nghiên cứu bệnh học và gây ra các quyết định sai lầm do các yếu tố như mệt mỏi, giảm chú ý và cũng khá tốn thời gian.

Mỗi ngực sẽ có hai hình ảnh được chụp, tạo thành hai hình chiếu khác nhau là Cranio-Caudal (CC) và Mediolateral Oblique (MLO) (Hình 6). Trong hình ảnh CC, ảnh được chụp từ phía trên, trong khi ở MLO hình được chụp từ một bên theo một góc sao cho nó có thể nhìn thấy cơ ngực.



Hình 1.6. Hình ảnh mô phỏng 2 góc chụp CC và MLO⁵

⁵ 乳房X線検査 (マンモグラフィ) | 国立がん研究センター 中央病院 (ncc.go.jp)

Có thể coi tỷ lệ dày của mỡ chỉ định mức độ dày đặc của vú. Thang đo thể hiện mức độ dày đặc của vú là “Breast Imaging Reporting and Data System” (BI-RADS). Trên nhũ ảnh, mô tuyến vú được phân làm 4 loại:

- Loại 1: Mô vú hoàn toàn là mỡ.
- Loại 2: Có 1 ít mô tuyến vú, phần lớn là mô mỡ.
- Loại 3: Có một ít mô mỡ, phần lớn là mô sợi tuyến.
- Loại 4: Hầu hết là mô sợi tuyến.

Vú thuộc loại 3, 4 là thuộc mô tuyến vú dày. Độ dày càng cao thì càng khó để đánh giá hình ảnh do những sự xuất hiện tương tự của mô bình thường và mô bất thường. Theo hiệp hội ung thư Hoa Kỳ thì độ hiệu quả của chụp nhũ ảnh giảm xấp xỉ 30% dẫn đến nguy cơ tăng ung thư vú.



Hình 1.7. Hình mô tuyến vú được chụp từ nhũ ảnh (Loại 1: Hình A; Loại 2: Hình B; Loại 3: Hình C; Loại 4: Hình D)⁶

Để giải quyết các thách thức của việc đánh giá X-quang tuyến vú, các hệ thống chẩn đoán có sự hỗ trợ của máy tính (CAD) được phát triển với các công cụ tự động hoặc bán tự động. Mục tiêu chính là phát hiện những bất thường mà dễ bị bỏ qua. Tuy vậy, đây vẫn còn là một trong những nhiệm vụ rất phức tạp vì các vùng tổn thương có xu hướng

⁶ <https://densebreast-info.org/>

có kích thước nhỏ, hình dạng phân bố khác nhau, độ tương phản thấp so với các mô vú bình thường.

CAD là viết tắt của cụm từ tiếng Anh "Computer-Aided Detection" (phát hiện hỗ trợ bởi máy tính) là một kỹ thuật trong lĩnh vực y học, sử dụng các phương pháp xử lý hình ảnh và máy học để giúp các chuyên gia y tế phát hiện và chẩn đoán các bệnh lý [5]. Kỹ thuật CAD thường được áp dụng để hỗ trợ trong việc chẩn đoán ung thư, bao gồm cả ung thư vú, ung thư phổi, ung thư ruột kết và nhiều bệnh lý khác.

Với kỹ thuật CAD, các hình ảnh y tế được tải lên máy tính và xử lý bằng các thuật toán phân tích hình ảnh để tạo ra một bản đồ mô tả các vùng bất thường trong hình ảnh. Kết quả này sau đó được trình bày cho các chuyên gia y tế để phân tích và đưa ra quyết định chẩn đoán cuối cùng.

1.3 Mục tiêu nghiên cứu

Mục tiêu nghiên cứu của bài toán là nhận dạng ung thư có phải là ác tính hay không với độ chính xác cao và tốc độ xử lý nhanh. Bài toán này là một trong những bài toán y tế quan trọng nhất hiện nay, vì ung thư được coi là một trong những căn bệnh gây tử vong hàng đầu trên toàn thế giới.

Nghiên cứu về phân loại ung thư là cực kỳ quan trọng trong việc phát hiện và chữa trị bệnh sớm, cũng như tăng cường khả năng sống sót của bệnh nhân. Bằng cách sử dụng các kỹ thuật hình ảnh y tế như siêu âm, chụp cắt lớp vi tính (CT) hay chụp cộng hưởng từ (MRI), bác sĩ có thể nhìn thấy bên trong cơ thể bệnh nhân và xác định sự tồn tại của khối u hay các mô bất thường khác.

Tuy nhiên, việc đánh giá hình ảnh y tế là một công việc tốn thời gian và tốn sức, và đòi hỏi kiến thức chuyên môn cao. Vì vậy, các kỹ thuật học máy và trí tuệ nhân tạo đang được áp dụng để phát triển các giải pháp tự động hóa việc phát hiện ung thư từ hình ảnh

y tế. Nghiên cứu về phân loại ung thư từ hình ảnh y tế đang trở thành một lĩnh vực nghiên cứu quan trọng trong lĩnh vực y tế và khoa học máy tính.

Nghiên cứu sẽ sử dụng các kỹ thuật học máy và trí tuệ nhân tạo để phát triển một mô hình dự đoán phân loại khối u trên các hình ảnh y tế. Các kết quả nghiên cứu sẽ có thể cung cấp một giải pháp đáng tin cậy và tiết kiệm thời gian để phân loại ung thư và cải thiện chẩn đoán của bác sĩ.

1.4 Lí do chọn đề tài

Nhận thấy được thực trạng của bệnh ung thư vú hiện nay trên thế giới và nhằm giảm bớt khối lượng công việc của đội ngũ y tế trong việc phát hiện bệnh, nhóm chúng tôi đã quyết định chọn đề tài nghiên cứu này.

Từ lâu, người ta đã công nhận rằng chụp nhũ ảnh có khả năng phát hiện sớm các tế bào ung thư. Do đó, kỹ thuật phân đoạn hình ảnh và trí tuệ nhân tạo (AI) có thể được áp dụng để chẩn đoán ung thư vú giai đoạn đầu.

Chúng tôi đã áp dụng các kỹ thuật trí tuệ nhân tạo để hỗ trợ chẩn đoán phân loại ung thư vú. Các mô hình học sâu được xây dựng dựa trên dữ liệu hình ảnh của tuyến vú giúp phát hiện bệnh nhân có ung thư hay không giúp các bác sĩ xác định những trường hợp cần phải tiếp tục xét nghiệm và chẩn đoán chính xác hơn...Các mô hình này có thể giúp tăng độ chính xác và giảm thời gian kiểm tra, đồng thời giúp giảm thiểu những sai sót trong quá trình chẩn đoán.

1.5 Phạm vi nghiên cứu

- Kiến thức và hiểu biết về CNNs, ResNet50, ConvNext, HRNet, pre-training và fine-tune model.
- Dữ liệu sử dụng: gồm 5 tập dữ liệu đó là DDSM, RSNA, INbreast, MIAS và CMMD.

1.6 Ý nghĩa khoa học và thực tiễn

Bài toán phân loại chẩn đoán ung thư từ nhũ ảnh có ý nghĩa vô cùng quan trọng từ khía cạnh khoa học và thực tiễn.

1.6.1 Ý nghĩa khoa học

- Nghiên cứu về phân loại ảnh ung thư vú sử dụng mô hình học sâu (deep learning) đóng góp vào việc phát triển và tiến bộ của lĩnh vực trí tuệ nhân tạo và học máy trong lĩnh vực y tế.
- Xây dựng mô hình học sâu để phân loại ảnh ung thư vú dựa trên ảnh y tế yêu cầu nghiên cứu và áp dụng các phương pháp tiên tiến như mạng nơ-ron, trí tuệ nhân tạo,...giúp nâng cao hiệu suất và chính xác của phương pháp phân loại.
- Khóa luận góp phần mở rộng kiến thức về ứng dụng deep learning trong lĩnh vực y tế, đặc biệt là trong phân loại ung thư vú và cung cấp những phương pháp tiên tiến mới cho cộng đồng nghiên cứu.

1.6.2 Ý nghĩa thực tiễn

- Phân loại ảnh ung thư vú sử dụng mô hình deep learning giúp nâng cao khả năng chẩn đoán và phát hiện ung thư vú sớm, từ đó cung cấp cơ hội điều trị bệnh sớm.
- Mô hình học sâu có thể được triển khai trong các hệ thống y tế tự động, giúp tiết kiệm thời gian và công sức cho các chuyên gia y tế trong việc đánh giá và phân loại ung thư vú.
- Kết quả của khóa luận sẽ góp phần tối thiểu được thời gian chẩn đoán bệnh, giúp các chuyên gia y tế đưa ra quyết định điều trị và giám sát bệnh nhân một cách chính xác và hiệu quả.

1.7 Bố cục luận văn

Chương 1: tổng quan về lĩnh vực nghiên cứu và sơ lược tổng quan về vấn đề nghiên cứu trên phương diện tổng quan. Nêu ra mục tiêu, đối tượng, phương pháp nghiên cứu và bố cục luận văn.

Chương 2: nêu các cơ sở lý thuyết và các nghiên cứu liên quan đến đề tài. Giới thiệu tổng quan về học sâu, các mô hình CNN, ResNet50, ConvNeXt và HRNet. Trình bày các nghiên cứu đã có về lĩnh vực phát hiện ung thư vú và nêu ra các phương pháp đánh giá kết quả.

Chương 3: trình bày tổng quan về mô hình đề xuất và hướng tiếp cận bài toán. Phân tích các đặc trưng của mô hình đề xuất, các thuật toán, kỹ thuật khi giải quyết bài toán.

Chương 4: giới thiệu về các bộ dữ liệu đã sử dụng, các phương pháp tiền xử lý dữ liệu và các kỹ thuật xử lý ảnh. Trình bày về các phương pháp xử lý dữ liệu mất cân bằng, một số công nghệ sử dụng trong quá trình huấn luyện bài toán. Bên cạnh đó còn nêu ra quá trình thực nghiệm và các kết quả của các mô hình sau nhiều lần thực nghiệm khác nhau.

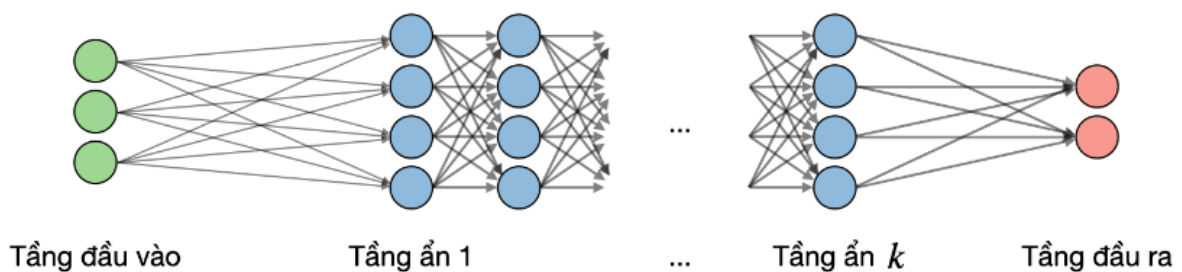
Chương 5: kết luận và hướng phát triển.

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN

2.1 Học sâu

Học sâu (còn được gọi là học cấu trúc sâu) là một phần trong một nhánh rộng hơn các phương pháp học máy dựa trên mạng thần kinh nhân tạo (mạng nơ-ron) kết hợp với việc học biểu diễn đặc trưng (representation learning). Việc học này có thể có giám sát, nửa giám sát hoặc không giám sát.

Mạng thần kinh nhân tạo là một lớp của các mô hình (models) được xây dựng với các tầng (layers). Các loại mạng thần kinh nhân tạo được sử dụng bao gồm: mạng thần kinh tích chập (Convolutional Neural Networks) và mạng thần kinh hồi quy (Recurrent Neural Networks). Một mạng nơ-ron bao gồm hơn ba lớp - sẽ bao gồm đầu vào và đầu ra - có thể được coi là một thuật toán học sâu.



Hình 2.1. Kiến trúc học sâu⁷

Một mạng thần kinh nhân tạo gồm có 3 phần:

- Lớp đầu vào (input layer): Một mạng nơ-ron nhân tạo sẽ có một số nút để nhập dữ liệu đầu vào. Các nút này tạo nên lớp đầu vào của hệ thống.
- Lớp ẩn (hidden layer): Đây là tầng nằm ở giữa, thể hiện cho quá trình xử lý thông tin và suy luận của mạng. Nó sẽ nhận các thông tin đầu vào ở lớp đầu vào và trả kết quả ở đầu ra thông qua chức năng của hàm kích hoạt.

⁷ [CS 229 - Deep Learning Cheatsheet \(stanford.edu\)](https://stanford.edu/~cs229/CS229-Deep-Learning-Cheatsheet)

- Lớp đầu ra (output layer): Lớp đầu ra bao gồm các nút xuất dữ liệu. Các mô hình học sâu xuất ra đáp án "có" hoặc "không" chỉ có hai nút trong lớp đầu ra. Mặt khác, các mô hình xuất ra nhiều đáp án hơn sẽ có nhiều nút hơn.

Bằng việc kí hiệu i là tầng thứ i của mạng, j là hidden unit (đơn vị ẩn) thứ j của tầng i , w , b , z tương ứng là trọng số (weight), bias và đầu ra ta có công thức khi đi qua nút mạng thứ i :

$$z_j^{[i]} = w_j^{[i]T} x + b_j^{[i]} \quad (PT\ 2.1)$$

2.2 Các nghiên cứu đã có

Về cơ bản có hai kỹ thuật được sử dụng để phát hiện ung thư vú, đó là học máy (machine learning) và học sâu. Học máy là một lĩnh vực của trí tuệ nhân tạo, tuy nhiên, học máy hiệu quả đối với lượng dữ liệu nhỏ, khi có nhiều dữ liệu hơn thì kết quả mô hình cũng không cải thiện hơn. Trong khi đó các thuật toán học sâu sẽ cho ra kết quả tốt hơn khi được đào tạo trên dữ liệu chất lượng cao theo số lượng lớn. Dưới đây là một số nghiên cứu đã có:

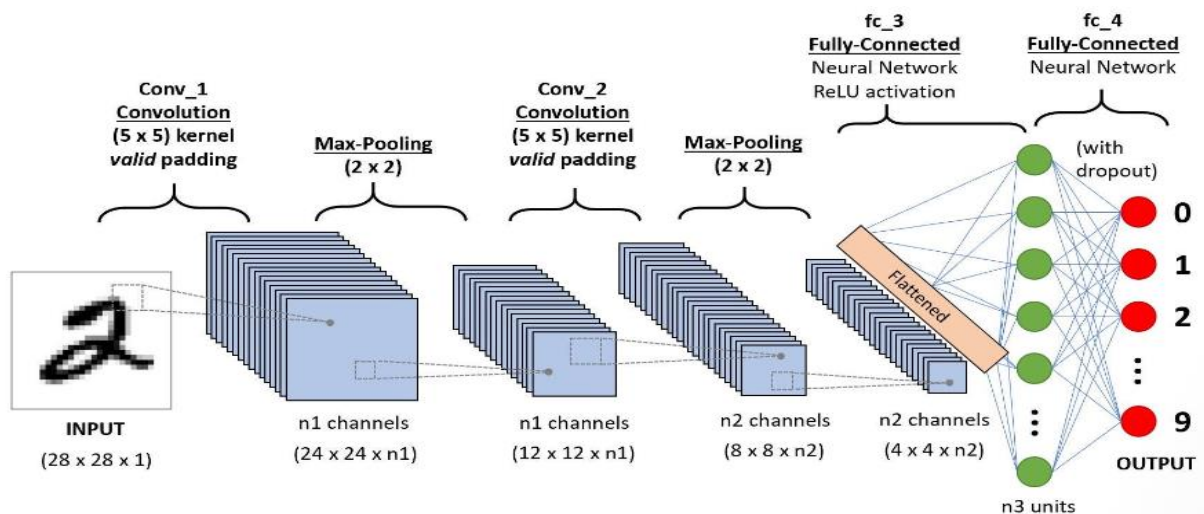
- Megha Rathi và cộng sự [6] đã đề xuất mô hình dựa trên phương pháp kết hợp sử dụng máy học để dự đoán ung thư vú. Ông đã thực hiện phương pháp này bằng cách sử dụng trích chọn đặc trưng dữ liệu bằng phương pháp MRMR kết hợp với bốn thuật toán phân loại để tìm ra kết quả tốt nhất. Tác giả đã sử dụng bốn bộ phân loại SVM, Naïve Bays, Function tree và End Meta và thực hiện so sánh giữa các mô hình với nhau. Kết quả chỉ ra rằng, SVM là thuật toán phân loại hiệu quả đối với bài toán này.
- M.Tahmooresi và cộng sự [7] so sánh SVM, KNN, ANN và cây quyết định (Decision Tree), kết quả cho thấy SVM đạt được kết quả tốt nhất, bài toán được khai triển trên dữ liệu chứa ảnh và dữ liệu máu.

- Muhammet Fatih Aslan và cộng sự [8] cũng đề xuất mô hình học máy nhưng sử dụng thuật toán phân loại khác, đó là Extreme Learning Machine (ELM) để so sánh với SVM, KNN và ANN. Kết quả có 1 chút thay đổi, đó là ELM cho kết quả tốt hơn.
- Anusha Bharat và cộng sự [9] đề xuất mô hình dùng bốn thuật toán phân loại là SVM, Decisiontree (CART), KNN, Naive Bayes. Theo tác giả thì KNN cho kết quả tốt hơn. SVM có một hạn chế đó là cho kết quả tốt hơn đối với bài toán phân loại nhị phân. Đó là lý do Multi-SVM được sử dụng.
- Ebru Ayndindag Bayrak và cộng sự [10] sử dụng WEKA với dữ liệu là Wisconsin Breast cancer. Kết quả cho thấy SVM đạt kết quả tốt hơn.
- Shewtha K và cộng sự [11] đề sử dụng 2 mô hình học sâu đó là Mobile Net và Inception V3. Kết quả cho thấy INception V3 độ chính xác cao hơn.
- Ch. Shravya và cộng sự [12] đề xuất mô hình học máy có giám sát. Nghiên cứu này được thực hiện sử dụng các thuật toán học máy phân loại có giám sát như Logistic Regression, SVM và KNN. Dữ liệu được tải từ UCI và cho thấy thuật toán SVM đạt kết quả cho tới 92.7 %, với nền tảng sử dụng là ngôn ngữ python.
- Sivapriya J và cộng sự [13] đề xuất sử dụng học máy sử dụng các thuật toán phân loại là Random Forest, SVM, Logistic và Naive Bayes. Thực hiện trên môi trường anaconda và cho thấy rằng Random Forest là 1 bộ phân loại tốt với độ chính xác tới 99.76%.
- Kalyani Wadkar và cộng sự [14] sử dụng mô hình ANN cơ bản và hiệu suất phân tích bằng với mô hình SVM. Kết quả cho thấy ANN có độ chính xác là 97% còn SVM chỉ 91%. Tác giả kêu rằng nếu ko có SVM thì ANN sẽ cho kết quả tốt hơn.
- Vishal Deshwal và cộng sự [15]: đầu tiên tác giả chỉ nghiên cứu sử dụng SVM và sau đó là SVM kết hợp Grid search. Kết quả so sánh cho thấy nếu có Grid search thì SVM sẽ hiệu quả hơn.

- S. Shamy và cộng sự [16]: sử dụng kmean - GMM và CNN. Đầu tiên tác giả tìm phần ROI sau đó áp dụng phương pháp trích xuất đặc trưng, cuối cùng là sử dụng CNN để đạt kết quả tốt nhất. Kết quả đạt độ chính xác 95.8% trên bộ MIAS.

2.3 Mô hình Convolution Neural Networks (CNNs)

CNNs/ConvNet là một thuật toán học sâu nhận dữ liệu đầu vào (input) là ảnh. Trong mạng thần kinh tích chập có các lớp chuyên dụng để xử lý dữ liệu ảnh như lớp tích chập (Convolutional), lớp tổng hợp (Pooling), lớp kích hoạt (Activation) và lớp kết nối đầy đủ (Fully connected). Lớp tích chập có thể tìm kiếm các đặc trưng quan trọng của ảnh thông qua tính toán nhờ các bộ lọc (filter), lớp Pooling để tổng hợp chỉ lấy các đặc trưng quan trọng nhất và các lớp activation như ReLU có thể giúp tăng tính phi tuyến của mô hình và học được các đặc trưng phức tạp của ảnh [17].



Hình 2.2. Minh họa kiến trúc CNNs giải quyết bài toán phân loại chữ viết tay [18]

Đầu tiên ta cần hiểu một vài khái niệm sau:

- Stride: Đối với phép tích chập hoặc phép pooling, độ trượt S ký hiệu số pixel mà cửa sổ sẽ di chuyển sau mỗi lần thực hiện phép tính:



Hình 2.3. Minh họa về stride $S=2$ [18]

- Zero padding: là tên gọi của quá trình thêm P số 0 vào các biên của đầu vào. Lí do chọn số 0 mà không phải số khác là do khi padding với giá trị 0, không có thay đổi về giá trị của các điểm ảnh trong ảnh gốc. Điều này đảm bảo rằng các điểm ảnh gốc không bị ảnh hưởng hoặc biến đổi khi thực hiện các phép toán nhân tích chập. Giá trị này có thể được lựa chọn thủ công hoặc một cách tự động bằng một trong ba phương pháp mô tả dưới đây:

Phương pháp	Valid	Same	Full
Giá trị	$P = 0$	$P_{start} = \left\lfloor \frac{\left\lceil \frac{I}{S} \right\rceil - I + F - S}{2} \right\rfloor$ $P_{end} = \left\lfloor \frac{S \left\lceil \frac{I}{S} \right\rceil - I + F - S}{2} \right\rfloor$	$P_{start} \in [0, F - 1]$ $P_{end} = F - 1$
Minh họa			
Mục đích	<ul style="list-style-type: none"> • Không sử dụng padding • Bỏ phép tích chập cuối nếu số chiều không khớp 	<ul style="list-style-type: none"> • Sử dụng padding để làm cho feature map có kích thước $\left\lceil \frac{I}{S} \right\rceil$ • Kích thước đầu ra thuận lợi về mặt toán học 	<ul style="list-style-type: none"> • Padding tối đa sao cho các phép tích chập có thể được sử dụng tại các rìa của đầu vào

		<ul style="list-style-type: none"> • Còn được gọi là 'half padding' 	<ul style="list-style-type: none"> • Bộ lọc 'thấy' được đầu vào từ đầu đến cuối
--	--	--	--

Tầng tích chập sử dụng các bộ lọc để thực hiện phép tích chập đưa chúng đi qua đầu vào I theo các chiều của nó. Các siêu tham số của tầng này bao gồm kích thước bộ lọc F (filter) và độ trượt (stride) S . Kết quả đầu ra O được gọi là feature map hay activation map. Công thức để tính tích chập là:

$$a \cdot b = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \quad (PT\ 2.2)$$

Với a là ma trận hình ảnh, b là ma trận bộ lọc (filter), a_i, b_i là các phần tử trong 2 ma trận. Công thức để tính kích thước O của feature map theo một chiều là:

$$O = \frac{I - F + P_{start} + P_{end}}{S} + 1 \quad (PT\ 2.3)$$

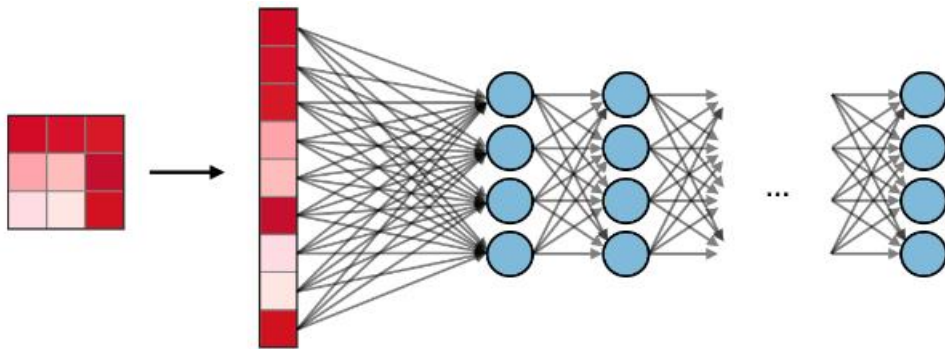
Trong đó P là số lượng zero padding, I là kích thước của hình ảnh, F là kích thước của filter.

*Lưu ý: trong một số trường hợp $P_{start} = P_{end} = \Delta$ ta có thể thay thế $P_{start} + P_{end}$ bằng $2P$.

Tầng Pooling là một phép downsampling, thường được sử dụng sau tầng tích chập, giúp tăng tính bất biến không gian. Cụ thể, max pooling và average pooling tương ứng là giá trị lớn nhất và giá trị trung bình được lấy ra. Lý do lớp Pooling cần để sau lớp convolution là vì lớp convolution tìm kiếm các đặc trưng quan trọng của ảnh và tạo ra đặc trưng mới, trong khi lớp Pooling giảm kích thước của đặc trưng này để giảm số lượng tham số và

tính toán trong mạng CNNs. Chính vì vậy, việc đặt lớp Pooling sau lớp Convolution là hợp lý.

Tầng kết nối đầy đủ (FC- Fully Connected) nhận đầu vào là các dữ liệu đã được làm phẳng (Flatten), mà mỗi đầu vào đó được kết nối đến tất cả nơ-ron. Trong mô hình mạng CNNs, các tầng kết nối đầy đủ thường được tìm thấy ở cuối mạng và được dùng trong bài toán phân lớp [18].



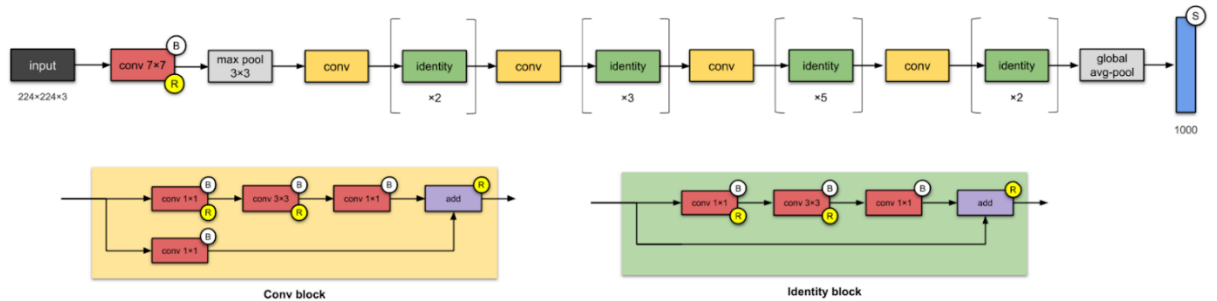
Hình 2.4. Minh họa về tầng kết nối đầy đủ

Tầng Rectified Linear Unit (ReLU) gồm một hàm kích hoạt ReLU được sử dụng trên tất cả các thành phần. Mục đích của nó là tăng tính phi tuyến tính cho mạng.

2.4 Mô hình Residual Network (ResNet)

ResNet (Residual Network) [19] là một mạng CNNs được thiết kế để làm việc với hàng trăm lớp. ResNet được giới thiệu đến công chúng vào năm 2015 và có rất nhiều biến thể của kiến trúc ResNet với số lớp khác nhau như ResNet-18, ResNet-50,... Một vấn đề xảy ra khi xây dựng mạng CNN nhiều lớp tích chập sẽ xảy ra hiện tượng Vanishing Gradient [20] và phân rã Gradient (gradient degradation). Khi mạng nơ-ron trở nên sâu hơn, việc huấn luyện trở nên khó khăn do gradient truyền ngược giảm dần theo các lớp, dẫn tới quá trình học tập của mô hình không tốt. Mạng ResNet ra đời cũng để giải quyết vấn đề đó.

ResNet có khối tích chập (Convolutional Block, chính là Conv Block trong hình) sử dụng bộ lọc kích thước 3x3 giống với của InceptionNet, gồm 2 nhánh tích chập. Trong đó một nhánh áp dụng tích chập 1x1 trước khi cộng trực tiếp vào nhánh còn lại và khối Identity (còn được gọi là skip connection hoặc khối residual) thì không áp dụng tích chập 1x1 mà cộng trực tiếp vào nhánh còn lại.

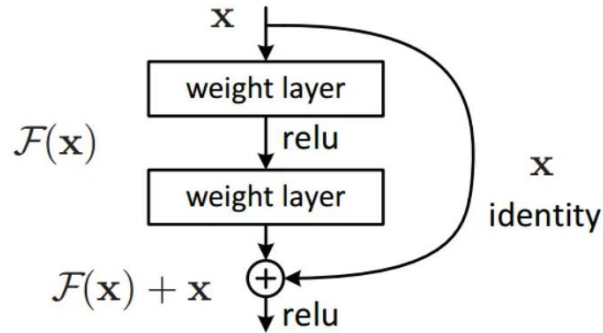


Hình 2.5. Kiến trúc mạng ResNet-50 gồm 2 khối đặc trưng là khối tích chập (Conv Block) và khối Identity [20]

Trong hình trên, Identity x 3 có nghĩa là có 3 khối Identity chồng lên nhau. Chi tiết kiến trúc ta có thể hiểu như sau:

- Dữ liệu đầu vào: là hình ảnh có kích thước 224x224x3.
- Conv 7x7 (B, R): là tích chập với 64 filter, sử dụng stride (2, 2), BatchNorm (B) và hàm kích hoạt ReLU (R) , MaxPooling (3,3).
- Identity x2: tích chập sử dụng 3 filter với kích thước 64x64x256, f=3, s=1. Có 2 khối Identity với kích thước filter 64x64x256, f=3.
- Identity x3: tích chập sử dụng 3 filter với kích thước 128x128x512, f=3, s=2. Có 3 khối Identity với kích thước filter 128x128x512, f=3.
- Identity x5: tích chập sử dụng 3 filter với kích thước 256x256x1024, f=3, s=2. Có 5 khối Identity với kích thước filter 256x256x1024, f=3.
- Identity x2: tích chập sử dụng 3 kích thước filter 512x512x2048, f=3, s=2. Có 2 khối Identity với kích thước filter 512x512x2048, f=3

- Global avg-pool: 2D Average Pooling sử dụng kích thước (2,2)
- Flatten
- Fully Connected (Dense): sử dụng hàm kích hoạt Softmax



Hình 2.6 Khối Residual [20]

Hình trên hiển thị khối Residual được sử dụng trong mạng. Xuất hiện một mũi tên cong xuất phát từ đầu và kết thúc tại cuối khối dư. Hay nói cách khác là sẽ bổ sung input \mathbf{X} vào đầu ra của layer, hay chính là phép cộng mà ta thấy trong hình minh họa, việc này sẽ giải quyết được vấn đề đạo hàm bằng 0, do vẫn còn cộng thêm \mathbf{X} . Tại một khối Identity sẽ áp dụng một lớp kích hoạt ReLU sau mỗi xen kẽ giữa những tầng trọng số.

Mặc dù có kiến trúc Residual Block từ GoogleNet nhưng ResNet lại dễ tóm tắt và triển khai hơn rất nhiều vì kiến trúc cơ sở của nó chỉ gồm các khối tích chập và khối identity.

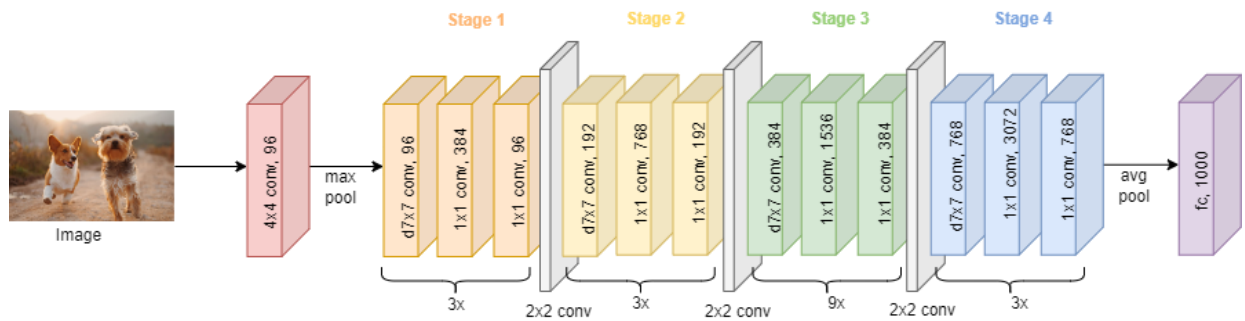
Cuối cùng, ở lớp Fully Connected sử dụng hàm kích hoạt softmax [21]. Hàm kích hoạt này giúp tính toán xác suất tối đa trong tham số của mô hình. Xác suất sẽ luôn nằm trong khoảng $(0,1]$ và tổng tất cả các xác suất bằng 1.

2.5 Mô hình ConvNeXt

Cuộc chạy đua công nghệ bắt đầu với sự ra đời của Vision Transformers (ViTs), nhanh chóng bị thay thế bởi kiến trúc ConvNets- được xem là mô hình "state-of-the-art" trong các mô hình phân loại ảnh. Mặc dù ViTs làm rất tốt bài toán phân loại hình ảnh, tuy nhiên ViTs lại gặp khó khăn trong xử lý bài toán nhận diện đối tượng hoặc phân đoạn

hình ảnh. Sự ra đời của kiến trúc Swin Transformer được giới thiệu như một sự kết hợp giữa kiến trúc CNNs và Transformer, và đem lại hiệu suất tốt hơn so với các phiên bản khác của Transformer trên nhiều tác vụ nhận dạng hình ảnh, bao gồm cả bài toán nhận diện đối tượng và phân đoạn hình ảnh. Tuy nhiên, tính hiệu quả của sự kết hợp này chủ yếu là do tính ưu việt của bản thân kiến trúc Transformer, chứ không phải do sự ảnh hưởng của các bias inductive của convolution. Sau đó nhóm tác giả đã cải tiến một mô hình ResNet theo hướng thiết kế của mô hình Swin Transformer, nghiên cứu những "khóa chính" tạo nên sự hiệu quả của mô hình. Kết quả của quá trình nghiên cứu này là một mô hình ConvNet được đặt tên là ConvNeXt được xây dựng hoàn toàn từ các mô-đun ConvNet thông thường [22].

ConvNeXt đã đạt state-of-the-art với độ chính xác cao trên tập ImageNet (87,8% - top 1 về độ chính xác) cũng như hiệu quả trong các tác vụ như phát hiện đối tượng trên tập COCO, phân đoạn hình ảnh trên tập ADE20K, đồng thời vẫn giữ được sự đơn giản và hiệu quả của các mô hình ConvNet. Điều này cho thấy ConvNeXt là một giải pháp hữu hiệu để giải quyết các vấn đề trong thị giác máy tính và cải thiện hiệu suất mô hình.



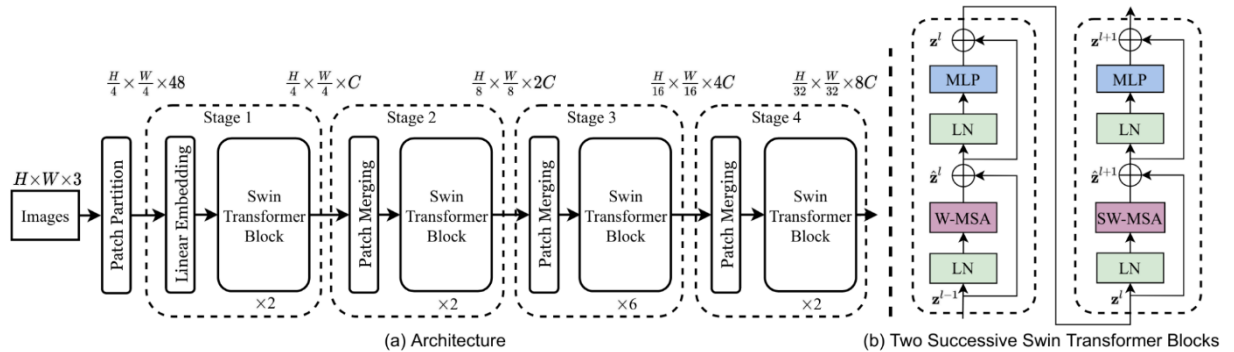
Hình 2.7. Kiến trúc ConvNeXt (trong đó bỏ qua không miêu tả các skip connection)⁸

Để tìm ra kiến trúc ConvNeXt, nhóm tác giả đã bắt đầu thử nghiệm với một mô hình ResNet50 và sử dụng các kỹ thuật huấn luyện tương tự như với mô hình ViTs. Kết quả

⁸ [ConvNeXt \(bertelsmann.com\)](https://arxiv.org/abs/2201.00097)

cho thấy cải thiện đáng kể so với mô hình ResNet50 gốc. Dưới đây là những cải tiến của ConvNeXt so với ResNet50 khiến mô hình ConvNeXt trở nên vượt trội (tổng quan độ chính xác tăng đến 3.2%):

- Về kỹ thuật training: Mô hình sử dụng hàm tối ưu là AdamW, với các kỹ thuật tăng cường ảnh như Mixup [23], Cutmix [24], RandAugment [25], Random Erasing [26], regularization schemes bao gồm Stochastic Depth và Label smoothing. Kết quả cho thấy, chỉ bằng cách cải thiện phương pháp huấn luyện, hiệu suất của mô hình ResNet-50 đã tăng từ 76,1% lên 78,8% (+2,7%). Điều này cho thấy rằng một phần đáng kể của sự khác biệt về hiệu suất giữa các ConvNets truyền thống và vision Transformers có thể do kỹ thuật huấn luyện.
- Điều chỉnh tỉ lệ giữa các stage: Cụ thể, đối với mô hình Swin-Transformer, tỉ lệ các stage trong các block là 1 : 1 : 3 : 1 và large-Swin-Transformer là 1 : 1 : 9 : 3. Tác giả đã đổi số tỉ lệ số stage trong ResNet50 theo tỉ lệ của Swin-Transformer từ (3, 4, 6, 3) thành (3, 3, 9, 3). Đáng chú ý, điều này đã cải thiện độ chính xác của mô hình từ 78.8 % thành 79.4 %.

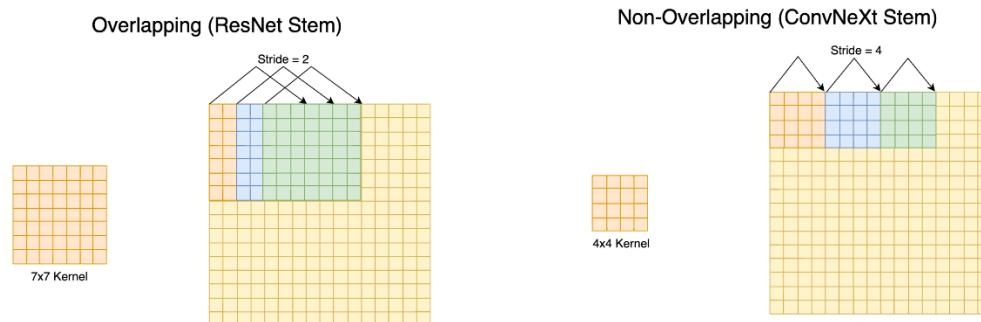


Hình 2.8. Kiến trúc Swin-Transformer tiêu chuẩn⁹.

- Thay đổi phương pháp “stem” thành “patchify”:
Đầu tiên, ta cần hiểu phương pháp stem và patchify là gì. Patchify là một phương pháp xử lý ảnh trong đó ảnh được chia thành các phần bằng nhau (gọi là các "patch")

⁹ [Swin-Transformer — MMLab 1.0.0rc8 documentation](#)

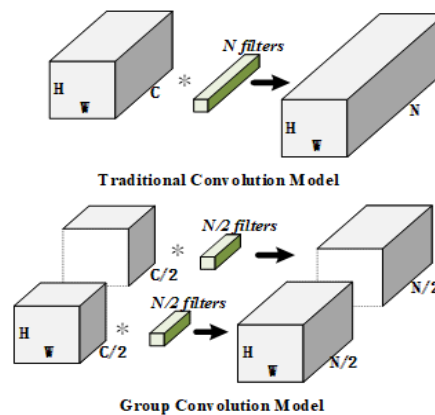
và mỗi patch được xử lý độc lập. Thay vì sử dụng lớp tích chập với một kernel lớn để xử lý toàn bộ ảnh (stem), Patchify sử dụng các kernel nhỏ hơn để xử lý các patch riêng lẻ.



Hình 2.9. Minh họa về phương pháp 'stem' (Resnet Stem) và patchify (ConvNeXt stem)

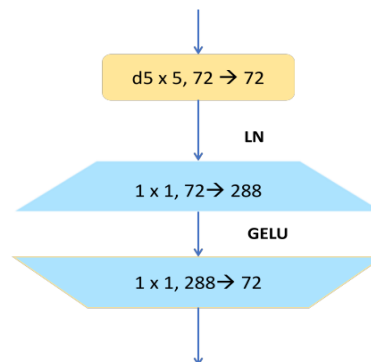
Đối với mô hình ResNet50 truyền thống, “stem cell” bao gồm một lớp tích chập 7x7 với stride bằng 2, theo sau là một lớp max pool, khiến cho kích thước ảnh giảm xuống 4 lần. Nhóm tác giả đã thay thế phương pháp stem của ResNet bằng phương pháp patchify với kernel size là 4, stride bằng 4 giống như của Swin Transformer sử dụng. Kết quả cho thấy độ chính xác mô hình thay đổi từ 79.4 % lên 79.5 %.

- Lấy ý tưởng từ ResNeXt: Nhóm tác giả thay đổi lớp tích chập 3x3 bằng lớp tích chập sâu theo chiều sâu (depthwise convolution). Lớp tích chập sâu theo chiều sâu thực chất là lớp tích chập nhóm (grouped convolution), trong đó số nhóm bằng số kênh của đầu vào. Thay thế lớp tích chập bằng lớp tích chập sâu giúp tăng độ chính xác của mô hình là 80.5% (tăng thêm 1,0%) và tăng GFLOPS thêm 0,9%.



Hình 2.10. Minh họa về group convolution¹⁰

- Inverted bottleneck, nghĩa là thay đổi kích thước của các lớp tích chập (convolutional layers) trong một khối để khối ở giữa có kích thước lớn nhất. Điều thú vị là, điều này giúp tăng độ chính xác của mô hình lên từ 80.5% thành 80.6%.



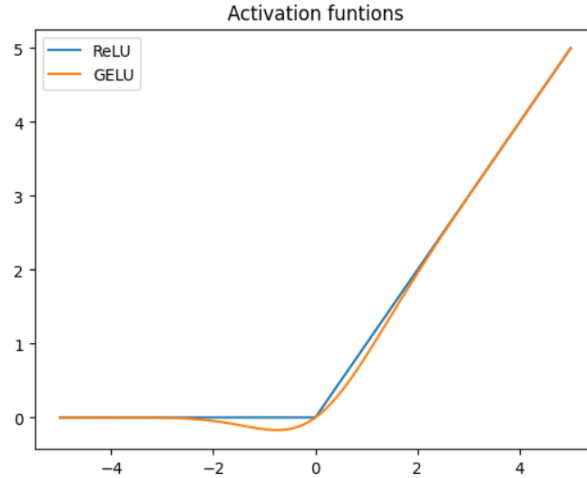
Hình 2.11. Minh họa một khối inverted bottleneck¹¹.

- Kích thước kernel tăng từ 3x3 thành 7x7 (giống với Swin Transformer).
- Sử dụng hàm kích hoạt GeLU thay vì ReLU. Hàm GELU (Gaussian Error Linear Unit) là một biến thể của hàm ReLU. Hàm GELU khá mượt mà và có tính chất khá tương đồng với hàm sigmoid, với giá trị đầu ra gần với 0 hoặc 1 khi đầu vào rất nhỏ

¹⁰ [Design of Efficient Convolutional Neural Module Based on An Improved Module - Advances in Science, Technology and Engineering Systems Journal \(astesj.com\)](#)

¹¹ [2301.01286.pdf \(arxiv.org\)](#)

hoặc rất lớn. Điều này giúp cho hàm GELU có thể học các mối quan hệ phức tạp giữa các đặc trưng đầu vào.



Hình 2.12. Minh họa hàm ReLU và GELU

Công thức của hàm GELU [27]:

$$g = 0.5x \left(1 + \frac{2}{\sqrt{\pi}} \int_0^{\frac{x}{\sqrt{2}}} e^{-t^2} dt \right) \quad (PT 2.4)$$

Có thể thực hiện xấp xỉ GELU bằng công thức sau:

$$g = 0.5x \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}} (x + 0.0447155x^3) \right] \right) \quad (PT 2.5)$$

Giảm số lượng hàm kích hoạt, chỉ giữ lại hàm kích hoạt sau inverted bottleneck. Độ chính xác +0.7%.

- Giảm số lượng lớp chuẩn hóa, chỉ giữ lại một lớp chuẩn hóa trước inverted bottleneck. Độ chính xác tăng 0.1%.
- Thay thế toàn bộ normalization (BN) thành Layer normalization (LN) +0.1%.
- Thêm các lớp downsampling vào giữa các stages của mạng Swin Transformer. Cụ thể, lớp downsampling này gồm một lớp chuẩn hóa (normalization), một lớp convolution với kích thước 2x2, bước nhảy (stride) bằng 2 và một lớp chuẩn hóa nữa.

Ngoài ra, thêm lớp chuẩn hóa trước giai đoạn đầu tiên và sau khi áp dụng lớp average pool sau stage 4. Kết quả là độ chính xác của mô hình được cải thiện thêm 0.5%.

Về tổng quan, có các kiến trúc ConvNeXt theo kích thước từ nhỏ tới lớn được giới thiệu dưới đây:

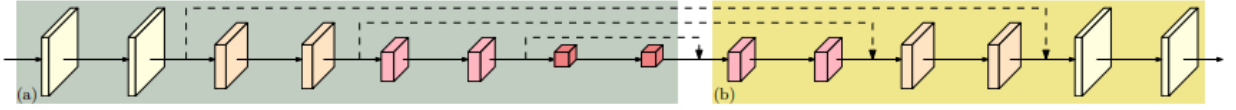
- ConvNeXt-tiny (ConvNeXt-T): $C = (96, 192, 384, 768)$, $B = (3, 3, 9, 3)$
- ConvNeXt-small (ConvNeXt-S): $C = (96, 192, 384, 768)$, $B = (3, 3, 27, 3)$
- ConvNeXt-base (ConvNeXt-B): $C = (128, 256, 512, 1024)$, $B = (3, 3, 27, 3)$
- ConvNeXt-large (ConvNeXt-L): $C = (192, 384, 768, 1536)$, $B = (3, 3, 27, 3)$
- ConvNeXt-Xtra large (ConvNeXt-XL): $C = (256, 512, 1024, 2048)$, $B = (3, 3, 27, 3)$

2.6 Mô hình High - Resolution Networks (HRNets)

2.6.1 Giới thiệu

Hầu hết mạng phân loại như Alexnet, VGGNet, GoogleNet, ResNet tuân theo quy tắc của Lenet (mô tả trong hình 2.13), tức là giảm dần kích thước không gian của các feature map, sau đó kết nối các tích chập từ độ phân giải cao đến độ phân giải thấp, sau đó mới được xử lý thêm bớt để phân loại.

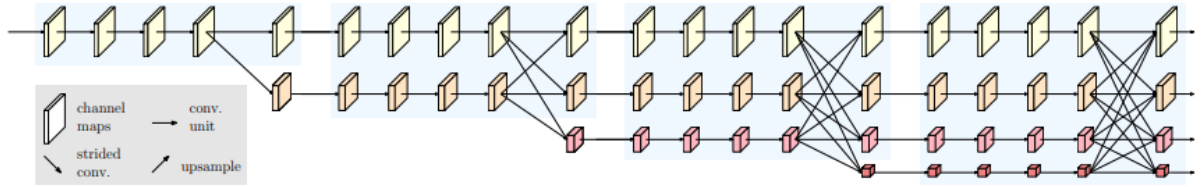
Mạng độ phân giải cao (HRNet) có khả năng duy trì các phản hồi có độ phân giải cao trong toàn bộ quá trình. Bắt đầu từ luồng tích chập có độ phân giải cao, dần dần thêm từng luồng có độ phân giải thấp và kết nối các luồng một cách song song. Do đó HRNet duy trì độ phân giải cao khi chuyển ảnh qua mạng, chứ không phục hồi độ phân giải như ý tưởng của top-down. Trước đây thường sử dụng phép Upsampling đặc trưng độ phân giải thấp lên cho bằng đặc trưng có độ phân giải cao và kết hợp lại bằng các phép cộng hoặc phép nối; HRNet sẽ thực hiện điều này nhiều lần, qua nhiều giai đoạn để các đặc trưng có độ phân giải cao, thấp sẽ bổ sung ngữ nghĩa cho nhau.



Hình 2.13. Cấu trúc phục hồi độ phân giải cao từ độ phân giải thấp.

2.6.2 Các tích chập song song đa độ phân giải

HRNet sẽ duy trì độ phân giải cao xuyên suốt quá trình ảnh truyền qua mạng. Kiến trúc sẽ chia thành nhiều tầng, tầng phía sau sẽ sinh ra đặc trưng có độ phân giải thấp và sâu hơn, tuy nhiên vẫn duy trì đặc trưng có độ phân giải ở tầng trước đó.



Hình 2.14. Minh họa mạng HRNet

Quan sát ở luồng số 1, qua nhiều tầng, đặc trưng độ phân giải cao (ký hiệu màu vàng) sẽ được duy trì. Bắt đầu từ tầng số 2, đặc trưng có độ phân giải thấp và sâu hơn (màu cam) sẽ được sinh ra, đặc trưng có độ phân giải cao màu vàng này sẽ được bổ sung ngữ nghĩa bởi các đặc trưng màu cam và ngược lại. Cứ như vậy qua nhiều tầng, các đặc trưng có các độ phân giải khác nhau sẽ được bổ sung ngữ nghĩa cho nhau, dẫn đến cuối cùng ta sẽ thu được đặc trưng độ phân giải cao, tuy nhiên chất lượng vẫn tốt.

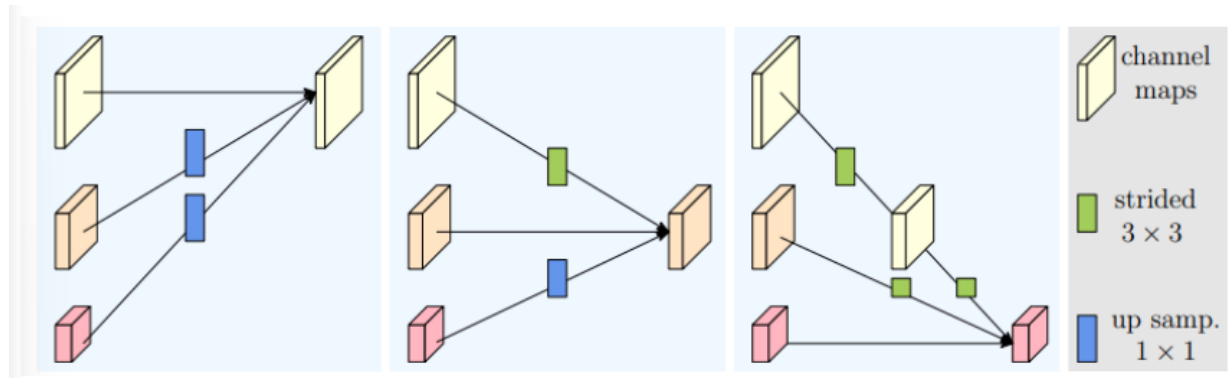
Hình 2.14 minh họa mạng HRNet bao gồm 4 giai đoạn. Luồng xử lý chính được mô tả một cách logic bằng công thức sau:

$$\begin{aligned}
 &\mathcal{N}_{11} \rightarrow \mathcal{N}_{21} \rightarrow \mathcal{N}_{31} \rightarrow \mathcal{N}_{41} \\
 &\quad \searrow \mathcal{N}_{22} \rightarrow \mathcal{N}_{32} \rightarrow \mathcal{N}_{42} \\
 &\quad \quad \searrow \mathcal{N}_{33} \rightarrow \mathcal{N}_{43} \\
 &\quad \quad \quad \searrow \mathcal{N}_{44}
 \end{aligned}$$

Trong đó, N_{st} là một nhánh đặc trưng tại giai đoạn thứ s , và r là chỉ mục độ phân giải. Độ phân giải ở giai đoạn đầu tiên sẽ là $r = 1$, độ phân giải thứ r sẽ bằng $\frac{1}{s^{r-1}}$ độ phân giải ở giai đoạn đầu tiên.

2.6.3 Các phép kết hợp đặc trưng

Việc kết hợp các đặc trưng là để trao đổi thông tin ngữ nghĩa giữa các đặc trưng có nhiều độ phân giải khác nhau. Nó được lặp đi, lặp lại nhiều lần. Ví dụ như ở hình 2.14, việc lặp đi lặp lại được thực hiện 4 lần.



Hình 2.15. Minh họa cách kết hợp thông tin ngữ nghĩa các đặc trưng độ phân giải cao, trung bình và thấp tương ứng từ trái sang phải.

Quan sát hình 2.15, bắt đầu từ luồng tích chập có độ phân giải cao làm giai đoạn đầu tiên, dần dần thêm từng luồng có độ phân giải cao xuống thấp, tạo thành các giai đoạn mới và kết nối song song các luồng đa độ phân giải. Do đó, độ phân giải cho các luồng song song của giai đoạn sau bao gồm độ phân giải từ giai đoạn trước và một mức thấp hơn.

2.6.4 Các dạng biểu diễn

HRNet có 3 dạng biểu diễn, được gọi là HRNetV1, HRNetV2, HRNetV2p

- HRNetV1: đầu ra là dạng biểu diễn chỉ lấy từ nhánh đặc trưng có độ phân giải cao (high-resolution). Ba đặc trưng của ba độ phân giải còn lại sẽ được lược bỏ.

- HRNetV2: các tác giả rescale các đặc trưng độ phân giải thấp lên độ phân giải cao bằng một phép upsampling song tuyến mà không thay đổi chiều sâu, và nối lại cả bốn đặc trưng. Sau đó đi qua một lớp tích chập $1 \times 1 \times 1$ để trộn cả bốn đặc trưng này lại.
- HRNetV2p: các tác giả xây dựng một biểu diễn đa tỷ lệ (multi-level features) bằng cách sử dụng đầu ra từ HRNetV2.

2.7 Phương pháp đánh giá kết quả

2.7.1 Confusion Matrix

Trong máy học, các mô hình phân loại được sử dụng để phân chia dữ liệu thành các danh mục khác nhau. Theo Jianfeng và các cộng sự [28], confusion matrix là phương pháp để tóm tắt hiệu suất của mô hình phân loại. Nó là một công cụ cần thiết để đánh giá độ chính xác của một mô hình và xác định mức độ phân loại dữ liệu, đặc biệt là các mô hình đánh giá rủi ro tín dụng khi mà các dữ liệu thường có xu hướng mất cân bằng rất lớn [29] [30] khiến cho kết quả của chỉ số đánh giá độ chính xác phân loại (classification accuracy) không còn đánh giá đúng chất lượng của mô hình huấn luyện [29].

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Hình 2.16. Confusion Matrix¹²

Confusion matrix có các thành phần sau:

¹² [Understanding Confusion Matrix | by Sarang Narkhede | Towards Data Science](#)

- TP: số lượng mẫu được mô hình phân loại đúng vào lớp dự đoán đúng.
- FP: số lượng mẫu được mô hình phân loại sai vào lớp dự đoán đúng.
- FN: số lượng mẫu được mô hình phân loại sai vào lớp dự đoán sai.
- TN: số lượng mẫu được mô hình phân loại đúng vào lớp dự đoán sai.

2.7.2 Độ chính xác (Accuracy)

Độ chính xác là thang đo đơn giản nhất giúp ta đánh giá hiệu quả dự báo của mô hình trên bộ dữ liệu. Bộ phân lớp của mô hình càng tốt sẽ cho kết quả độ chính xác càng cao. Độ chính xác được tính bằng tổng số trường hợp dự đoán đúng trên cả âm tính và dương tính chia cho tổng số mẫu.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (\text{PT 2.6})$$

2.7.3 Recall

Recall (TPR): Recall được định nghĩa là tỉ lệ số điểm true positive trong số những điểm thực sự là positive (TP + FN). Trong mô hình phân loại sẽ kỳ vọng chỉ số này lớn nhất.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{PT 2.7})$$

2.7.4 Precision

Precision: Precision được định nghĩa là tỉ lệ số điểm true positive trong số những điểm được phân loại là positive (TP + FP). Trong mô hình phân loại sẽ kỳ vọng chỉ số này lớn nhất.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (\text{PT 2.8})$$

2.7.5 F1 score

F1-score là trung bình điều hòa (harmonic mean) của precision và recall (giả sử hai đại lượng này khác 0). F1-score được tính theo công thức:

$$F1\ score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (PT\ 2.9)$$

2.8 Các phương pháp fine-tuning model

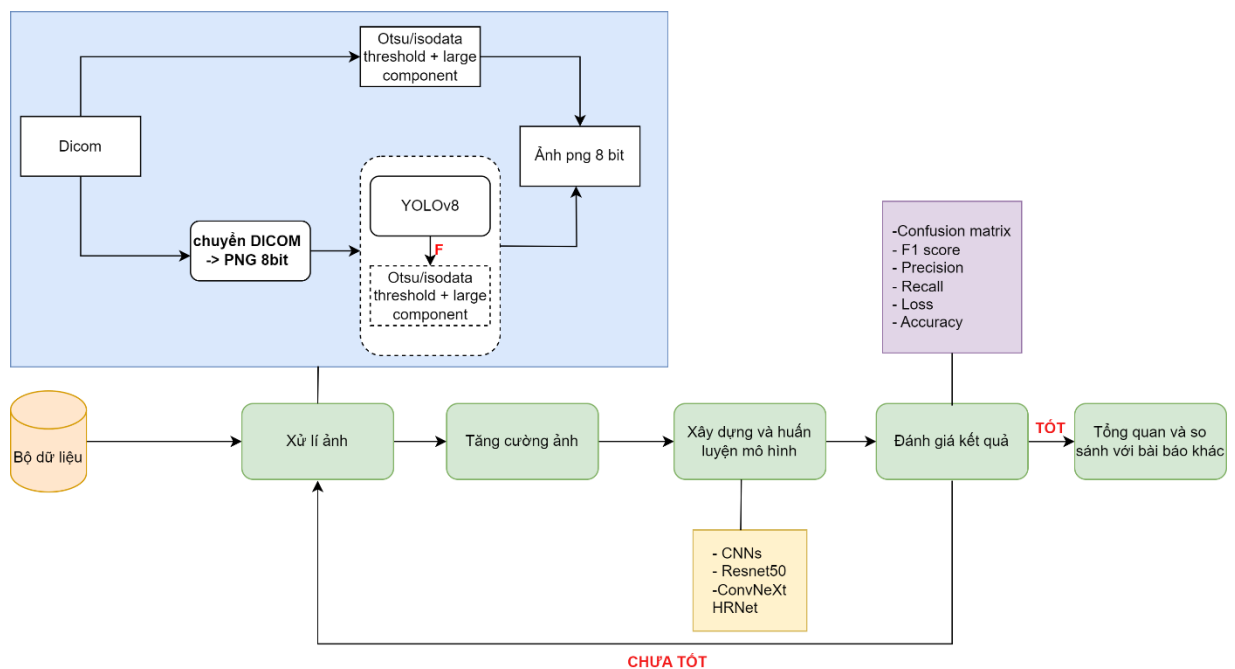
Fine-tuning là một kỹ thuật trong học sâu (deep learning) được sử dụng để tinh chỉnh một mô hình mạng nơ-ron (neural network model) đã được huấn luyện trước đó trên một tập dữ liệu khác hơn, thường là nhỏ hơn. Việc fine-tuning cho phép tái sử dụng các trọng số của mô hình đã được huấn luyện trước đó, từ đó giảm thiểu thời gian huấn luyện và chi phí tính toán. Dưới đây là các cách fine-tune:

- Freeze weight (đóng băng trọng số) một hoặc nhiều phần: Khi freeze weight của các lớp này đồng nghĩa với việc khi train thì trọng số của lớp này sẽ được giữ nguyên, chỉ cập nhật các lớp còn lại.
- Thay đổi kiến trúc: có thể là thêm layer hoặc chỉ lấy một phần layer.
- Sử dụng cả kiến trúc và trọng số để tiếp tục đào tạo.

CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT

3.1 Tổng quan hướng tiếp cận

Chúng tôi sẽ thực nghiệm bài toán phân loại ung thư vú với các mô hình như: CNN, ResNet50, ConvNeXt, HRNet, sau đó tiến hành tinh chỉnh tham số của những mô hình này để so sánh và rút ra được kết luận lựa chọn những tham số mô hình nào là hợp lý nhất. Để đạt được kết quả tốt, chúng tôi sẽ tiến hành cắt ROI - region of interest (đối với bài toán này là vùng nhũ) bằng YOLOv8, sử dụng pre-train và các phương pháp xử lý ảnh hiệu quả hiện nay.



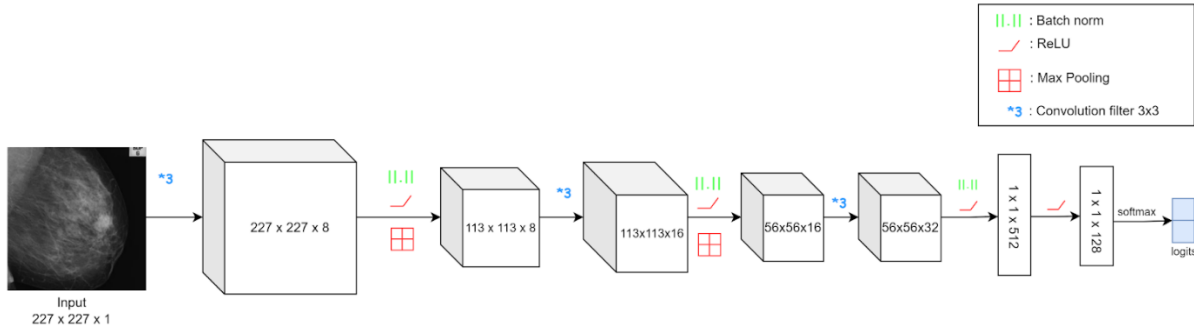
Hình 3.1. Sơ đồ quá trình giải quyết bài toán

3.2 Đặc trưng mô hình đề xuất

3.2.1 Mô hình CNNs

Trong luận án này, nhóm chúng tôi sử dụng kiến trúc CNNs được tham khảo từ [31] gồm ba lớp tích chập, hai lớp max pooling, 5 lớp kích hoạt ReLU, ba lớp chuẩn hóa và ba lớp fully connected. Mỗi layer chúng tôi đều sử dụng stride bằng 2, kích thước 2x2 cho mỗi

lớp max pooling. Với mỗi lớp convolution, chúng tôi đều sử dụng stride bằng 1, kích thước kernel 3x3 và có padding ‘same’ để trích xuất đặc trưng của ảnh. Theo sau đó là lớp fully connected với số nút lần lượt là 512, 128 sử dụng hàm kích hoạt ReLU. Cuối cùng là lớp softmax để trả về logits dự đoán các lớp. Kiến trúc này được rút ra từ thực nghiệm của nhóm tác giả, chi tiết được thể hiện dưới hình sau:



Hình 3.2. Kiến trúc CNNs sử dụng cho bài toán chẩn đoán nhũ ảnh.

3.2.2 Mô hình ResNet50

Trong khóa luận này, chúng tôi sử dụng kiến trúc mô hình ResNet50 được tham khảo từ [20] kiến trúc mô hình như đã được nêu ở chương 2. Ở lớp fully connected cuối cùng, chúng tôi đã sửa lại số lớp đầu ra bằng 2.

3.2.3 Mô hình ConvNeXt

Nhóm đã fine-tune sử dụng trọng số đã qua đào tạo của mô hình “convnext_small.fb_in22k_ft_in1k_384” sử dụng API của HuggingFace. Mô tả thông số mô hình:

Bảng 3.1. Thông số mô hình convnext_small.fb_in22k_ft_in1k_384

Thông số mô hình	50.2 triệu parameters
	GMACs: 45.2
	Size ảnh mặc định: 384x384

Dữ liệu fine-tune	ImageNet-1k
Dữ liệu pretrained	ImageNet-22k

Vì mô hình sử dụng cho bài toán phân loại 1000 lớp, nên để sử dụng mô hình cho bài toán phân loại nhị phân, nhóm chúng tôi đã sửa đổi lớp Linear ở cuối với đầu ra bằng 2.

3.2.4 Mô hình HRNet

Đối với bài toán phân loại ảnh ung thư, chúng tôi sử dụng kiến trúc HRNet được tham khảo ở [32]. Ở lớp fully connected, chúng tôi tiến hành thay đổi đầu ra (output features) bằng 2 để phù hợp với bài toán.

3.3 Hàm mất mát Cross Entropy

Cross-entropy là một khái niệm quan trọng trong lĩnh vực học máy và thống kê, đặc biệt trong bài toán phân loại. Nó được sử dụng để đo lường sự khác biệt giữa hai phân phối xác suất. Trong bài toán phân loại, cross-entropy thường được sử dụng để đo lường hiệu suất của mô hình phân loại bằng cách so sánh giữa các xác suất dự đoán của mô hình với các nhãn thực tế của dữ liệu. Cross-entropy loss được sử dụng khi điều chỉnh trọng số của mô hình trong quá trình đào tạo. Mục đích là để giảm thiểu tổn thất, nghĩa là tổn thất càng nhỏ thì mô hình càng tốt. Mô hình hoàn hảo có Cross-entropy loss bằng 0.

Theo Saurav Maheshkar [33] Cross-Entropy Loss có công thức như sau:

$$L_{CE} = - \sum_{i=1}^n t_i \log(p_i) \quad (PT\ 3.1)$$

Trong đó:

- t_i là xác suất thực tế (nhãn) của mẫu thuộc vào lớp i
- p_i là xác suất dự đoán của mô hình cho mẫu thuộc vào lớp i
- n : Số lớp

Đối với bài toán phân loại nhị phân, công thức trên có thể được rút gọn thành:

$$L_{CE} = -y \log(p) + (1 - y)\log(1 - p) \quad (PT\ 3.2)$$

Trong đó:

- y là nhãn thực tế của mẫu (0 hoặc 1)
- p là xác suất dự đoán của mô hình

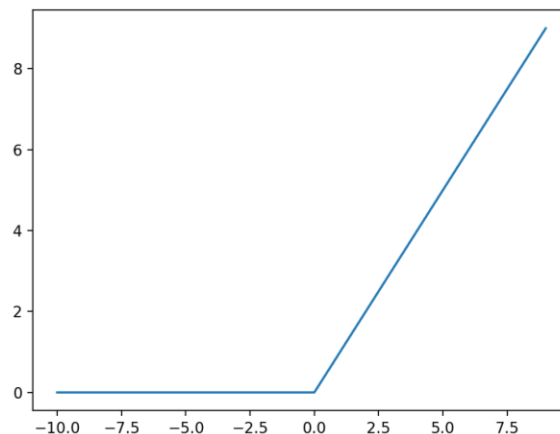
Cross-Entropy là một hàm mất mát (loss function) phổ biến trong việc huấn luyện mô hình phân loại trong học máy. Các thuật toán tối ưu hóa được sử dụng để điều chỉnh các tham số của mô hình (như gradient descent) thường dựa trên việc tính toán đạo hàm của cross-entropy để cập nhật trọng số của mô hình.

3.4 Hàm kích hoạt

3.4.1 ReLU

Hàm ReLU đang được sử dụng dụng khá nhiều trong những năm gần đây khi huấn luyện các mạng nơ ron. ReLU là một hàm tuyến tính từng phần sẽ xuất đầu vào trực tiếp nếu nó dương, nếu không nó sẽ xuất ra số 0. Nó đã trở thành chức năng kích hoạt mặc định cho nhiều mạng thần kinh vì khi sử dụng nó, mô hình sẽ đạt hiệu suất tốt hơn và dễ đào tạo hơn. Công thức của hàm kích hoạt ReLU:

$$f(x) = \max(0, x) \quad (PT\ 3.3)$$



Hình 3.3. Hàm kích hoạt ReLU

3.4.2 Softmax

Hàm softmax tính toán xác suất xảy ra của một sự kiện. Nói một cách khái quát, hàm softmax sẽ tính khả năng xuất hiện của một class trong tổng số tất cả các class có thể xuất hiện. Sau đó, xác suất này sẽ được sử dụng để xác định class mục tiêu cho các input.

Cụ thể, hàm softmax biến vector k chiều có các giá trị thực bất kỳ thành vector k chiều có giá trị thực có tổng bằng 1. Giá trị nhập có thể dương, âm, bằng 0 hoặc lớn hơn 1, nhưng hàm softmax sẽ luôn biến chúng thành một giá trị nằm trong khoảng $(0;1]$.

Như vậy, chúng có thể được gọi là “xác suất”. Nếu một trong các giá trị nhập rất nhỏ hoặc âm, hàm softmax biến chúng thành 1 xác suất nhỏ. Còn nếu một giá trị nhập lớn thì nó sẽ được chuyển thành một xác suất lớn. Nhưng xác suất luôn lớn hơn 0 và nhỏ hơn 1, hoặc bằng 1.

Công thức của hàm softmax được tính như sau:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (PT\ 3.4)$$

Trong đó:

- z : Là giá trị vector nhập vào cho hàm Softmax (từ z_0 đến z_n)
- z_i : Giá trị vector thứ i nhập vào cho hàm Softmax
- e^{z_i} : Hàm lũy thừa tiêu chuẩn được áp dụng cho mỗi giá trị nhập. Nó sẽ đưa ra một giá trị tương đương lớn hơn 0. Giá trị này sẽ rất nhỏ nếu giá trị nhập là âm, và rất lớn nếu giá trị nhập là dương. Tuy nhiên nó không cố định trong khoảng $(0,1]$. Đây là yêu cầu của một xác suất.
- $\sum_{j=1}^n e^{z_j}$: Đảm bảo rằng tổng của các giá trị đầu ra sẽ luôn bằng 1 và nằm trong khoảng $(0,1]$. Như vậy, sẽ xuất hiện phân phối xác suất chính xác.
- $\sum_{j=1}^n e^{z_j}$: Số class trong một phân loại nhiều class

3.5 Các thuật toán tối ưu

Thuật toán tối ưu (optimizer) là cơ sở để xây dựng mô hình Neural Network với mục đích “học” được các features (hay pattern) của dữ liệu đầu vào, từ đó có thể tìm 1 cặp trọng số và bias phù hợp để tối ưu hóa model. Vì vậy chúng ta cần tìm một thuật toán để cải thiện trọng số và bias theo từng bước, đó là lý do các thuật toán optimizer ra đời.

3.5.1 Stochastic Gradient Descent (SGD)

SGD là một biến thể của Gradient Descent (GD) [34]. Thay vì mỗi epoch chúng ta sẽ cập nhật trọng số (weight) 1 lần thì mỗi epoch có n điểm dữ liệu chúng ta sẽ cập nhật trọng số n lần. Nhìn vào 1 mặt, SGD sẽ làm giảm đi tốc độ của 1 epoch. Công thức SGD cũng tương tự như GD nhưng thực hiện trên từng điểm dữ liệu. Trong thuật toán này, tại một thời điểm, ta chỉ tính đạo hàm mất mát dựa trên chỉ một điểm dữ liệu x_i rồi cập nhật θ dựa trên đạo hàm này. Việc này được thực hiện với từng điểm trên toàn bộ dữ liệu, sau đó lặp lại quá trình trên.

Một lưu ý là: sau mỗi epoch, chúng ta cần shuffle (xáo trộn) thứ tự của các dữ liệu để đảm bảo tính ngẫu nhiên. Việc này cũng ảnh hưởng tới hiệu năng của SGD. Một cách toán học, quy tắc cập nhật SGD là:

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x_i; y_i) \quad (PT\ 3.5)$$

Trong đó $J(\theta; x_i; y_i)$ là hàm mất mát với chỉ một cặp điểm dữ liệu (input, label) là $(x_i; y_i)$

- Ưu điểm: Giải quyết được đối với cơ sở dữ liệu lớn mà GD không làm được.
- Nhược điểm: chưa giải quyết được 2 nhược điểm lớn của Gradient Descent: phụ thuộc vào điểm khởi tạo ban đầu và learning rate.

3.5.2 Adam

Adam là một thuật toán tối ưu được giới thiệu lần đầu bởi Kingma & Ba vào năm 2014, tại International Conference on Learning Representations [35]. Adam là sự kết hợp của Momentum [34] và RMSprop [34]. Theo như bài báo giới thiệu, Adam có khá nhiều điểm mạnh, đặc biệt là dễ sử dụng, tính toán hiệu quả, và phù hợp với các dữ liệu lớn và rời rạc. Đây cũng là thuật toán thường được sử dụng trong lĩnh vực xử lý ảnh và xử lý ngôn ngữ tự nhiên.

Khác với SGD, một phương pháp để tối ưu hàm mục tiêu trong học sâu, Adam áp dụng các learning rate khác nhau cho mỗi tham số dựa vào các tham số β_1 (first momentum of gradient) và β_2 (second momentum of gradient). Theo như tác giả, Adam có ưu điểm của hai thuật toán tối ưu phổ biến khác là Adaptive gradient algorithm và Root mean square propagation. Việc sử dụng Adam từ thư viện của các framework khá là tiện lợi, chúng ta chỉ cần lưu ý đến bốn tham số chính như sau:

- α : hay còn được biết đến là learning rate; tham số này phản ánh tỷ lệ cập nhật của các trọng số trong quá trình huấn luyện; giá trị của tham số càng lớn, thì tốc độ học ban đầu càng nhanh.
- β_1 : tỉ lệ giảm dần theo cấp số nhân cho the first moment estimates.
- β_2 : tỉ lệ giảm dần theo cấp số nhân cho the second moment estimates.
- ϵ : một giá trị lớn hơn 0 và rất nhỏ nhằm tránh lỗi chia cho 0 trong quá trình thực thi.

3.5.3 AdamW

AdamW là một cải tiến của thuật toán tối ưu Adam, AdamW thêm một hệ số trọng lượng giải quyết vấn đề liên quan đến việc điều chỉnh các trọng số trên các tham số thưa (sparse parameters) trong mô hình. Việc thêm hệ số trọng lượng giúp regularization mạnh mẽ hơn và ngăn chặn các trọng số không cần thiết tăng lên quá cao, từ đó giảm khả năng overfitting.

3.6 Các phương pháp tiền xử lý dữ liệu

Bảng 3.2. Bảng tổng quan số lượng mẫu ảnh đã sử dụng trong bài toán này

Tên bộ dữ liệu	Số lượng bệnh nhân	Số lượng ảnh từng bộ	Tổng số lượng ảnh là ung thư ác tính
INbreast	115	410	343 (83.66%)
MIAS		322	51 (15,8%)
RSNA	11913	54706	1158 (2.1%)
DDSM	2500	7808	2716 (34.78%)
CMMD	1775	5202	4094 (78.7%)
Tổng		68448	8362

3.6.1.1 Chuyển ảnh DICOM sang ảnh PNG

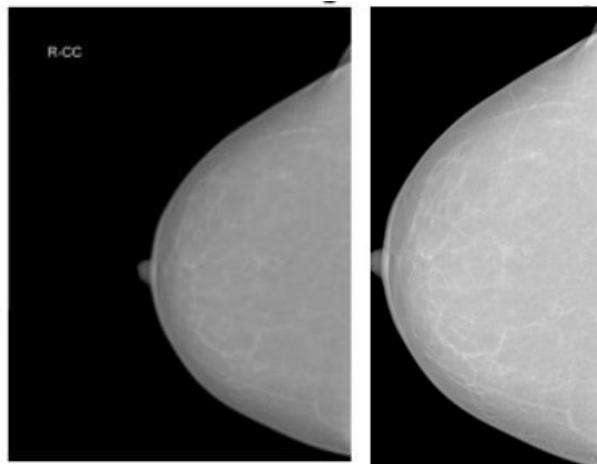
Ảnh Digital Imaging and Communications in Medicine - Dicom là một dạng ảnh tiêu chuẩn sử dụng trong lĩnh vực y tế để trao đổi và lưu trữ các hình ảnh CT, MRI, siêu âm và X-quang. Bên cạnh đó, ảnh này còn chứa các thông tin chi tiết của bệnh nhân nên dung lượng tệp rất lớn. Chính vì vậy để giảm dung lượng và dễ dàng xử lý chúng em đã chuyển sang định dạng png. Công thức chuyển đổi hình ảnh:

$$image_{new} = \frac{image - \min(image)}{\max(image) - \min(image)} * 255 \quad (PT\ 3.6)$$

Bên cạnh đó, dữ liệu dicom có 2 dạng MONOCHROME-1 và MONOCHROME-2 nó dùng để xác định cách hiển thị các pixel dưới dạng đơn sắc. Hai thông số này được lưu vào PhotometricInterpretation. Với MONOCHROME-1 nghĩa là các pixel cao được hiển thị dưới dạng đen hơn, ngược lại với MONOCHROME-2 các pixel cao được hiển thị trắng hơn.

3.6.1.2 Cắt ảnh

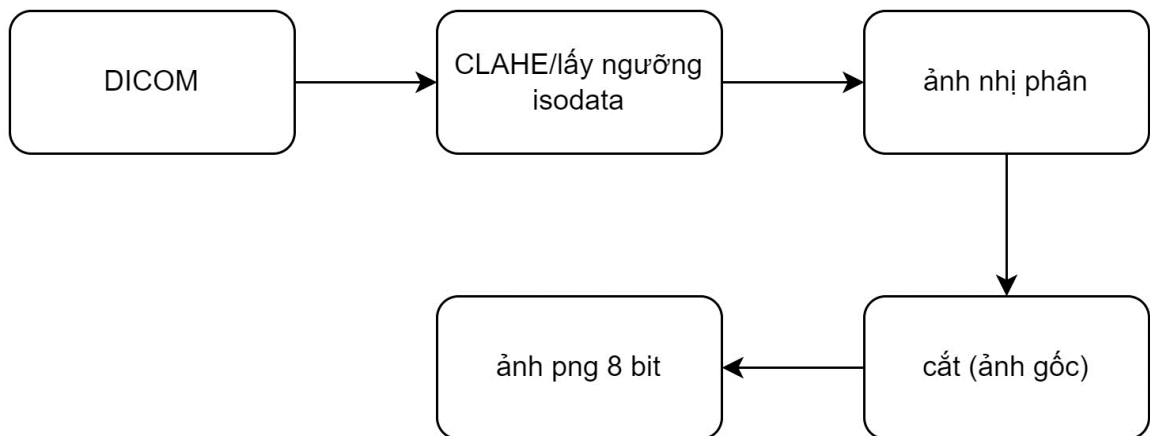
Mục đích của việc cắt ảnh là để loại bỏ phần ảnh đen dư thừa và giảm bớt nhiễu ảnh, để mô hình chỉ tập trung học phần nhũ ảnh. Hơn nữa, ảnh y khoa thường nặng và lớn, việc giảm bớt kích thước ảnh cũng sẽ tiết kiệm được tài nguyên tính toán và đạt được kết quả nhanh hơn. Trong quá trình giải quyết bài toán, giai đoạn đầu nhóm chúng tôi sử dụng ảnh cắt bằng liên thông lớn nhất, thời gian còn lại sử dụng ảnh cắt sử dụng YOLOv8.



Hình 3.4. Nhũ ảnh trước và sau khi cắt

3.6.1.3 Cắt ảnh sử dụng liên thông lớn nhất

Đầu tiên sử dụng cắt ngưỡng isodata để cắt ngưỡng phân vùng nhũ ảnh. Sau đó chuyển đổi ảnh từ ảnh xám thành ảnh nhị phân. Khi đó, ảnh sẽ chia thành 2 phần, một phần là các đối tượng trong ảnh mang giá trị 255 và phần còn lại là phần nền mang giá trị 0. Sau đó ta sử dụng ý tưởng của thuật toán loang để tìm các thành phần liên thông. Phần liên thông có diện tích lớn nhất sẽ là phần nhũ ảnh.



Hình 3.5. Sơ đồ cắt ảnh sử dụng phương pháp tìm liên thông lớn nhất

- Nhược điểm:
 - Phụ thuộc vào việc lấy ngưỡng ảnh, ngưỡng ảnh không chuẩn sẽ không cắt được như ý muốn.
 - Ảnh nhiễu nhiễu sẽ bị cắt sai hoặc không cắt được.

Trong quá trình cắt, các bộ ít nhiễu ảnh là INbreast, Mias, RSNA, tỉ lệ cắt đúng lên đến 90%, tuy nhiên bộ DDSM tỉ lệ cắt đúng chỉ 80%.

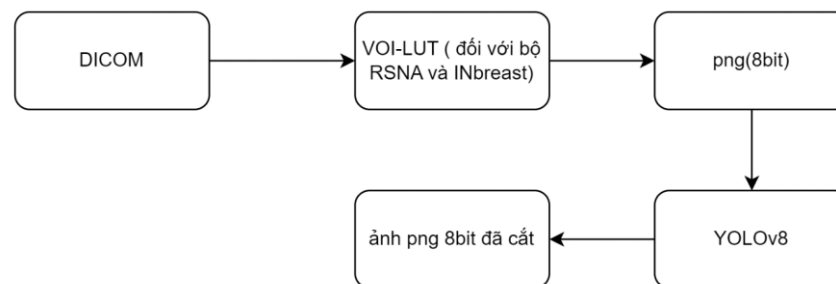
3.6.1.4 Cắt ảnh sử dụng YOLOv8

YOLOv8 là một trong những mô hình giải quyết bài toán phát hiện đối tượng (detect objection) tốt nhất hiện tại, được phát triển từ YOLOv5, tuy nhiên, có một vài điểm thay đổi so với mô hình YOLOv5 như sau:

- Thay thế module C3 bằng module C2f.
- Thay thế lớp tích chập đầu tiên trong kiến trúc backbone 6 x 6 Conv bằng 3 x 3 Conv.
- Xóa 2 Conv (No.10 và No. 14 trong YOLOv5).
- Tại Bottleneck, thay thế 1 x 1 Conv thành 3 x 3 Conv.
- Sử dụng “decouple head” thay vì “objectness”.

Hiện nay có 5 phiên bản YOLOv8 là YOLOv8-nano, YOLOv8-small, YOLOv8-medium, YOLOv8-Large và YOLOv8-Xtra Large. Trong đó, YOLOv8-nano nhanh và nhỏ gọn nhất, YOLOv8-XL cho kết quả độ chính xác cao nhất. YOLOv8 sử dụng trong bài này là YOLOv8-nano.

Vì phần này nhóm đi theo hướng sử dụng chứ không đi sâu vào kiến trúc, chi tiết thêm tham khảo ở [36] [37].



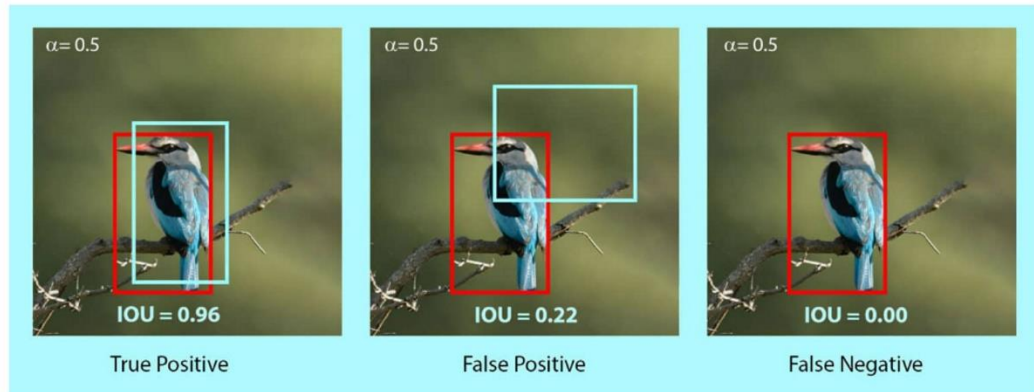
Hình 3.6. Sơ đồ quá trình cắt ảnh sử dụng YOLOv8

Để sử dụng mô hình, nhóm em sử dụng 1000 ảnh lấy từ 4 bộ dữ liệu (Mias, INbreast, RSNA và mini-DDSM) mỗi bộ lấy số lượng ảnh trái phải với tỉ lệ MLO, CC bằng nhau. Trong đó có 810 ảnh để train, 100 ảnh test và 90 ảnh sử dụng để validation. Chi tiết dưới bảng sau:

Bảng 3.3. Dữ liệu huấn luyện mô hình YOLOv8

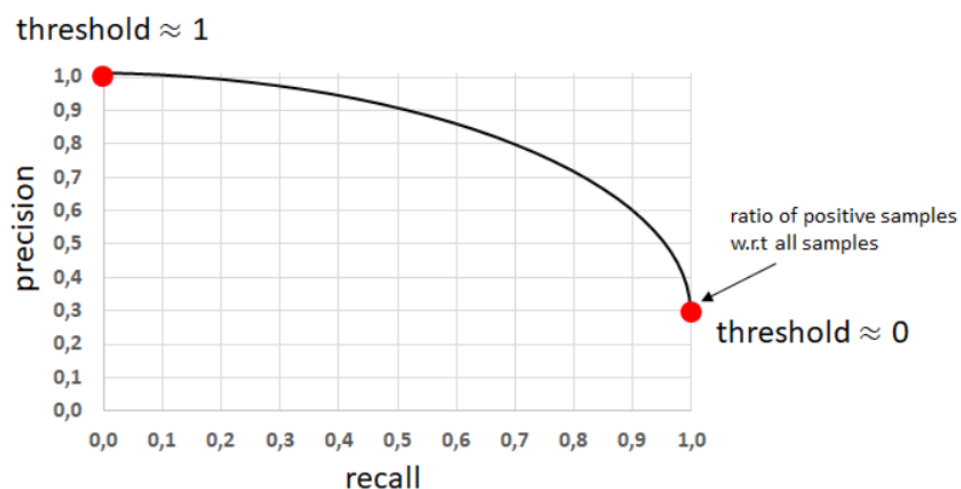
Tên bộ dữ liệu	Số ảnh	Tỉ lệ %
INbreast	57	7.04%
Mias	89	10.99%
Mini-DDSM	117	21.85%
RSNA	487	60.12%

Trong phát hiện đối tượng, tính chính xác của dự đoán (TP, FP hoặc FN) được quyết định với sự trợ giúp của ngưỡng IoU (Intersection Over Union) - Phần giao / phần hợp.



Hình 3.7. Minh họa về tính chính xác của dự đoán với sự trợ giúp của ngưỡng đánh giá IoU¹³

Nếu IoU lớn hơn threshold đó thì được coi là True Positive (nhận chuẩn) và ngược lại nếu $<$ threshold thì sẽ coi là False Positive (nhận là Positive nhưng bị sai). Từ đó, chúng ta thay đổi IoU threshold, tính toán ra Precision và Recall tại các mức đó và vẽ ra được Precision Recall Curve như sau:



Hình 3.8. Precision Recall Curve¹⁴

¹³ [Intersection Over Union IoU in Object Detection Segmentation \(learnopencv.com\)](https://learnopencv.com/intersection-over-union-iou-in-object-detection-segmentation/)

¹⁴ [Thủ tìm hiểu về mAP - đo lường Object Detection model - Mì AI \(mimai.vn\)](https://mimai.vn/thu-tim-hieu-ve-mAP-do-luong-object-detection-model/)

Tiếp theo ta sẽ tìm hiểu về độ đo đánh giá AP – Average Precision. AP chính là vùng diện tích nằm dưới đường cong Precision Recall nói trên.

AP lớn nếu vùng này lớn, suy ra đường cong có xu hướng gần góc trên bên phải và có nghĩa là tại các threshold khác nhau thì Precision và Recall đều khá cao. Từ đó suy ra model tốt.

AP nhỏ thì cả Precision và Recall đều khá thấp và model không tốt. mAP là Mean Average Precision là trung bình cộng giá trị AP của các class khác nhau. Vì trong bài số lớp của nhóm chỉ có 1 nên mAP chính là AP. Độ chính xác mAP khi nhận diện nhũ ảnh của nhóm tới 99.5%.

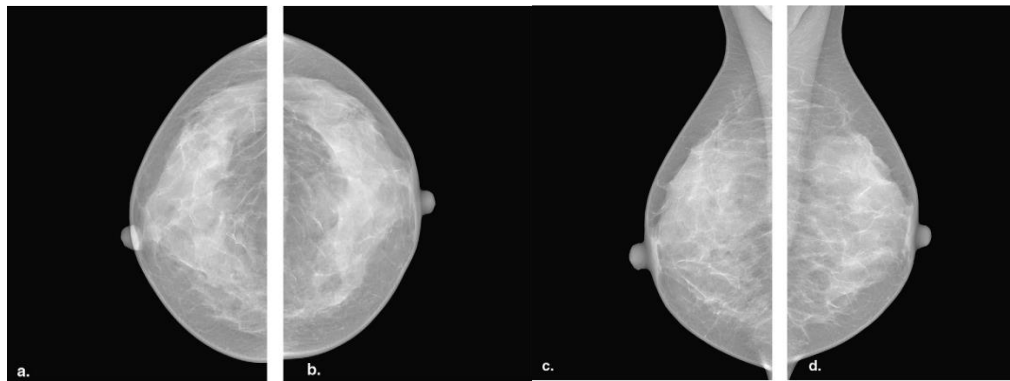
CHƯƠNG 4. THỰC NGHIỆM VÀ KẾT QUẢ

4.1 Dữ liệu

4.1.1 Tập dữ liệu *INbreast*

Bộ dữ liệu được mua tại Breast Centre của CHSJ, Porto, dưới sự cho phép của Hội đồng Khoa học và Đạo đức và Ủy ban Quốc gia về Bảo vệ Dữ liệu. Các hình ảnh được chụp từ tháng 04/2008 đến tháng 07/2010, với thiết bị chụp là MammoNovation Siemens FFDM, sử dụng chất truyền dẫn như Selen vô định để thu nhận và chuyển đổi trực tiếp các photon của tia X thành tín hiệu điện nạp (detector of amorphous selenium), kích thước pixel là 70 microns, với độ phân giải tương phản là 14-bit. Ma trận hình ảnh có kích thước 3328 x 4084 hoặc 2560 x 3328 pixels tùy thuộc tấm nén được mua (theo kích thước ngực của bệnh nhân). Nhữ ảnh được lưu dưới định dạng DICOM. Tất cả các thông tin bảo mật y tế đều đã được xóa khỏi file DICOM, theo Bổ sung 55 của chuẩn DICOM [38].

Bộ dữ liệu INbreast có tổng cộng 115 trường (410 ảnh), trong đó có 90 trường hợp được thu thập cả 2 vú (mỗi trường hợp bao gồm 4 ảnh) và 25 trường hợp là từ bệnh nhân cắt bỏ tuyến vú (gồm 2 ảnh cho mỗi trường hợp). Thông tin bao gồm mô tả chi tiết tổn thương (masses), vôi hóa (calcifications), không đối xứng (asymmetries) và biến dạng (distortions). Các đường viền xác định tổn thương chính xác do các chuyên gia thực hiện cũng được cung cấp ở định dạng XML [39].



Hình 4.1. Nhũ ảnh trong bộ dữ liệu INbreast: a) Góc chụp CC vú phải, b) Góc chụp CC vú trái, c) Góc chụp MLO vú phải, d) Góc chụp MLO vú trái

4.1.2 Tập dữ liệu *Mammographic Image Analysis Society (MIAS)*

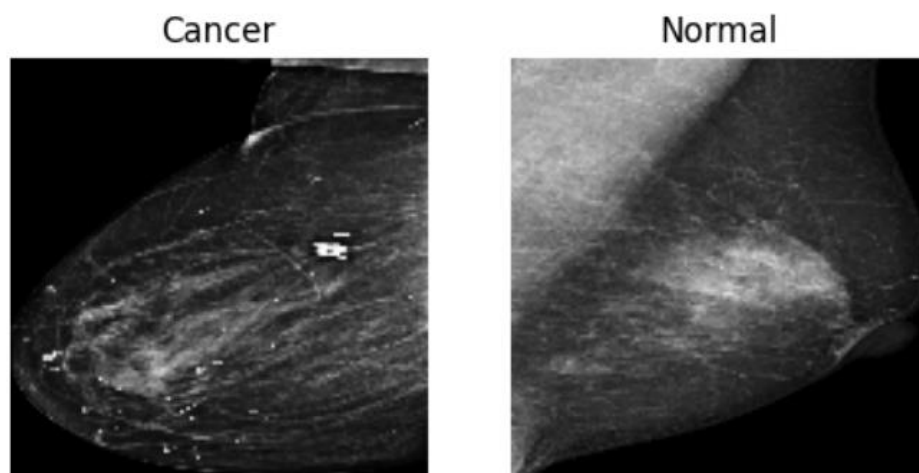
Hiệp hội phân tích hình ảnh chụp quang tuyến vú (MIAS) là một tổ chức gồm các nhóm nghiên cứu của Vương quốc Anh quan tâm đến việc hiểu về chụp quang tuyến vú và đã tạo ra một cơ sở dữ liệu về chụp quang tuyến vú kỹ thuật số.

Phim lấy từ chương trình Khám nghiệm vú Quốc gia của Vương quốc Anh đã được số hóa thành ảnh kỹ thuật số kích thước cạnh pixel là 50 micron. Tập dữ liệu chứa 322 phim đã được số hóa. Bộ dữ liệu cũng bao gồm các đánh dấu chuẩn xác của bác sĩ về các vị trí của các bất thường có thể có. Bộ dữ liệu đã được giảm xuống có kích thước cạnh 200 micron và được thêm/bớt viền để tất cả hình ảnh đều có kích thước 1024 x 1024. Trong đó, có 207 ảnh mang nhãn bình thường (normal), 64 ảnh mang nhãn ung thư lành tính (benign) và 51 ảnh mang nhãn ung thư ác tính (malignant) [40].

4.1.3 Tập dữ liệu *RSNA Screening Mammography Breast Cancer Detection*

RSNA (Radiological Society of North America) là một tập dữ liệu liên quan trọng trong lĩnh vực chẩn đoán ung thư vú bằng hình ảnh mammography. Được tổ chức và phát triển bởi Hội Nghị Y Học Hình Ảnh Bắc Mỹ (Radiological Society of North America), tập dữ liệu này nhằm hỗ trợ nghiên cứu và phát triển các thuật toán và hệ thống hỗ trợ chẩn đoán ung thư vú từ hình ảnh mammography.

Tập dữ liệu này bao gồm một lượng lớn các hình ảnh mammography từ hàng ngàn phụ nữ tham gia vào các chương trình sàng lọc ung thư vú. Đây là các hình ảnh chụp màng phim hoặc hình ảnh kỹ thuật số (Digital mammography) của vùng vú, nhằm phát hiện các dấu hiệu và khối u có thể là ung thư vú. Tập dữ liệu chứa thông tin về các bệnh nhân, kết quả chẩn đoán, các khối u đã được xác định và các dữ liệu khác liên quan đến quá trình sàng lọc ung thư vú. Hình ảnh được lưu dưới dạng dicom, có 53540 ảnh không ung thư và 1158 ảnh ung thư.



Hình 4.2. Hình ảnh minh họa dữ liệu trong tập RSNA

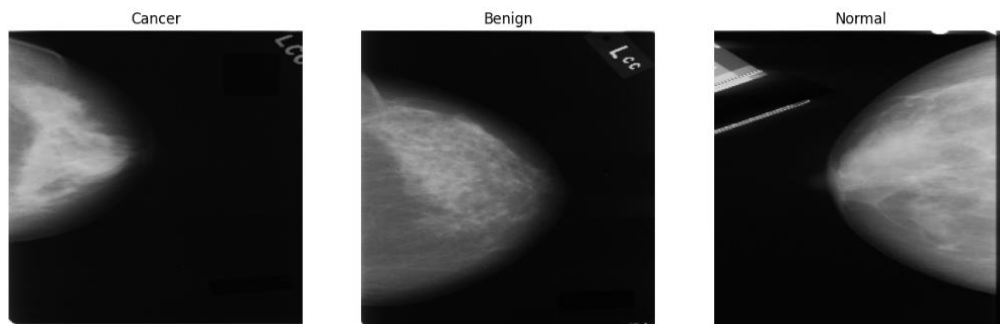
4.1.4 Tập dữ liệu *Digital Database for Screening Mammography (DDSM)*

The Digital Database for Screening Mammography (DDSM) là một bộ dữ liệu được sử dụng bởi cộng đồng nghiên cứu phân tích hình ảnh chụp nhũ ảnh. Hỗ trợ chính cho tập dữ liệu này là một khoản trợ cấp từ chương trình nghiên cứu Ung thư Vú của Bộ Tư lệnh Nghiên cứu Y tế và Vật tư Quân đội Hoa Kỳ.

Mục đích của tập dữ liệu là tạo điều kiện thuận lợi cho nghiên cứu hợp lý trong việc phát triển các thuật toán máy tính để hỗ trợ trong việc chẩn đoán bệnh. Các mục đích phụ của cơ sở dữ liệu có thể bao gồm phát triển các công cụ hỗ trợ giảng dạy hoặc đào tạo. Cơ sở dữ liệu chứa khoảng 2500 bệnh nhân. Mỗi nghiên cứu bao gồm hai hình ảnh của mỗi

bên vú, cùng với một số thông tin liên quan đến bệnh nhân (tuổi tại thời điểm nghiên cứu, mật độ vú ACR,...) và thông tin hình ảnh (máy quét, độ phân giải không gian,...).

Dữ liệu trong DDSM được thu thập từ nhiều hệ thống hình ảnh y tế, bao gồm cả các kỹ thuật chụp phim và kỹ thuật chụp kỹ thuật số. Nó bao gồm các hình ảnh phim và các hình ảnh số, cả hai được chụp với các thư thể và góc chụp khác nhau. DDSM cũng cung cấp các thông tin metadata, bao gồm kết quả kiểm tra và mô tả của bác sĩ.



Hình 4.3. Hình ảnh mô phỏng của tập dữ liệu DDSM.

Hình 26 mô phỏng của tập dữ liệu DDSM, bên trái là hình ảnh vú ung thư ác tính, giữa là hình ảnh ung thư lành tính, hình bên phải là ảnh không ung thư. Tập dữ liệu có chứa 2716 ảnh ung thư ác tính, 2684 ảnh lành tính và 2408 ảnh bình thường.

4.1.5 Tập dữ liệu *The Chinese Mammography Database (CMMD)*

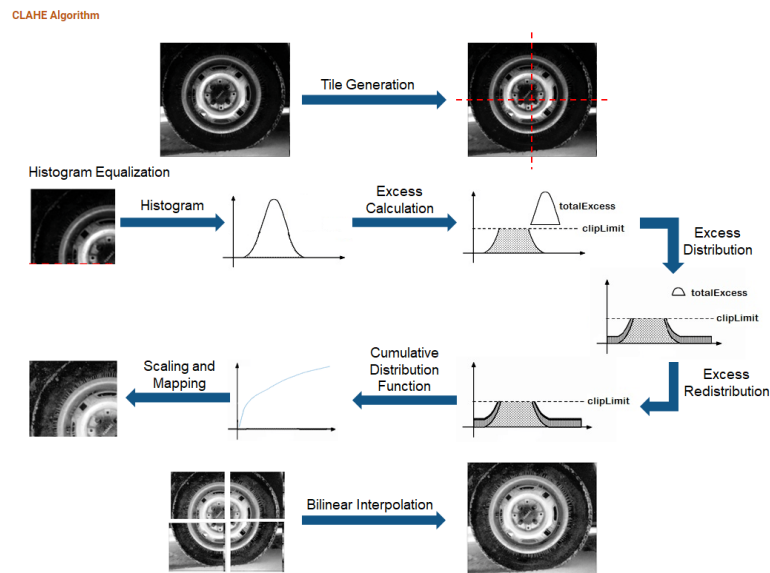
Bộ dữ liệu được thu thập từ 1775 bệnh nhân đến từ Trung Quốc mắc bệnh u vú (bao gồm cả lành tính và ác tính) được thu thập từ tháng 7 năm 2012 đến tháng 1 năm 2016. Hình ảnh chụp nhũ ảnh được thu thập ở định dạng .TIFF và được chuyển đổi sang DICOM 8 bit, bao gồm 5202 ảnh trong đó có 4094 là ung thư ác tính và 1108 ảnh là ung thư lành tính.

4.1.6 Các kỹ thuật xử lý ảnh

4.1.6.1 Tăng cường độ tương phản ảnh (CLAHE)

Các thuật toán tăng cường độ tương phản của ảnh đã phát triển trong vài thập kỷ qua để đáp ứng cho các bài toán xử lý ảnh. Có hai mục tiêu chính trong việc nâng cao độ tương phản ảnh. Thứ nhất, là cải thiện độ sắc nét của hình ảnh để dễ dàng trực quan. Thứ hai là hỗ trợ phân tích hình ảnh, phát hiện đối tượng và phân đoạn hình ảnh một cách dễ dàng hơn.

CLAHE - Contrast limited adaptive histogram equalization là một phương pháp cân bằng hình ảnh, nó là một biến thể của phương pháp Adaptive Histogram Equalization (AHE) [41] nhằm ngăn hiện tượng khuếch đại quá mức độ tương phản.

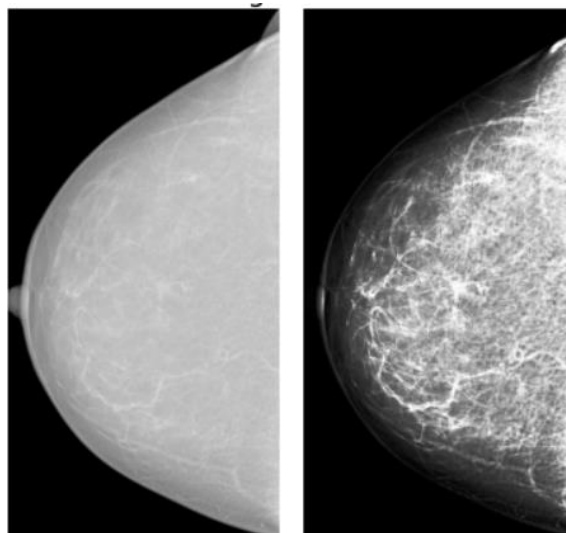


Hình 4.4. Hình ảnh minh họa sơ đồ quá trình CLAHE [42]

Thuật toán CLAHE gồm có 3 bước chính [42]:

- Tile Generation: ảnh đầu vào được chia thành các phần nhỏ gọi là tiles hoặc ô. Trong ví dụ này, ảnh đầu vào được chia thành bốn tiles.

- Histogram Equalization: Tiến hành việc cân bằng histogram cho từng tile sử dụng một ngưỡng giới hạn (clip limit) được xác định trước. Cân bằng histogram gồm năm bước sau:
 - Tính toán histogram: tính toán các giá trị bình của các ô trong histogram cho từng tile.
 - Tính toán giá trị vượt mức: tính toán giá trị vượt mức bằng cách tính tổng các giá trị vượt ngưỡng giới hạn.
 - Phân phối giá trị vượt mức: chia các giá trị vượt mức vào các ô khác.
 - Tái phân phối giá trị vượt mức: tái phân phối các giá trị đã phân phối vào các ô khác.
 - Scale và ánh xạ sử dụng hàm phân phối tích lũy (CDF): tính toán hàm phân phối tích lũy (CDF) cho các giá trị trong histogram. Giá trị CDF của mỗi tile được tỷ lệ và ánh xạ sử dụng các giá trị pixel của ảnh đầu vào.
- Bilinear Interpolation: các tiles sau khi qua các bước cân bằng histogram được nối lại với nhau bằng phương pháp nội suy tuyến tính (bilinear interpolation) để tạo ra một ảnh đầu ra với độ tương phản tốt hơn.



Hình 4.5. Ảnh minh họa ảnh nguyên bản (trước) và sau khi CLAHE (phải)

Thư viện skimage hỗ trợ rất tốt cho việc sử dụng CLAHE, dưới đây là các thông số cần chú ý:

- Giới hạn cắt (clip_limit): giới hạn cắt xác định giới hạn tối đa của độ sáng được phân bố lại trong mỗi ô. Nếu giá trị độ sáng vượt quá giới hạn này, nó sẽ bị cắt giảm để tránh việc tăng cường nhiều.
- Số lượng bin (n_bins): đây là số lượng khoảng giá trị độ sáng trong histogram được sử dụng trong quá trình cân bằng lược đồ. Số lượng bin càng lớn, cân bằng lược đồ càng chính xác, nhưng cũng tăng độ phức tạp tính toán.

4.1.6.2 Các kỹ thuật tăng cường ảnh

Trong nghiên cứu này, chúng tôi sử dụng một số các kỹ thuật liên quan đến việc xoay, lật ảnh của Pytorch. Việc xoay, lật ảnh có nhiều ứng dụng trong việc xử lý ảnh. Bao gồm:

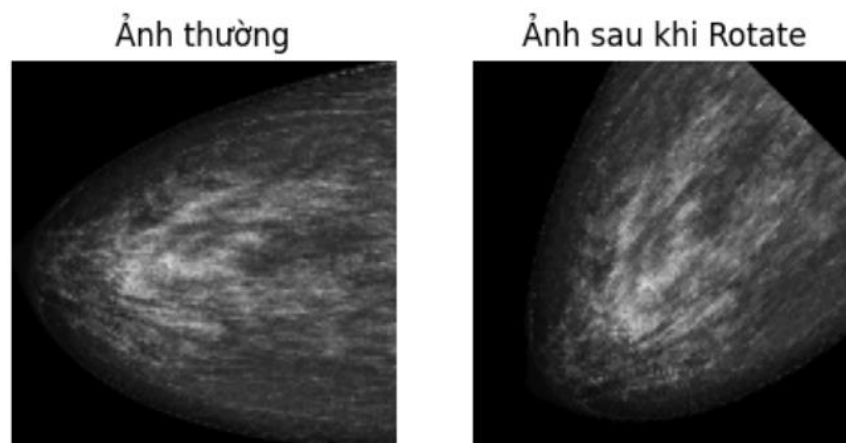
- Tăng cường dữ liệu (Data augmentation): xoay hình ảnh một góc ngẫu nhiên có thể giúp tăng cường dữ liệu huấn luyện. Điều này có thể giúp mô hình học sâu học được các đặc trưng không gian từ các góc nhìn khác nhau, giảm hiện tượng overfitting và cải thiện khả năng tổng quát hóa.
- Điều chỉnh hướng chính (Orientation adjustment): xoay hình ảnh để điều chỉnh hướng chính của đối tượng. Điều này có thể hữu ích trong các bài toán nhận dạng đối tượng hoặc phân loại, nơi việc xác định hướng chính của đối tượng có ý nghĩa quan trọng.
- Chỉnh sửa góc nhìn (Viewpoint adjustment): xoay hình ảnh để điều chỉnh góc nhìn của đối tượng.

Một số hàm chúng tôi sử dụng trong quá trình xử lý ảnh bao gồm:

- Rotate: là một hàm xử lý dữ liệu hình ảnh của thư viện Albumentations. Hàm này được sử dụng để thực hiện việc xoay hình ảnh. Hàm Rotate có thể bao gồm:
 - Limit (giới hạn): đây là một giá trị đối số để chỉ định phạm vi xoay hình ảnh. Ví dụ limit=45 cho phép xoay ảnh trong khoảng từ -45 độ đến +45 độ.

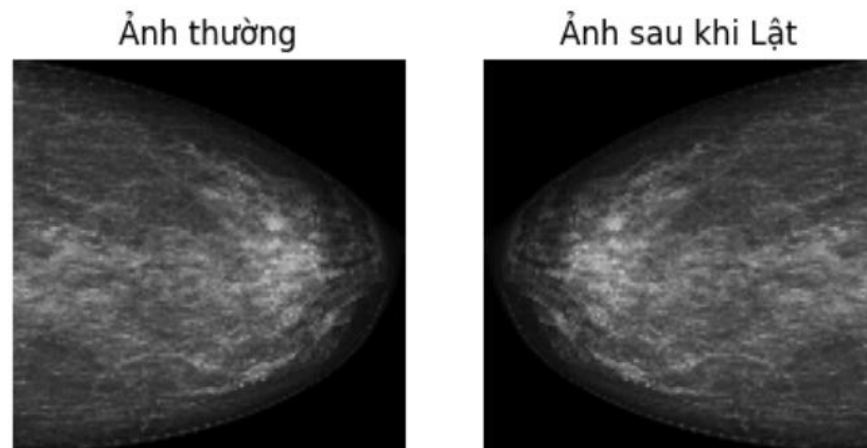
- interpolation (phương pháp nội suy): đây là một tham số để xác định phương pháp nội suy được sử dụng khi xoay hình ảnh. Các phương pháp nội suy thông thường bao gồm ‘bilinear’ (nội suy song tuyến) và ‘nearest’ (nội suy gần nhất) và ‘cubic’ (nội suy ba chức năng).
- p: xác suất áp dụng phép biến đổi trên hình ảnh. Tham số này thường được đặt trong khoảng từ 0 đến 1. Nếu $p=1$: phép biến đổi được áp dụng cho toàn bộ hình ảnh, $p=0$: phép biến đổi sẽ không được áp dụng cho bất kỳ hình ảnh nào.

Với các tham số trên, hàm Rotate cho phép xoay hình ảnh trong khoảng giới hạn cụ thể và sử dụng phương pháp nội suy chỉ định. Mặc định là $\text{limit}=45$, $p=1$ và interpolation là ‘bilinear’.



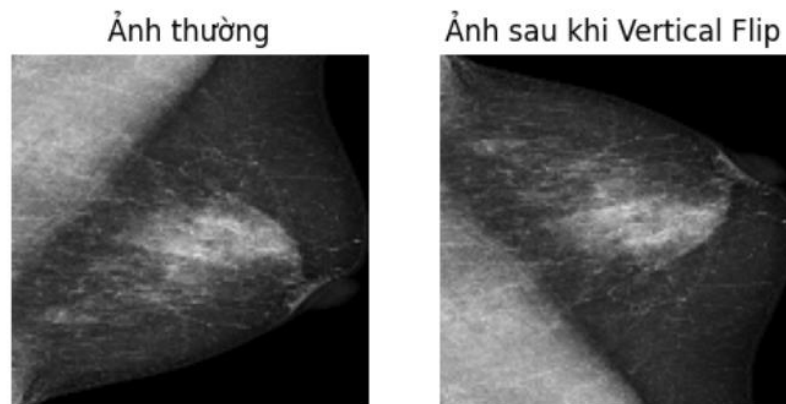
Hình 4.6. Hình ảnh minh họa sử dụng rotate với góc xoay 45 độ

- HorizontalFlip: thực hiện việc lật ngang hình ảnh. Hàm này nhận vào một đối tượng hình ảnh và trả về một hình ảnh mới sau khi đã được lật ngang. Nó sẽ đảo ngược hình ảnh qua trục hoành, tạo ra một phiên bản lật ngang của hình ảnh ban đầu. Hàm HorizontalFlip áp dụng thông số mặc định với $p=0.5$. Tức là có 50% hình ảnh sẽ được lật ngang.



Hình 4.7. Hình ảnh minh họa sử dụng kỹ thuật HorizontalFlip

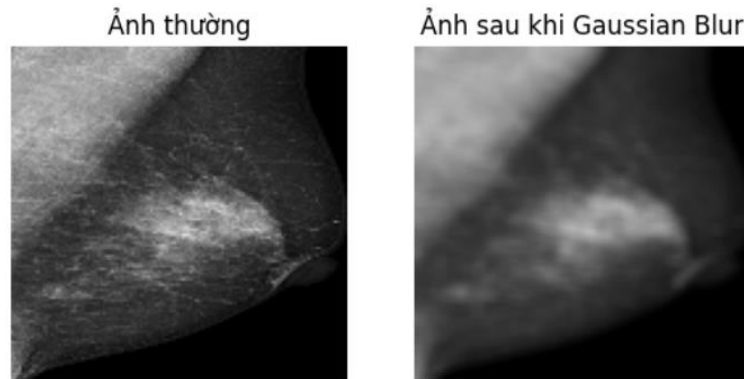
- VerticalFlip: dùng để thực hiện phép lật dọc trên một hình ảnh, tức là các điểm ảnh ở phía trên trở thành phía dưới và ngược lại. Hàm VerticalFlip không có các tham số đặc biệt và mặc định sử dụng xác suất là 0.5, tức là có 50% hình ảnh sẽ được lật dọc ($p=0.5$).



Hình 4.8. Hình ảnh mô tả sử dụng kỹ thuật VerticalFlip

- GaussianBlur: là một phương pháp xử lý ảnh thường được sử dụng để làm mờ hình ảnh bằng cách áp dụng một bộ lọc Gauss. Nó là một bộ lọc giảm nét hình ảnh bằng cách làm mờ các điểm ảnh theo phân phối Gauss. Quá trình làm mờ này giúp loại bỏ nhiễu trong ảnh. Hàm GaussBlur có các thông số blur_limit (float hoặc tuple): xác định mức độ làm mờ của hình ảnh. Nếu chỉ định một giá trị float thì mức độ làm mờ

sẽ được chọn ngẫu nhiên từ khoảng [0, 'blur_limit']. Nếu chỉ định một tuple (min, max) thì mức độ làm mờ sẽ được chọn ngẫu nhiên từ khoảng [min, max]. Giá trị mặc định là [3, 7].



Hình 4.9. Hình ảnh mô tả sử dụng kỹ thuật GaussianBlur

- Ngoài ra nhóm còn thực hiện các phép tăng cường ảnh như dưới đây:

```
Resize(height=224,width=224,always_apply=True),
    # crop, tweak from A.RandomSizedCrop()
    # flip
    HorizontalFlip(p=0.5),
    VerticalFlip(p=0.5),
    # contrast
    # A.OneOf([
    # # RandomToneCurve(scale=0.3, p=0.5),
    # RandomBrightnessContrast(brightness_limit=(-0.1, 0.2), co
ntrast_limit=(-0.4, 0.5), brightness_by_max=True, always_apply=False, p=0.5
    )
    # ], p=0.5),
    # geometric
    A.OneOf(
        [
            ShiftScaleRotate(shift_limit=None, scale_limit=[-0.15,
0.15], rotate_limit=[-30, 30], interpolation=cv2.INTER_LINEAR,
                                border_mode=cv2.BORDER_CONSTANT, va
lue=0, mask_value=None, shift_limit_x=[-0.1, 0.1],
                                shift_limit_y=[-0.2, 0.2], rotate_m
ethod='largest_box', p=0.6),
            A.ElasticTransform(alpha=1, sigma=20, alpha_affine=10,
interpolation=cv2.INTER_LINEAR, border_mode=cv2.BORDER_CONSTANT,
                                value=0, mask_value=None, approxima
te=False, same_dxdy=False, p=0.2),
```

```
A.GridDistortion(num_steps=5, distort_limit=0.3, interpolation=cv2.INTER_LINEAR, border_mode=cv2.BORDER_CONSTANT, value=0, mask_value=None, normalized=True, p=0.2),
                ], p=0.5),
            # random erase
            A.CoarseDropout(max_holes=6, max_height=0.15, max_width=0.25, min_holes=1, min_height=0.05, min_width=0.1, fill_value=0, mask_fill_value=None, p=0.25),
            ], p=0.9)
```

4.1.7 Các phương pháp xử lý dữ liệu mất cân bằng

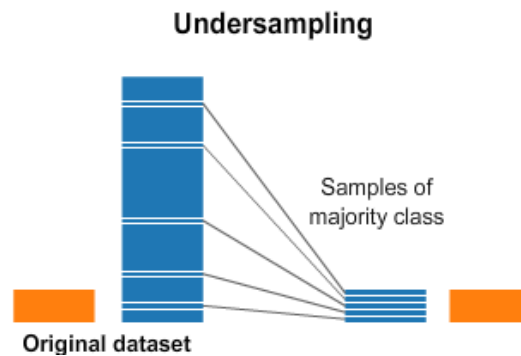
Mất cân bằng dữ liệu là một trong những hiện tượng phổ biến của bài toán phân loại nhị phân (binary classification) như spam email, phát hiện gian lận, dự báo vỡ nợ, chẩn đoán bệnh lý,... Trong trường hợp tỷ lệ dữ liệu giữa 2 classes là 50:50 thì được coi là cân bằng. Khi có sự khác biệt trong phân phối giữa 2 classes, chẳng hạn 60:40 thì dữ liệu có hiện tượng mất cân bằng.

Hầu hết các bộ dữ liệu đều khó đạt được trạng thái cân bằng mà luôn có sự khác biệt về tỷ lệ giữa 2 classes. Đối với những trường hợp dữ liệu mất cân bằng nhẹ như tỷ lệ 60:40 thì sẽ không ảnh hưởng đáng kể tới khả năng dự báo của mô hình.

4.1.8 Resampling

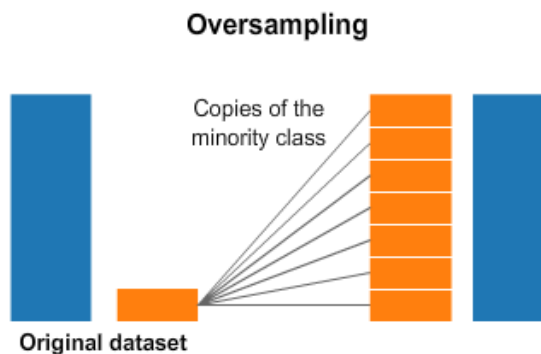
Resampling dữ liệu là một trong những cách tiếp cận được ưa thích nhất để đối phó với tập dữ liệu không cân bằng. Nói chung, có hai loại phương pháp cho việc này:

- Undersampling: giảm số lượng mẫu thu thập từ lớp thiểu số để cân bằng với lớp đa số.



Hình 4.10. Minh họa kỹ thuật undersampling¹⁵

- Oversampling: tăng số lượng mẫu của lớp thiểu số để cân bằng với lớp đa số.



Hình 4.11. Minh họa kỹ thuật oversampling¹¹

Trong quá trình giải quyết bài toán này, nhóm chúng tôi nhận thấy dữ liệu bị mất cân bằng khá lớn (88% dữ liệu là không ung thư và chỉ khoảng 12% dữ liệu mang nhãn ung thư), trong quá trình thực nghiệm, nhóm chúng tôi đã thực hiện thử nghiệm cả 2 phương pháp là undersampling và oversampling.

4.1.9 Trọng số phân lớp (Class weighting)

Là một trong những cách để giải quyết vấn đề mất cân bằng dữ liệu, class weight (hay pos weight) sử dụng để tăng độ ưu tiên cho lớp thiểu số trong quá trình huấn luyện. Các trọng số được gán cho mỗi lớp thường tỷ lệ nghịch với tần số của chúng trong tập dữ

¹⁵ [Undersampling and oversampling: An old and a new approach | by Nour Al-Rahman Al-Serw | Analytics Vidhya | Medium](#)

liệu, có nghĩa là nếu số lượng nhãn thiếu số càng ít, thì trọng số nhãn đó sẽ càng cao và ngược lại [43]. Và trọng số này sẽ được thêm vào hàm mất mát. Để tạo sự cân bằng giữa các lớp, ta thêm weight vào hàm mất mát, ta sẽ được công thức như sau [44]:

$$L_{CE} = - \sum_{i=1}^n w_i t_i \log(p_i) \quad (PT\ 4.1)$$

Trong đó, trọng số w_i được tính bằng công thức:

$$E = 1 - \frac{\sum_n p_{l(x)}}{N} \quad (PT\ 4.2)$$

4.1.10 Làm mịn nhãn (Labels smoothing)

Labels smoothing đã được sử dụng thành công để cải thiện độ chính xác của các mô hình học sâu trong nhiều tác vụ, bao gồm phân loại hình ảnh, nhận dạng giọng nói và dịch máy [45]. Label smoothing đề xuất thay vì sử dụng các giá trị nhãn 0 hoặc 1, ta sẽ sử dụng một phân phối mịn hơn, ví dụ như phân phối đều trên các lớp (uniform distribution) hoặc phân phối Gaussian, làm cho mô hình phân loại trở nên khó hơn và có khả năng tổng quát hóa tốt hơn trên dữ liệu mới. Là một trong những kỹ thuật regularization nên nó có thể làm giảm overfitting.

4.2 Dữ liệu thực nghiệm với mô hình ConvNeXt và CNNs:

Trong quá trình thực nghiệm nhiều lần, đối với mô hình ConvNeXt và CNNs, nhóm chúng tôi thực hiện 2 cách chia dữ liệu như sau:

- Cách 1 (TrainSet1): Chỉ sử dụng cố định 4 tập dữ liệu mini-DDSM, Mias, INbreast và CMMD để huấn luyện. Tập dữ liệu huấn luyện bao gồm 13 056 ảnh trong đó tỉ lệ ảnh 1 : 0 là 0.52 : 0.48. Số ảnh validate lấy từ mỗi bộ 5% số ảnh, tỉ lệ tương tự bằng cách lấy sample. Số ảnh test gồm 54 703 ảnh của bộ RSNA (bảng 2) và test riêng từng bộ dữ liệu:

Bảng 4.1. Chia dữ liệu TrainSet1

Dataset	Tổng ảnh	Ảnh ung thư	Train	Validate	Test
DDSM	7808	2716	7417	390	7808
Mias	322	51	306	16	322
INbreast	410	343	390	20	410
CMMD	5202	4094	4945	260	5202
RSNA	54703	1158	0	0	54703
Tổng	13742	7204	13056	686	68445

- Cách 2 (TrainSet2): Sử dụng 5 tập dữ liệu, gồm mini-DDSM, Mias, INbreast, CMMD và RSNA để huấn luyện. Riêng với bộ RSNA chỉ sử dụng 600 ảnh mang nhãn 0.

Bảng 4.2. Chia dữ liệu TrainSet2

Dataset	Tổng ảnh	Ảnh ung thư	Train	Validate	Test
DDSM	7808	2716	7417	390	7808
Mias	322	51	306	16	322
INbreast	410	343	390	20	410
CMMD	5202	4094	4945	260	5202
RSNA	54703	1158	600	0	54703
Tổng	13742	7204	13656	686	68445

4.3 Dữ liệu thực nghiệm với mô hình HRNet

4.3.1.1 Dữ liệu huấn luyện

Bảng 4.3. Dữ liệu huấn luyện mô hình HRNet

Tên Dataset	Số lượng nhãn Cancer	Số lượng nhãn Normal	Tổng số ảnh
DDSM	2172	4073	6245
Mias	40	216	256
INbreast	274	53	327
Tổng số	2486	4342	6828

4.3.1.2 Dữ liệu đánh giá

Bảng 4.4. Dữ liệu đánh giá mô hình HRNet

Tên Dataset	Số lượng nhãn Cancer	Số lượng nhãn Normal	Tổng số ảnh
DDSM	554	1019	1573
Mias	11	55	66
INbreast	69	14	83
Tổng số	634	1088	1722

4.4 Dữ liệu thực nghiệm với mô hình ResNet50

4.4.1.1 Dữ liệu huấn luyện

Bảng 4.5. Dữ liệu huấn luyện mô hình ResNet50

Tên Dataset	Số lượng nhãn Cancer	Số lượng nhãn Normal	Tổng số ảnh
DDSM	2172	2000	4172
Mias	40	216	256
INbreast	274	53	317
Tổng số	2486	2269	4755

4.4.1.2 Dữ liệu đánh giá

Bảng 4.6. Dữ liệu đánh giá mô hình ResNet50

Tên Dataset	Số lượng nhãn Cancer	Số lượng nhãn Normal	Tổng số ảnh
DDSM	544	500	1044
Mias	11	55	66
INbreast	69	14	83
Tổng số	624	569	1248

4.5 Công nghệ sử dụng

Toàn bộ quá trình thực nghiệm huấn luyện mô hình nhóm chúng tôi đều thực hiện trên kaggle với thông số phần cứng như sau:

Bảng 4.7. Mô tả phần cứng được sử dụng trong quá trình thực nghiệm

Thành phần phần cứng	Năm ra mắt	Số lõi	Bộ nhớ GPU	Số giờ khả dụng trên tuần
NVIDIA Tesla P100 GPU	2016	3584 Cuda cores	16 GB	30-32 h

4.6 Kết quả

4.6.1 Kết quả sau khi thực nghiệm

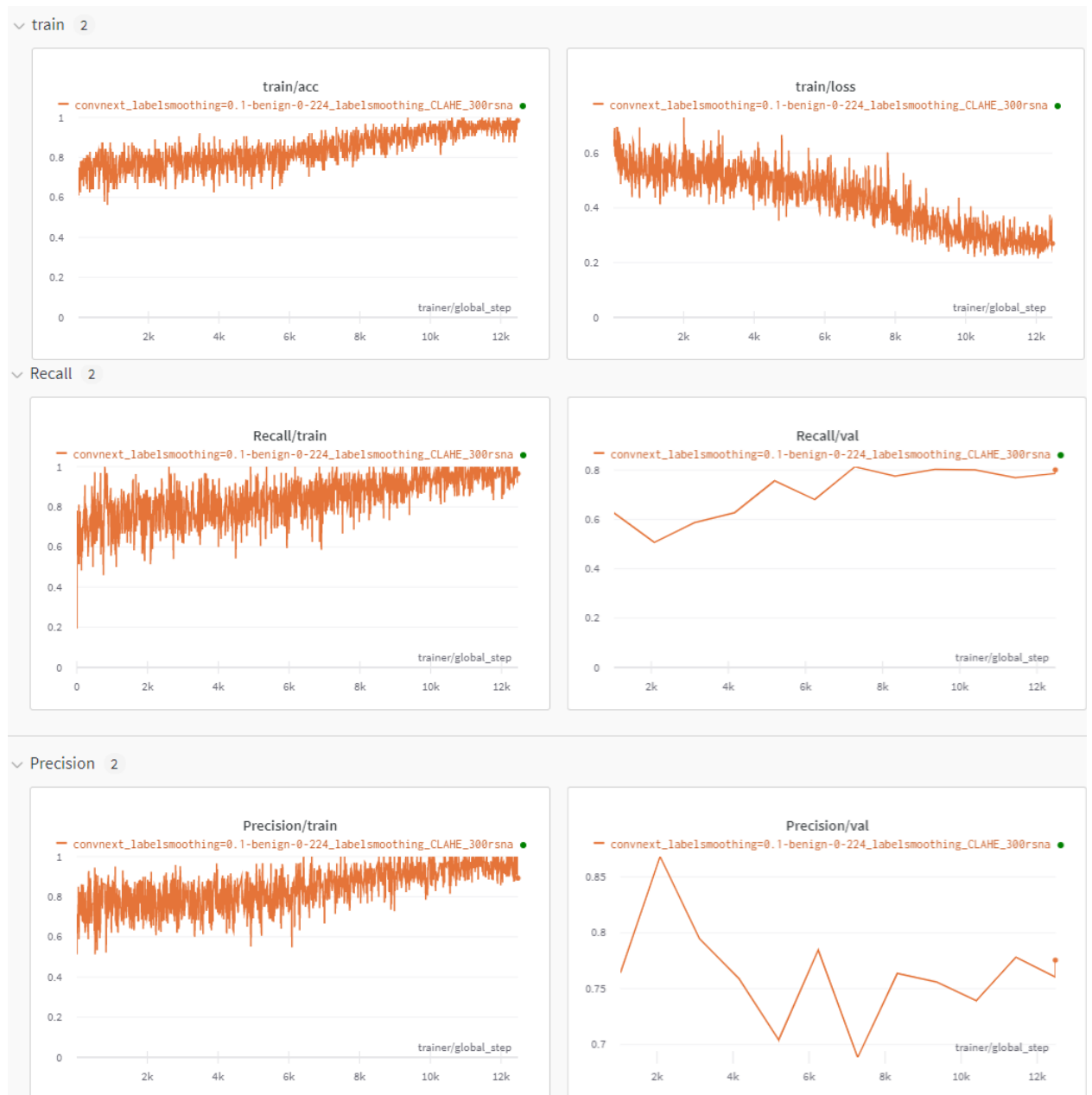
a. Kết quả mô hình ConvNeXt

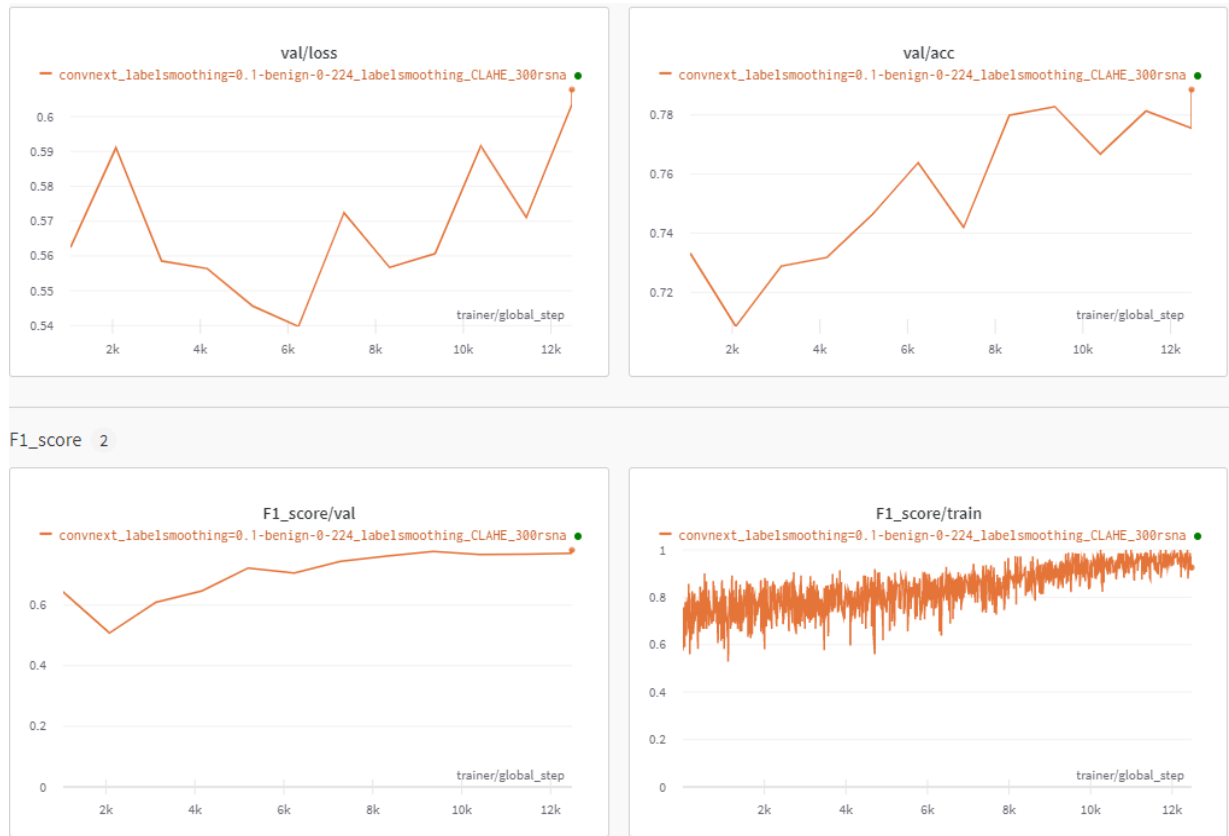
Tập dữ liệu train	Tập dữ liệu test	Tham số tinh chỉnh	Acc train	Acc test	Pretrained
Mias, RSNA(2316 ảnh tỉ lệ 1:0= 50:50, ta gọi bộ này là RSNA1158) ,	RSNA(2316 ảnh)	- Optimizer : SGD - loss: CrossEntropy() - lr: 0.01 - batch: 8 - size ảnh: 227x227x1 -scheduler: get_cosine_scheduler_with_warmup() - epoch: 20 epoch	62%	59%	Không
Mias, RSNA1158, DDSM (chỉ	RSNA1158	- Optimizer : RMSprop() - loss: CrossEntropy() - lr: 0.001	50%	44%	Không

lấy ảnh có nhãn 1)		<ul style="list-style-type: none"> - batch: 32 - size ảnh: 384x384x1 -scheduler: OneCycleLR() - init weight: xavier uniform - epoch : 20 epoch 			
Mias, INbreast, DDSM, RSNA1158,	Mias	<ul style="list-style-type: none"> - Optimizer : AdamW - loss: CrossEntropy(pos_weight) - lr: 0.0001 - batch: 128 - size ảnh: 227x227x3 -scheduler: OneCycleLR() - epoch: 10 epoch - oversampling data 	93%	57.7%	Có
RSNA, DDSM, CMMD, INbreast (95% tổng bộ dữ liệu)	RSNA, DDSM, CMMD, INbreast (5%)	<ul style="list-style-type: none"> - Optimizer : AdamW - loss: CrossEntropy(pos_weight, label_smoothing=0.1) - lr: 0.001 - batch: 16 - size ảnh: 384x384x3 -scheduler: CosineAnnealingWarmupRestarts () - epoch: 10 epoch - undersampling data 	87.5%	56.97%	Có
RSNA, DDSM,	RSNA, DDSM,	<ul style="list-style-type: none"> - Optimizer : AdamW 	50%	52.77%	Có

CMMD (đánh lại nhãn 'benign' thành 0), INbreast (95% tổng bộ dữ liệu)	CMMD, INbreast (5%)	-loss: CrossEntropy(pos_weight, label_smoothing=0.1) - lr: 0.001 - batch: 16 - size ảnh: 384x384x3x -scheduler: CosineAnnealingWarmupRestarts () - epoch: 10 epoch			
DDSM, INbreast, CMMD, Mias (95% tổng bộ dữ liệu)	DDSM, INbreast, CMMD, Mias (5% tổng bộ dữ liệu)	- Optimizer : AdamW -loss: CrossEntropy(pos_weight, label_smoothing=0.1) - lr: 0.001 - batch: 16 - size ảnh: 224x224x3x -scheduler: CosineAnnealingWarmupRestarts () - epoch: 15 epoch	87.5%	73.7%	Có
DDSM, INbreast, CMMD, Mias (95% tổng bộ dữ liệu)	DDSM, INbreast, CMMD, Mias (5% tổng bộ dữ liệu)	- Optimizer : AdamW -loss: CrossEntropy(pos_weight) - lr: 0.001 - batch: 64 - size ảnh: 224x224x3 -scheduler: CosineAnnealingWarmupRestarts () - epoch: 30 epoch	88.27%	77.25%	Có

DDSM, INbreast, CMMD, Mias (95% tổng bộ dữ liệu)	DDSM, INbreast, CMMD, Mias (5% tổng bộ dữ liệu)	- Optimizer : AdamW -loss: CrossEntropy(pos_weight, label_smoothing=0.1) - lr: 0.001 - batch: 64 - size ảnh: 224x224x3 -scheduler: CosineAnnealingWarmupRestarts () - epoch: 60 epoch	93.97%	77.84%	Có
DDSM, INbreast, CMMD, Mias (95% tổng bộ dữ liệu) + 600 ảnh nhãn 0 của bộ RSNA	DDSM, INbreast, CMMD, Mias (5% tổng bộ dữ liệu)	- Optimizer : AdamW -loss: CrossEntropy(pos_weight, label_smoothing=0.1) - lr: 0.001 - batch: 64 - size ảnh: 224x224x3 -scheduler: CosineAnnealingWarmupRestarts () - epoch: 60 epoch	93.87%	77%	Có



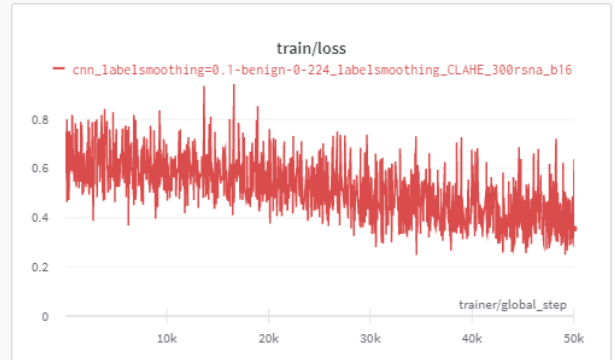
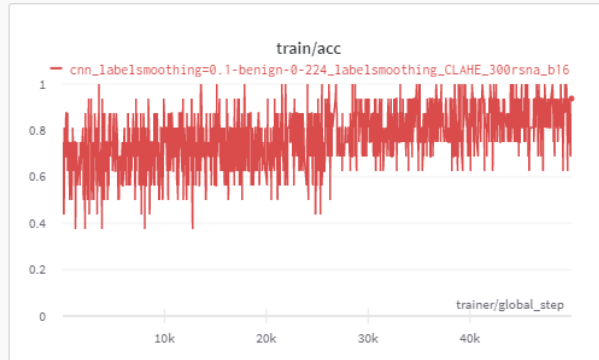


b. Kết quả mô hình CNN

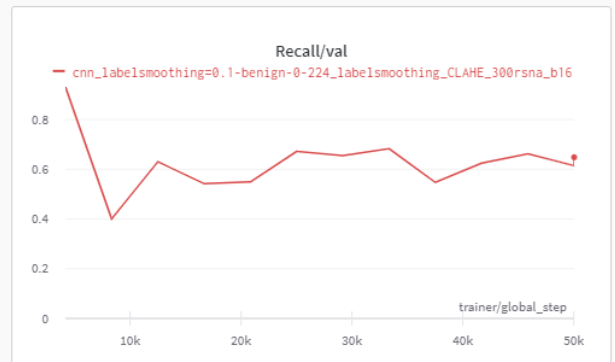
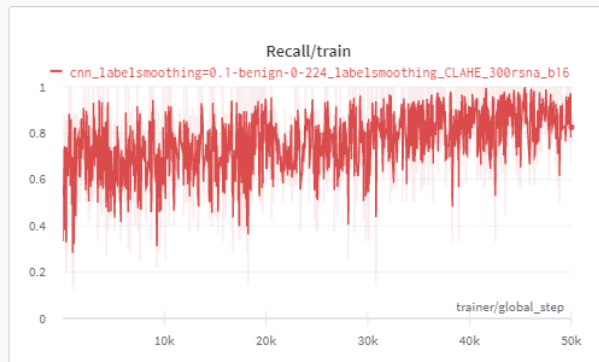
Tập dữ liệu train	Tập dữ liệu test	Tham số tinh chỉnh	Acc train	Acc test	Pretrained
RSNA1158, Mias, INbreast, DDSM(chỉ lấy ảnh nhãn 1)	RSNA	- Optimizer : SGD -loss: CrossEntropy() - lr: 0.1 - batch: 64 - size ảnh: 224x224 -scheduler: OneCycleLR () - epoch: 20 epoch	57.8%	50%	Không
DDSM, INbreast,	DDSM, INbreast, CMMD,	- Optimizer : AdamW -loss: CrossEntropy(pos_weight) - lr: 0.0001	67%	59%	Không

CMMD, Mias, RSNA	Mias (5% tổng bộ dữ liệu)	- batch: 128 - size ảnh: 227x227 -scheduler: OneCycleLR () - epoch: 10 epoch - oversampling			
DDSM, INbreast, CMMD, RSNA	DDSM, INbreast, CMMD, (5% tổng bộ dữ liệu)	- Optimizer : AdamW -loss: CrossEntropy(pos_weight, labels_smoothing=0.1) - lr: 0.0001 - batch: 128 - size ảnh: 227x227 -scheduler: OneCycleLR () - epoch: 20 epoch - undersampling	86.72%	58.93%	Không
DDSM, INbreast, CMMD, Mias (95% tổng bộ dữ liệu) + 300 ảnh nhãn 0 của bộ RSNA	DDSM, INbreast, CMMD, (5% tổng bộ dữ liệu)	- Optimizer : AdamW -loss: CrossEntropy(pos_weight, labels_smoothing=0.1) - lr: 0.001 - batch: 16 - size ảnh: 224x224 -scheduler: CosineAnnealingWarmupRestarts - epoch: 60 epoch - undersampling	84.63%	72.15%	Không

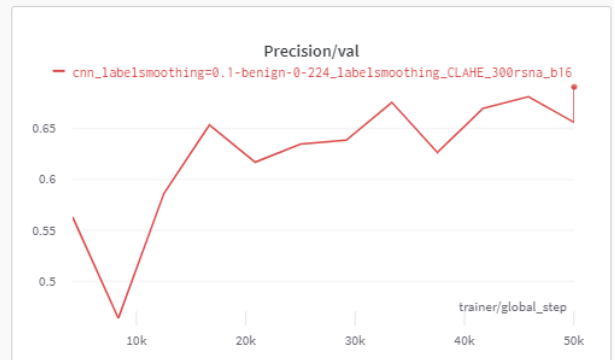
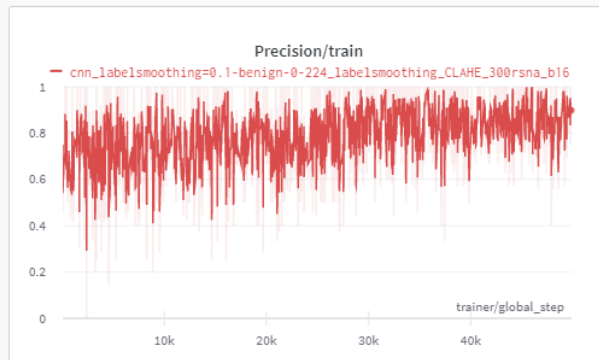
▼ train 2



▼ Recall 2



▼ Precision 2



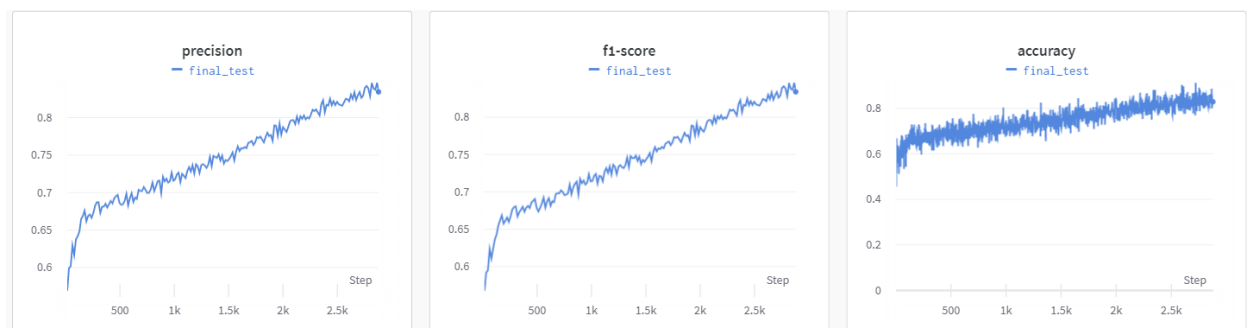


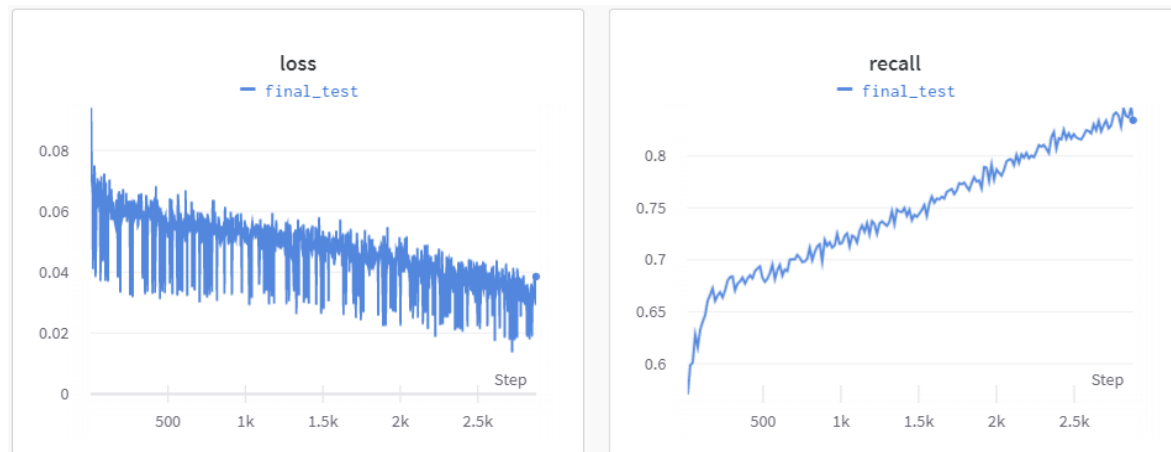
c. Kết quả mô hình ResNet50

Tập dữ liệu train	Tập dữ liệu test	Tham số tinh chỉnh	Acc train	Acc test	Pretrained
DDSM (4172 ảnh), Mias (256 ảnh), INbreast (327 ảnh)	DDSM (1044 ảnh), Mias (66 ảnh), INbreast (83 ảnh)	- Optimizer : SGD -loss: CrossEntropy() - lr: 0.005 - batch: 32 - size ảnh: 224x224 -scheduler: ReduceLROnPlateau - epoch: 180 epoch - weight_decay=0.0001 -momentum=0.9	85.74%	69.40%	Không

		- Không sử dụng GaussianBlur trên ảnh -factor = 0.1 -patience=5			
DDSM (4172 ảnh), Mias (256 ảnh), INbreast (327 ảnh)	DDSM (1044 ảnh), Mias (66 ảnh), INbreast (83 ảnh)	- Optimizer : SGD -loss: CrossEntropy() - lr: 0.005 - batch: 16 - size ảnh: 224x224 - Có sử dụng GaussianBlur trên ảnh -scheduler: ReduceLROnPlateau - epoch: 180 epoch - weight_decay=0.0001 -momentum=0.9 -factor = 0.1 -patience=5 -epoch: 80	99.99%	64.49%	Không

Kết quả tập train:

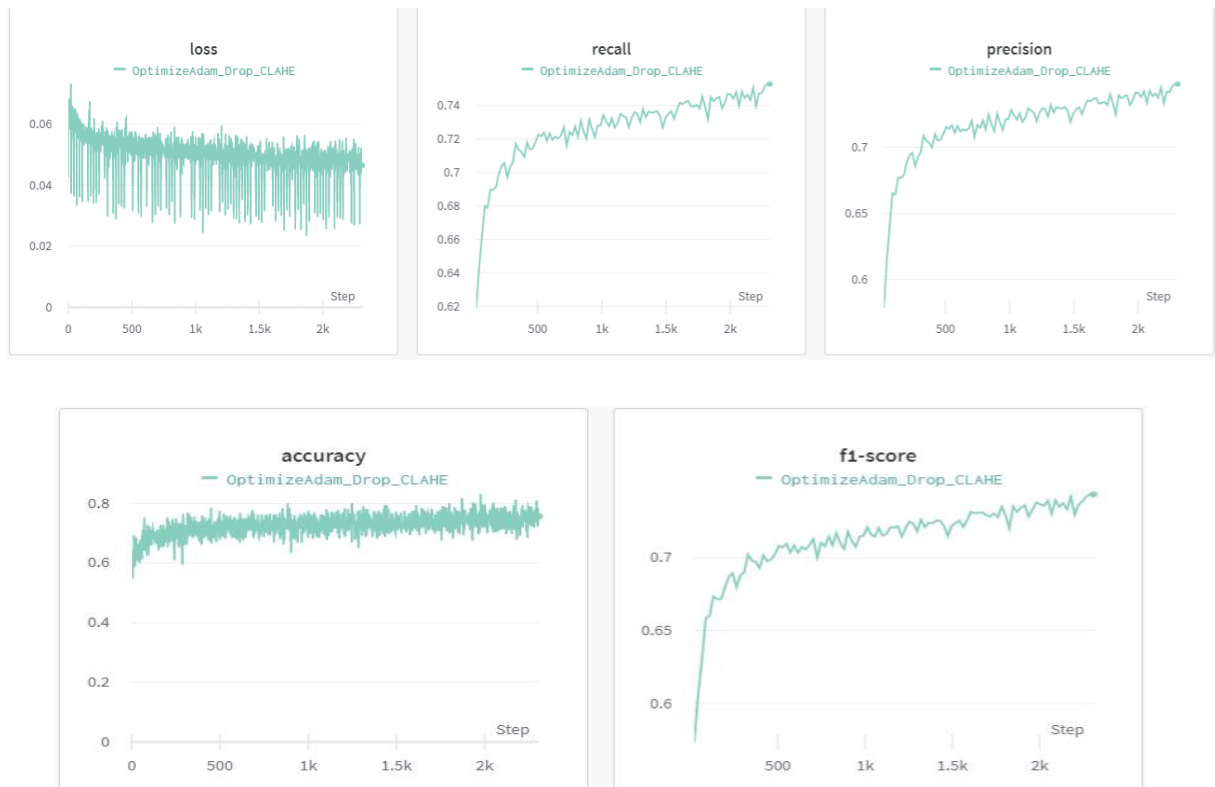




d. Kết quả mô hình HRNet

Tập dữ liệu train	Tập dữ liệu test	Tham số tinh chỉnh	Acc train	Acc test	Pretrained
DDSM (4172 ảnh), Mias (256 ảnh), INbreast (327 ảnh)	DDSM (1044 ảnh), Mias (66 ảnh), INbreast (83 ảnh)	- Optimizer : SGD -loss: CrossEntropy() - lr: 0.01 - batch: 32 - size ảnh: 512x512 - Có GaussianBlur trên ảnh -epoch: 105 -Không sử dụng CLAHE	74.55%	73.31%	Không
DDSM, Mias, INbreast (6848 ảnh)	DDSM, Mias, INbreast (1728 ảnh)	- Optimizer : SGD -loss: BCEWithLogitsLoss() - lr: 0.01 - batch: 32 - size ảnh: 512x512 - Có GaussianBlur trên ảnh -epoch: 40	64.4%	63.11%	Không

Kết quả tập train:



e. Kết quả đánh giá chung:

Mô hình	Train					Validate			
	Acc	Precision	F1 score	Recall	loss	Acc	F1 score	Recall	Precision
ConvNeXt	93.87%	95.95%	94.14%	95.5%	0.297	77%	78.98%	82.04%	76.16%
CNN	84.63%	85.07%	84.82%	84.78%	0.421	72.15%	63.72%	64.82%	69.12%
HRNet	74.37%	68.91%	60.51%	53.94%	0.04	71.20%	55.78%	49.83%	63.34%
ResNet50	84.48%	81.12%	86.05%	91.63%	0.03	69.15%	72.65%	78.36%	67.72%

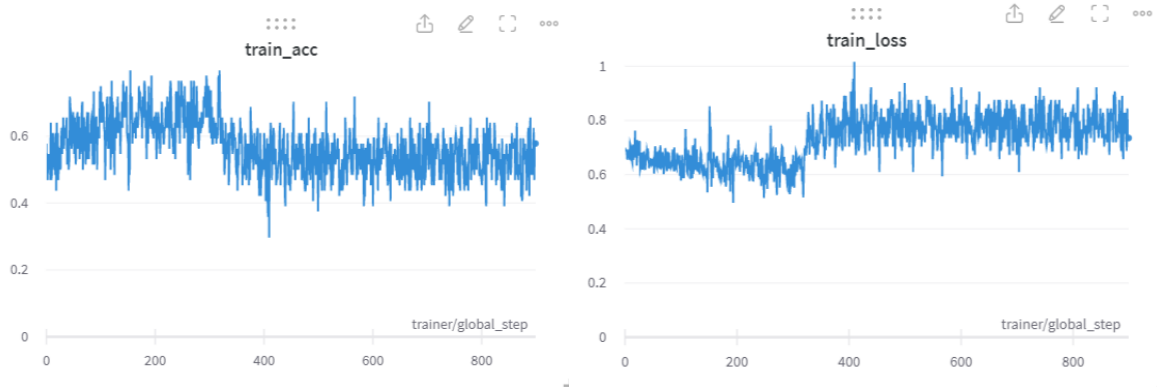
f. Kết quả độ chính xác trên từng bộ dữ liệu:

Mô hình	Mias				INbreast			
	Acc	Recall	Precision	F1 score	Acc	Recall	Precision	F1 score
CNN	77.01%	15.68%	20.51%	17.78%	38.29%	29.15%	90.90%	44.15%
ResNet50	48.48%	63.63%	18.91%	29.16%	53.01%	59.42%	78.84%	67.76%
ConvNeXt	97.52%	92.16%	92.16%	92.16%	96.58%	98.25%	97.68%	97.97%
HRNet	66.77%	41.17%	21.42%	28.18%	41.22%	35.27%	86.42%	50.10%
CNN (kết quả bài báo [19])	95.95%	-	-	-	96.3%	-	-	-

Mô hình	DDSM				CMMD			
	Acc	Recall	Precision	F1 score	Acc	Recall	Precision	F1 score
CNN	73.82%	45.51%	68.67%	54.74%	83.41%	96.68%	84.48%	90.17%
ResNet50	71.07%	71.32%	72.65%	71.98%	-	-	-	-
ConvNeXt	95.8%	92.56%	95.23%	93.88%	93.65%	96.68%	95.33%	96%
HRNet	65.30%	46.98%	50.13%	48.50%	43.98%	39.03%	79.26%	52.30%
CNN (kết quả bài báo [19])	99.39%	-	-	-	-	-	-	-

4.6.2 Nhận xét về kết quả đạt được

Sau vài lần train mà kết quả thấp và không cải thiện, nhóm em nhận thấy có một vấn đề đó là cả giá trị train và giá trị của hàm loss đều không giảm và giao động rất nhiều (hình 4.22).



Hình 4.12. Loss và accuracy của mô hình CNNs train ở giai đoạn đầu

Qua quá trình tham khảo các tài liệu và tham khảo cách làm trên các cuộc thi về xử lý bài toán phân loại ảnh ung thư, nhóm chúng tôi đưa ra các nguyên nhân dẫn đến kết quả ban đầu thấp và loss không giảm có thể do:

a. Dữ liệu không sạch

Ban đầu nhóm sử dụng train mô hình trên các tập dữ liệu được cắt bằng phương pháp sử dụng liên thông lớn nhất. Tuy nhiên ảnh vẫn còn tỉ lệ sai sót cao (10-20% là ảnh cắt sai hoặc không cắt được), giải pháp là sử dụng YOLOv8 để cắt ảnh (tỉ lệ chính xác khi xác định nhũ để cắt ảnh tới 99%).

Thứ 2, về đánh label cho ảnh, lúc đầu, các nhãn ‘benign’ và ‘malignant’ đều là 1 (có nghĩa đều là ung thư, chỉ nhãn ‘normal’ sẽ được hiểu là nhãn ‘0’, sau đó nhóm đã tìm hiểu lại thì nhận thấy ‘benign’ thường ít được quan tâm trừ khi nếu chúng xâm lấn qua các mô xung quanh hoặc gây chèn ép tới các cơ quan quan trọng gần đó như mạch máu hoặc dây thần kinh. Vì vậy nhóm em đã đổi các ảnh có nhãn ‘benign’ thành 0. Có nghĩa là:

- Nhãn 0: đại diện cho ‘benign’ (tổn thương lành tính) và ‘normal’ (bình thường).
- Nhãn 1: đại diện cho ‘malignant’ (tổn thương ác tính).

b. Mất cân bằng dữ liệu:

Dữ liệu của nhóm là khá lớn (hơn 64 nghìn ảnh), tuy nhiên số ảnh mang nhãn 1 gần như gấp 8 lần nhãn 0, mất cân bằng quá lớn, điều này có thể dẫn đến việc model học không đồng đều giữa 2 lớp 0 và 1, xác suất dẫn đến kết quả dự đoán là 0 (không ung thư) sẽ nhiều hơn. Điều này dẫn đến tình trạng âm tính giả, trên thực tế, điều này có thể dẫn đến bệnh nhân mắc bệnh nhưng không được điều trị kịp thời. Để giải quyết vấn đề mất cân bằng dữ liệu nhóm đã sử dụng các phương pháp sau:

- Resampling
- Pos weight
- c. Optimizer và scheduler

Tham khảo bài báo [22], trong quá trình cải tiến từ model ResNet50 lên một model “state-of-the-art”, bằng cách cải thiện phương pháp huấn luyện, hiệu suất của mô hình ResNet-50 đã tăng từ 76,1% lên 78,8% (+2,7%). Từ điều này, chúng tôi quyết định thực nghiệm với optimizer AdamW và scheduler là CosineAnnealingWarmupRestarts giống như trong bài báo sử dụng.

d. Mô hình chưa đủ tốt

Sử dụng pretrained thay vì chỉ sử dụng kiến trúc mô hình. Thay mô hình ConvNeXt của framework pytorch bằng các model pretrained được lấy từ HuggingFace.

Sau khi kết hợp các kỹ thuật và tinh chỉnh mô hình, kết quả đã cải thiện một cách đáng kể:

- Mô hình ConvNeXt tăng 30.23% về độ chính xác trên tập train, tăng 32.06% về độ chính xác trên tập test, tuy nhiên vẫn xảy ra tình trạng overfitting.
- Kết quả đạt được tốt nhất khi huấn luyện ảnh với kích thước 224x224x3 đối với mô hình ConvNeXt và 224x224x1 đối với mô hình CNNs.

- Phương pháp huấn luyện cũng đóng một vai trò đáng kể trong việc tăng độ chính xác cho mô hình (việc sử dụng optimizer, scheduler, chia batch size và chia dữ liệu huấn luyện).
- Đối với mô hình HRNet và ResNet50, mặc dù kết quả không cao nhưng lại không có hiện tượng bị overfitting. Để tăng độ chính xác, cần phải cải thiện mô hình với nhiều thông số khác nhau. Việc thử nghiệm các optimizer khác nhau vào hai mô hình này có sự thay đổi đáng kể.
- Đối với mô hình ResNet, vì kiến trúc mô hình khá lớn và phức tạp trong khi phần cứng có giới hạn, nên với kích thước ảnh 224x224 đang là tốt nhất hiện giờ của nhóm, theo những nghiên cứu khác của chúng tôi, việc tăng size ảnh có thể mang lại kết quả tốt hơn.

CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 Kết luận

Trong khóa luận này, chúng tôi đã thử nghiệm với 4 mô hình khác nhau là: CNN, HRNet, ResNet, ConvNeXt để hỗ trợ tiếp cận chẩn đoán ung thư từ nhũ ảnh sử dụng các bộ dữ liệu huấn luyện là Mias, DDSM, INbreast. So với bài báo gốc, phương pháp của chúng tôi đã có sự cải thiện trên 2 bộ dữ liệu là Mias và INbreast. Nhóm chúng tôi nhận thấy rằng, cách tiếp cận này phù hợp với bài toán và có ích trong việc chẩn đoán ung thư, giúp giảm thiểu thời gian và nguồn nhân lực của đội ngũ y tế.

Qua quá trình làm, nhóm chúng tôi cũng đã rút ra được nhiều kinh nghiệm quý báu trong bài toán phân loại ung thư từ nhũ ảnh nói riêng cũng như xử lý ảnh nói chung.

5.2 Kiến thức đạt được

Qua luận văn, nhóm chúng tôi đã có thêm được nhiều kinh nghiệm quý báu trong lĩnh vực xử lý ảnh nói chung cũng như xử lý ảnh y khoa nói riêng. Hiểu rõ hơn về các thuật toán CNNs, ConvNeXt, HRNet, ResNet50. Học hỏi và áp dụng được các cách cải thiện kết quả của một mô hình, hiểu được khi tối ưu kết quả một mô hình đòi hỏi kết hợp nhiều yếu tố như hàm mất mát, độ cân bằng dữ liệu, cách xử lý dữ liệu đầu vào,... Biết được thêm những kỹ thuật để xử lý dữ liệu ảnh và dữ liệu mất cân bằng, mà trên thực tế sẽ rất thường xuyên gặp dạng dữ liệu mất cân bằng như vậy.

5.3 Kỹ năng đạt được

Việc thực hiện luận văn đã góp phần nâng cao khả năng thực nghiệm kết quả, đọc và tham khảo các nghiên cứu khác của chúng tôi.

5.4 Hướng phát triển trong tương lai

Batch size và kích thước ảnh là hai thông số quan trọng trong quá trình huấn luyện mô hình. Khi batch size lớn, đồng nghĩa với việc đưa vào mô hình nhiều ảnh một lúc, giúp

mô hình có thể học tổng quát hơn. Khi kích thước ảnh tăng, có nghĩa là lượng thông tin lưu trữ cũng sẽ tăng, mô hình học được nhiều đặc trưng hơn sẽ dẫn đến kết quả tốt hơn. Tuy nhiên, để có thể đồng thời cả batch size lớn và kích thước ảnh lớn ta cũng cần lượng tài nguyên lớn hơn để đảm bảo tính ổn định về thời gian huấn luyện cũng như dự đoán, cho kết quả cao sẽ phù hợp hơn với những doanh nghiệp lớn có đội ngũ duy trì, nghiên cứu phát triển.

Trong quá trình nghiên cứu, nhóm em chỉ thực nghiệm chỉ được thử nghiệm trên phần cứng có giới hạn, vì vậy nếu được tham gia vào đội ngũ R&D ở các công ty hoặc có điều kiện nghiên cứu tiếp nhóm chúng tôi sẽ thử với batch size lớn hơn và kích thước ảnh lớn hơn. Trong tương lai, có thể sử dụng các phương pháp transfer learning khác như knowledge distillation để đào tạo mô hình, đồng thời tối ưu mô hình về cả kích thước và hiệu suất để có thể triển khai được vào dự án thực tế. Ngoài ra, để nâng cao độ chính xác cho hai mô hình HRNet và ResNet50, nhóm sẽ thử áp dụng pretrain cho 2 mô hình này.

TÀI LIỆU THAM KHẢO

- [1] D. Tính, “Tỷ lệ mắc ung thư mới và tử vong do ung thư tiếp tục tăng,” 1 12 2022. [Trực tuyến]. Available: <https://thanhnien.vn/ty-le-mac-ung-thu-moi-va-tu-vong-do-ung-thu-tiep-tuc-tang-1851527371.htm>.
- [2] Kosmia Loizidou, Rafaella Elia, Costas Pitris, “Computer-aided breast cancer detection and classification in mammography: A comprehensive review,” *Computers in Biology and Medicine*, tập 153, 2023.
- [3] "vinmec.com," 2019. [Online]. Available: <https://www.vinmec.com/vi/oncology-radiotherapy/health-news/distinguish-between-benign-and-malignant-tumors/>.
- [4] "nibib.nih.gov," 2023. [Online]. Available: <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>.
- [5] Macedo Firmino, Giovani Angelo, Higor Moraes, Marcel R. Dantas và Ricardo Valentim, "Computer-aided detection (CADE) and diagnosis (CADx) system for lung cancer with likelihood of malignancy," in *BioMed Eng OnLine*, 06 January 2016, 06 January 2016, p. 15.
- [6] Megha Rathi, Vikas Pareek, "Hybrid approach to predict breast cancer using machine," *International Journal of Computer Science Engineering*, vol. vol. 5, pp. 125-136, 2016.
- [7] M. Tahmooresi, A. Afshar, B. Bashari Rad, K. B. Nowshath, M. A. Bamiah, "Early detection of breast cancer using machine learning techniques," *Journal of Telecommunication*, , vol. 10, pp. 21-27, 2018.

- [8] Muhammet Fatih Aslam, YunusCelik, KadirSabanci, AkifDurdu, ""Breast cancer diagnosis by different machine learning method using blood analysis data,," *International Journal of Intelligent* , vol. 6, no. no. 04, pp. 289-293, 2018.
- [9] Anusha bharat, Pooja N, R Anishka Reddy, "Using machine learning algorithms for breast," *IEEE 3rd International Conference on Circuits, Control* , pp. 1-4, 2018.
- [10] Ebru Aydindag Bayrak, Pinar Kirci, TolgaEnsari, "Comparison of machine learning methods for breast cancer diagnosis," *Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, pp. 1-3, 2019.
- [11] S. M. S. S. S. C. D. Shwetha K, "Breast cancer detection using deep learning technique," *International Journal of Engineering Research & Technology*, pp. 1-4, 2018.
- [12] Ch. Shravya, K. Pravalika, ShaikSubhani, ""Prediction of breast cancer using supervised machine learning techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 1106-1110, 2019.
- [13] Sivapriya J, Aravind Kumar V, Siddarth Sai S, Sriram S, "Breast cancer prediction using machine learning," *International Journal of Recent Technology and Engineering*, vol. 8, pp. 4879-4881, 2019.
- [14] Kalyani Wadkar, Prashant Pathak, Nikhil Wagh, "Breast cancer detection using ANN network and performance analysis with SVM," *International Journal of Computer Engineering and Technology*, vol. 8, no. 6S, pp. 501-505, 2019.

- [15] Vishal Deshwal, Mukta Sharma, "Breast cancer detection using SVM classifier with grid search techniques,," *International Journal of Computer Application*,, vol. 178, pp. 18-23, 2019.
- [16] V Sansya Vijayan, Lekshmy P L, "Deep learning based prediction of breast cancer in histopathological images,," *International Journal of Engineering Research & Technology*, vol. 8, pp. 148-152, 2019.
- [17] S. Saha, "saturncloud.io," 15 12 2018. [Trực tuyến]. Available: <https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/>.
- [18] Afshine Amidi, Shervine Amidi, "starnford.edu," [Trực tuyến]. Available: <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks>.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.
- [20] T. Jacob, 25 2 2022. [Online]. Available: <https://www.kdnuggets.com/2022/02/vanishing-gradient-problem.html#:~:text=When%20there%20are%20more%20layers,this%20the%200vanishing%20gradient%20problem..>
- [21] Chigozie Enyinna Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," 2018.

- [22] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie, "A ConvNet for the 2020s," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022.
- [23] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz, mixup: Beyond empirical risk minimization, *International Conference on Learning Representations (ICLR)*, 2018.
- [24] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo, In *ICCV*, 2019.
- [25] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le, Randaugment: Practical automated data augmentation with a reduced search space, *CVPR Workshops*, 2020.
- [26] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang, Random erasing data augmentation, *AAAI*, 2019.
- [27] Dan Hendrycks, Kevin Gimpel, GAUSSIAN ERROR LINEAR UNITS (GELUS), *arXiv*, 2020.
- [28] J. Xu, Y. Zhang and D. Miao, "Three-way confusion matrix for classification: A measure driven view,," *Information sciences*, vol. 507, pp. 772-794, 2020.
- [29] J. Brownlee, "Machine Learning Mastery,," August 2020. [Online]. Available: <https://machinelearningmastery.com/confusion-matrix-machine-learning/>.
- [30] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, pp. 86-874, 2006.

- [31] R. Karthiga, K. Narasimhan, Rengarajan Amirtharajan, "Diagnosis of breast cancer for modern mammography using artificial," *Mathematics and Computers in Simulation*, vol. 202, pp. 316-330, 2022.
- [32] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu,, "Deep High-Resolution Representation Learning for Visual Recognition for Visual Recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 43, pp. 3349-3364, 2021.
- [33] "Weight & Bias," Saurav Maheshkar, 15 3 2022. [Online]. Available: <https://wandb.ai/sauravmaheshkar/cross-entropy/reports/What-Is-Cross-Entropy-Loss-A-Tutorial-With-Code--VmlldzoxMDA5NTMx>.
- [34] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv*, 2017.
- [35] Diederik P. Kingma, Jimmy Ba, Adam: A Method for Stochastic Optimization, 2017.
- [36] Ultralytics, "Github," 10 1 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics/issues/189?fbclid=IwAR0JnGZWKL6tc5aPzVst3aRUb1V1qITuXwZdRnU3EXOpDrHjUXQy7AOzleU>.
- [37] "Ultralytics," [Online]. Available: https://docs.ultralytics.com/yolov5/tutorials/architecture_description/#44-build-targets.
- [38] "National Library of Medicine," [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22078258/>.
- [39] "INbreast: Toward a Full-field Digital Mammographic Database".

- [40] "MedicMind," [Online]. Available: <https://www.medicmind.tech/cancer-imaging-data#:~:text=The%20Mammographic%20Image%20Analysis%20Society,a%20database%20of%20digital%20mammograms..>
- [41] Stephen M. Pizer *, †, E. Philip Amburn ‡, John D. Austin *, Robert Cromartie *, Ari Geselowitz §, Trey Greer *, Bart ter Haar Romeny ", John B. Zimmerman , Karel Zuiderveld ", "Adaptive histogram equalization and its variations," *Science Direct*, vol. 39, no. 3, pp. 355-368, 1987.
- [42] "MathWorks," [Online]. Available: <https://www.mathworks.com/help/visionhdl/ug/contrast-adaptive-histogram-equalization.html>.
- [43] A. Kumar, "Data Analytics," 1 8 2023. [Online]. Available: <https://vitalflux.com/class-imbalance-class-weight-python-sklearn/#:~:text=Class%20weighting%3A%20Using%20class%20weight,represented%20relative%20to%20the%20other..>
- [44] "StackExchange," 2020. [Online]. Available: <https://stats.stackexchange.com/questions/439491/understanding-the-weighted-cross-entropy-method-of-u-net>.
- [45] Rafael Müller*, Simon Kornblith, Geoffrey Hinton, "When Does Label Smoothing Help?," *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019.
- [46] Kaiming He, Xiangyu Zhang, Shaoqing Ren Jian Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016.

- [47] T. Jacob, "kdnuggets.com," 25 2 2022. [Online]. Available: <https://www.kdnuggets.com/2022/02/vanishing-gradient-problem.html#:~:text=When%20there%20are%20more%20layers,this%20the%20vanishing%20gradient%20problem..>
- [48] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," *CVPR*, 2017.
- [49] P. Đ. Khánh, "Khoa học dữ liệu - Khanh's blog," 23 Aug 2020. [Online]. Available: <https://phamdinhhkhanh.github.io/2020/08/23/FocalLoss.html>.
- [50] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 318-327, 2020.
- [51] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 730-734, 2015.
- [52] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2481-2495, 2017.
- [53] H. Noh, S. Hong and B. Han, "Learning Deconvolution Network for Semantic Segmentation," *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1520-1528, 2015.
- [54] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv*, 2015.

- [55] Alejandro Newell, Kaiyu Yang, Jia Deng, "Stacked Hourglass Networks for Human Pose Estimation," *arXiv*, 2016.
- [56] Xi Peng, Rogerio S. Feris, Xiaoyu Wang, Dimitris N. Metaxas, "A Recurrent Encoder-Decoder Network for Sequential Face Alignment," *arXiv*, 2016.
- [57] Bin Xiao, Haiping Wu, Yichen Wei, "Simple Baselines for Human Pose Estimation and Tracking," *arXiv*, 2018.
- [58] Jucelino Cardoso Marciano dos Santos, Gilberto Arantes Carrijo, Cristiane de Fátima dos Santos Cardoso, Júlio César Ferreira, Pedro Moises Sousa & Ana Cláudia Patrocínio , "Fundus image quality enhancement for blood vessel detection via a neural network using CLAHE and Wiener filter," *Research on Biomedical Engineering* , pp. 107-119, 2020.

NHẬT KÍ LÀM VIỆC

Tuần	Từ ngày	Đến ngày	Nội dung
1	01/01/2023	08/08/2023	Đọc hiểu đề tài và các bài báo liên quan
2	08/01/2023	16/01/2023	Tìm kiếm nguồn dữ liệu giải quyết bài toán
3	17/01/2023	22/01/2023	EDA dữ liệu
4	18/01/2023	31/01/2023	Tiền xử lí dữ liệu
5	01/02/2023	06/02/2023	Đọc và tìm hiểu các mô hình (CNNs, ResNet50)
6	08/02/2023	21/02/2023	Xây dựng mô hình
7	17/02/2023	19/02/2023	Huấn luyện mô hình CNNs với bài toán phân loại
8	17/02/2023	19/02/2023	Huấn luyện mô hình ResNet50 với bài toán phân loại
9	19/02/2023	20/02/2023	Đánh giá kết quả
10	21/02/2023	05/03/2023	Tìm hiểu các phương pháp xử lý ảnh khác
11	07/03/2023	10/03/2023	Chuẩn bị lại dữ liệu
12	07/03/2023	18/03/2023	Cắt ảnh bằng phương pháp liên thông
13	20/03/2023	21/03/2023	Huấn luyện lại mô hình, đánh giá mô hình
14	21/03/2023	31/03/2023	Tìm hiểu mô hình YOLOv8 để cắt ảnh

15	31/03/2023	04/04/2023	Huấn luyện mô hình YOLOv8 để cắt ảnh
16	05/04/2023	07/04/2023	Đánh giá kết quả
17	08/04/2023	11/04/2023	Tìm hiểu các kỹ thuật cải tiến mô hình
18	10/04/2023	16/04/2023	Tìm hiểu mô hình HRNet và ConvNeXt cho bài toán phân loại nhũ ảnh
19	14/04/2023	28/04/2023	Huấn luyện mô hình HRNet, ConvNeXt, CNNs và ResNet50 trên dữ liệu cắt bằng YOLOv8
20	29/04/2023	30/04/2023	Đánh giá mô hình
21	01/05/2023	15/05/2023	Viết báo cáo