

Trường Đại học Khoa học Tự nhiên - Đại học Quốc gia Hà Nội

Khoa Toán - Cơ - Tin học



Ứng dụng học sâu phân loại ảnh vi thể tế bào học tuyến giáp theo thang chuẩn Bethesda

Môn học: Phương pháp nghiên cứu khoa học

Sinh viên thực hiện:

Phạm Ngọc Hải (21002139)

Giảng viên hướng dẫn:

PGS.TS Lê Hồng Phương

Ngày 7 tháng 6 năm 2024

Lời cảm ơn

Trước hết, em xin bày tỏ lòng biết ơn sâu sắc đến giảng viên hướng dẫn, **PGS.TS Lê Hồng Phương**, đã tận tình hướng dẫn, chỉ bảo và đưa ra những ý kiến quý báu giúp em hoàn thành nghiên cứu này. Đặc biệt, xin cảm ơn thầy đã liên kết với Bệnh viện Quân đội Trung ương 108 để cung cấp dữ liệu cần thiết, tạo điều kiện thuận lợi cho quá trình nghiên cứu của em.

Em cũng xin cảm ơn các thầy cô giáo trong khoa Toán - Cơ - Tin học - Trường Đại học Khoa học Tự nhiên - Đại học Quốc gia Hà Nội, đã truyền đạt kiến thức, kỹ năng cần thiết trong suốt quá trình học tập, tạo nền tảng vững chắc để em thực hiện nghiên cứu.

Bên cạnh đó, em xin cảm ơn các bạn bè, anh/chị khóa trước đã có những chia sẻ kinh nghiệm làm nghiên cứu khoa học; giúp em chủ động lập kế hoạch nghiên cứu một cách rõ ràng.

Cuối cùng, em xin gửi lời cảm ơn sâu sắc đến gia đình đã luôn ủng hộ, động viên và tạo điều kiện tốt nhất cho quá trình nghiên cứu của em.

Em xin chân thành cảm ơn!

Phân công việc và đánh giá

Trong quá trình thực hiện nghiên cứu, em đã đảm nhận các công việc một cách độc lập, từ lập kế hoạch đến viết báo cáo. Dưới đây là chi tiết về các công việc và nhận xét tự đánh giá.

Kế hoạch và phân công việc

Lập kế hoạch: Xác định mục tiêu nghiên cứu. Lập kế hoạch công việc và phân bổ thời gian hợp lý cho từng giai đoạn.

Tìm hiểu tài liệu: Nghiên cứu và tổng hợp các tài liệu liên quan từ các nguồn tin cậy như sách, bài báo khoa học, và các tài liệu trên internet.

Triển khai nghiên cứu: Tìm hiểu dữ liệu, tiền xử lý dữ liệu. Triển khai thuật toán. Kiểm thử và đánh giá kết quả.

Viết báo cáo: Tổng hợp và viết các phần của báo cáo nghiên cứu.

Đánh giá công việc

Khó khăn và thách thức:

- Triển khai nghiên cứu một mình nên khi gặp vấn đề sức khỏe dễ ảnh hưởng tiến độ nghiên cứu.
- Cần phải cân đối thời gian giữa việc học và việc thực hiện nghiên cứu.

Kết quả đạt được:

- Hoàn thành đúng kế hoạch đã đề ra.
- Đạt được những kết quả quan trọng, đóng góp vào mục tiêu nghiên cứu ban đầu.

Bài học kinh nghiệm:

- Học được cách lập kế hoạch và quản lý thời gian hiệu quả.
- Nâng cao kỹ năng phân tích và viết báo cáo khoa học.

Nhìn chung, quá trình thực hiện nghiên cứu này không chỉ giúp em hoàn thành báo cáo mà còn mang lại nhiều kinh nghiệm quý báu và kỹ năng mới, góp phần vào sự phát triển cá nhân và học tập của em.

Mục lục

Tóm tắt	1
1 Giới thiệu	1
1.1 Bối cảnh và lý do chọn đề tài	1
1.1.1 Tổng quan về ung thư tuyến giáp	1
1.1.2 Tổng quan về các phương pháp chẩn đoán ung thư tuyến giáp	1
1.1.3 Lý do chọn đề tài	2
1.2 Mục tiêu nghiên cứu	2
1.3 Cấu trúc phần còn lại của báo cáo	3
2 Tổng quan tài liệu	3
3 Mô tả bộ dữ liệu	4
3.1 Nguồn gốc dữ liệu	4
3.2 Số lượng và đặc điểm ảnh	5
3.3 Cách truy cập và sử dụng	5
3.4 Cách thức thu thập	5
4 Phương pháp nghiên cứu	5
4.1 Giới thiệu	5
4.2 Bài toán cắt ra vùng quan tâm của ảnh	6
4.2.1 Giả thuyết đặt ra và ý nghĩa bài toán	6
4.2.2 Các bước thuật toán	7
4.2.3 Mô hình xác định/phân đoạn tế bào	8
4.2.4 Thuật toán phân cụm để chọn ra những cụm tối ưu	10
4.3 Bài toán phân nhãn ảnh sinh thiết	11
5 Kết quả	13
5.1 Các thang đo đánh giá	13
5.1.1 Thước đo đánh giá mô hình xác định tế bào	13
5.1.2 Thước đo đánh giá giải thuật phân cụm	15
5.1.3 Thước đo đánh giá mô hình phân loại ảnh	15
5.2 Kết quả cụ thể	17
5.2.1 Mô hình xác định tế bào	17
5.2.2 Kết quả giải thuật phân cụm	18
5.2.3 Đánh giá kết quả ảnh cắt thu được	18

5.2.4	Kết quả mô hình phân loại nhân ảnh sinh thiết	19
6	Thảo luận	21
6.1	Về kết quả đạt được	21
6.2	Về mục tiêu hiện tại và phương hướng tương lai	21
7	Kết luận	22
	Tài liệu tham khảo	23

Danh sách hình vẽ

1	Trực quan hóa bộ dữ liệu trên 569 ảnh	7
2	Kiến trúc mạng YOLO version 8 cho tác vụ xác định đối tượng	9
3	Kiến trúc chung của 1 mạng VGG16	12
4	Cách tính IoU	14
5	Giá trị mAP của mô hình xác định tế bào trong quá trình huấn luyện	18
6	Đánh giá các hàm loss trong quá trình huấn luyện mô hình xác định tế bào	18
7	Ví dụ về cách các ảnh được chọn cắt	19
8	Mô hình H1	20
9	So sánh kết quả dùng ảnh gốc và ảnh cắt với mô hình H1	20
10	Kiến trúc mô hình H2	20
11	Kết quả mô hình H2 trên tập kiểm tra của bộ dữ liệu ảnh cắt	21

Tóm tắt

Mục tiêu: Phân loại ảnh vi thể tế bào tuyến giáp theo thang chuẩn Bethesda để xây dựng hệ thống chẩn đoán hỗ trợ bằng máy tính (CAD) cho các bác sĩ. **Phương pháp:** Sử dụng các mô hình học sâu xác định vùng quan tâm của ảnh và phân loại nhân tương ứng của ảnh. **Kết quả:** Mô hình được huấn luyện trên tập 7698 ảnh và kiểm chứng trên tập 954 ảnh. Kết quả thực nghiệm trên 1491 ảnh của tập kiểm tra độc lập cho độ chính xác đạt được tới 87%. **Kết luận:** Kết quả nghiên cứu hiện tại cho thấy tiềm năng của việc áp dụng mô hình học sâu trong phân loại ảnh vi thể tế bào tuyến giáp theo thang chuẩn Bethesda. Đây là cơ sở giúp xây dựng hệ thống hỗ trợ bằng máy tính (CAD) hiệu quả.

Từ khóa: Ảnh sinh thiết chọc hút bằng kim nhỏ (FNAB), Trí tuệ nhân tạo (AI), Học sâu (DL), Mạng tích chập (CNN)

1 Giới thiệu

1.1 Bối cảnh và lý do chọn đề tài

1.1.1 Tổng quan về ung thư tuyến giáp

Ung thư là nguyên nhân gây tử vong hàng đầu, ảnh hưởng đến nhiều bộ phận cơ thể khác nhau. Trong các loại ung thư thì ung thư tuyến giáp là loại ung thư nội tiết phổ biến nhất [1]. Chiếm 52,070 ca trong tổng số 54,740 ca bệnh ung thư nội tiết mới theo ước tính tại Mỹ năm 2019 [2]. Đặc biệt ở nữ giới có tỷ lệ mắc bệnh phổ biến thứ 5 với 4% ca bệnh mới hàng năm [2]. Số ca được chẩn đoán mắc bệnh đang gia tăng do sự phổ cập dịch vụ y tế hoặc cũng có thể từ việc chẩn đoán quá mức [3]. Điều này đặt ra khối lượng công việc lớn với các bác sĩ do thiếu nhân lực chất lượng cao hoặc/và tăng nguy cơ bệnh nhân phải thực hiện nhiều biện pháp chẩn đoán, phẫu thuật không cần thiết.

1.1.2 Tổng quan về các phương pháp chẩn đoán ung thư tuyến giáp

Hiện tại ung thư tuyến giáp có thể xác định qua sờ nắn bằng tay, xét nghiệm máu, phân tích ảnh chụp siêu âm, ảnh sinh thiết chọc hút bằng kim nhỏ (FNAB), phân tích ảnh phẫu thuật cắt bỏ mô và phân tích phân tử [4].

Trong đó các phương pháp như sờ nắn, xét nghiệm máu, siêu âm có chi phí rẻ, dễ thực hiện, ít xâm lấn. Đặc biệt phương pháp siêu âm đã có nhiều nghiên cứu giúp tối ưu hóa quá trình [5, 6, 7].

Siêu âm là phương pháp hình ảnh ưa thích để xác định xem có cần sinh thiết hay không. Hút kim nhỏ (FNAB) rất quan trọng trong việc quyết định liệu nên phẫu thuật hay giám sát định kỳ qua

siêu âm. Ung thư có đặc điểm đáng ngờ trên siêu âm có thể cần phân tích tế bào học để đánh giá nguy cơ ác tính. Hiệu quả của một số xét nghiệm phân tử bổ sung vẫn chưa chắc chắn, mặc dù một số nghiên cứu báo cáo kết quả đầy hứa hẹn. Phương pháp quản lý và điều trị chủ yếu phụ thuộc vào kết quả tế bào học (FNAB) và đặc điểm siêu âm. Chiến lược điều trị tối ưu bao gồm từ theo dõi đơn giản cho các trường hợp có nguy cơ thấp đến can thiệp phẫu thuật cho bệnh nhân có nguy cơ cao [8].

Và vì thế, hiện nay trước khi thực hiện phẫu thuật loại bỏ tuyến giáp, sinh thiết chọc hút bằng kim nhỏ (FNAB) với đánh giá theo thang chuẩn Bethesda vẫn đang là tiêu chuẩn vàng [9]. Thang Bethesda phân loại 1 mẫu sinh thiết vào 1 trong 6 nhãn lần lượt là: không chẩn đoán (ND), lành tính, không điển hình có ý nghĩa không xác định/tổn thương nang trứng có ý nghĩa không xác định (AUS/FLUS), u nang/nghi ngờ u nang (FN/SFN), nghi ngờ ác tính (SUS) và ác tính [9]. Với nhiều mức phân loại theo chuẩn thống nhất chung đã giúp cung cấp phương pháp điều trị linh hoạt và quy trình chẩn đoán rõ ràng.

1.1.3 Lý do chọn đề tài

Dựa theo bối cảnh nêu trên, việc xây dựng 1 hệ thống hỗ trợ bác sĩ phân loại ảnh mẫu sinh thiết theo thang chuẩn Bethesda bằng máy tính (CAD) có thể giúp:

- Giảm tải áp lực công việc cho các bác sĩ
- Hạn chế việc chẩn đoán quá mức với bệnh nhân, tránh phẫu thuật chi phí tốn kém và hệ lụy sức khỏe về sau

1.2 Mục tiêu nghiên cứu

Nghiên cứu này có mục tiêu chính là xây dựng hệ thống chẩn đoán hỗ trợ bằng máy tính (CAD) cho các bác sĩ trong việc khám lâm sàng mẫu sinh thiết tuyến giáp hút bằng kim nhỏ.

Để đạt được mục tiêu trên các công việc, bài toán cần giải quyết là:

- Xác định được vùng quan trọng trong ảnh tự động để giảm thời gian quan sát của bác sĩ cũng như giúp thuật toán phân loại ở mục tiêu 2 có kết quả dự đoán nhanh hơn
- Phân nhãn được ảnh sinh thiết thuộc vào mức độ tương ứng dựa theo thang chuẩn Bethesda (6 nhãn - tuy nhiên do tính chất của bộ dữ liệu nên mục tiêu hiện tại của nghiên cứu là phân loại 3 nhãn)
- Kết hợp 2 mục tiêu đầu để xây dựng được hệ thống máy hỗ trợ

1.3 Cấu trúc phần còn lại của báo cáo

- Tổng quan tài liệu: Phần này xem xét và tóm tắt các tài liệu liên quan đến đề tài nghiên cứu. Bao gồm trình bày các công trình nghiên cứu trước đây, các khía cạnh lý thuyết và phương pháp đã được sử dụng, và các kết quả đã đạt được trong lĩnh vực tương tự
- Mô tả bộ dữ liệu: Cung cấp thông tin chi tiết về bộ dữ liệu được sử dụng trong nghiên cứu. Nó mô tả các nguồn dữ liệu, quy trình thu thập và quá trình tiền xử lý dữ liệu
- Phương pháp triển khai: Chi tiết về phương pháp và quy trình được sử dụng trong nghiên cứu
- Kết quả: Cung cấp thông tin chi tiết về các kết quả quan trọng
- Thảo luận: Giải thích và phân tích kết quả nghiên cứu, liên kết với các kiến thức nền tảng và tài liệu tham khảo. Cung cấp một cái nhìn sâu hơn về ý nghĩa của kết quả, so sánh với các nghiên cứu trước đây, và thảo luận về các khía cạnh quan trọng và hạn chế của nghiên cứu.
- Kết luận: Tóm tắt kết quả chính của nghiên cứu và rút ra các kết luận quan trọng. Cung cấp một cái nhìn tổng quan về đóng góp của nghiên cứu, nhấn mạnh các điểm mạnh và yếu của nghiên cứu, và đề xuất các hướng nghiên cứu tiếp theo.
- Tài liệu tham khảo

2 Tổng quan tài liệu

Savala và cộng sự [10] đã sử dụng mạng thần kinh nhân tạo (ANN) để phân biệt 2 loại ung thư tuyến giáp là FA và FTC. Kết quả phân loại thu được có độ chính xác 78%, độ nhạy 75% và độ đặc hiệu là 80%. Mặc dù nghiên cứu làm việc với ảnh WSI - tức ảnh mô học của toàn bộ sinh thiết, chứ không phải ảnh tế bào học của nghiên cứu này. Xong đây cũng là bằng chứng về sự hiệu quả của các phương pháp học sâu so với các phương pháp học máy truyền thống cho tác vụ xử lý ảnh y tế. Hạn chế hiện tại của nghiên cứu này là mới chỉ phân loại nhãn đầu ra là ung thư FA hoặc ung thư FTC. Ngoài ra kết quả trên tập kiểm tra và tập kiểm thử có chênh lệch lớn dù phân phối của 2 tập là không quá khác biệt.

Sanyal và cộng sự [11] đã sử dụng mạng tích chập (CNN) để phân loại ung thư tuyến giáp loại PTC và không phải PTC. CNN đã đạt được mức độ độ chính xác 85,1%, độ nhạy 90,5% và độ đặc hiệu 83,3%. Hạn chế của nghiên cứu vẫn là phân loại nhị phân và bộ dữ liệu còn nhỏ.

Quan và cộng sự [12] đã sử dụng CNN dựa trên VGG16 thông qua học chuyển tiếp để phân biệt ung thư tuyến giáp loại PTC với các nốt tuyến giáp lành tính. Dữ liệu gồm 279 tiêu bản tế bào được nhuộm H&E (159 ung thư PTC, 120 lành tính). Mỗi hình ảnh lớn hơn được phân chia thủ

công thành nhiều hình ảnh nhỏ hơn kích thước 224x224. Từ đó thu được 887 hình ảnh (476 PTC, 411 lành tính) và tập dữ liệu này đã được tăng cường thêm 8 lần thông qua các phép lật và phép quay. CNN đã đạt được độ chính xác 97,7%, độ nhạy 100% và độ đặc hiệu 94,9% trên ứng độ từng ảnh cắt và độ chính xác 95% ở mức độ bệnh nhân. Hạn chế nghiên cứu khi chỉ phân biệt ung thư PTC với u lành tính. Ngoài ra cách trích xuất thủ công từng vùng quan tâm của ảnh mặc dù có thể giúp hạn chế nhiều trong dữ liệu nhưng sẽ gây khó khi muốn tự động hóa quá trình trích xuất vùng quan tâm trong ảnh.

Lee và các cộng sự [13] sử dụng CNN dựa trên mạng DenseNet cho cả ảnh màu và ảnh thang màu xám. Độ chính xác của bộ phân loại sử dụng ảnh màu, ảnh RI, và cả hai lần lượt là 98,0%, 98,0% và 100%. Kết quả thu được thực sự ấn tượng nhưng bộ dữ liệu hạn chế, chỉ bao gồm 887 ảnh cắt. Mặt khác dữ liệu đầu vào một lần nữa vẫn là do các bác sĩ thủ công chọn ra các cụm tế bào quan tâm để khử nhiễu. Đầu ra bài toán vẫn chỉ là phân loại nhị phân nên khó ứng dụng cho các tác vụ chẩn đoán y tế lâm sàng.

Nhìn chung, các nghiên cứu theo hướng sử dụng học sâu đối với ảnh y tế đang giúp hỗ trợ các bác sĩ rất tốt [14]. Đối với lĩnh vực xử lý ảnh tế bào học từ kim hút nhỏ (FNAB), nhiều nghiên cứu đã được triển khai xong vẫn còn tồn tại 1 số hạn chế bao gồm:

- Trích rút vùng quan tâm của ảnh để giảm nhiễu vẫn thực hiện thủ công
- Bộ dữ liệu nhỏ
- Phân loại nhị phân

Như vậy với mục tiêu trích xuất vùng ảnh quan trọng tự động, bộ dữ liệu đa dạng cùng mục tiêu phân loại đa nhãn theo thang chuẩn Bethesda; nghiên cứu này kỳ vọng sẽ khắc phục các hạn chế của các nghiên cứu hiện nay.

3 Mô tả bộ dữ liệu

Bộ dữ liệu này gồm các ảnh sinh thiết tế bào được lấy từ các mẫu bệnh phẩm bằng phương pháp chọc hút kim tiêm nhỏ. Phương pháp này được sử dụng phổ biến trong việc chẩn đoán ung thư do ít xâm lấn và dễ thực hiện.

3.1 Nguồn gốc dữ liệu

Các ảnh được thu thập tại Bệnh viện Trung ương Quân đội 108 khi thực hiện chọc hút tế bào để chẩn đoán ung thư tuyến giáp.

3.2 Số lượng và đặc điểm ảnh

Tổng số ảnh trong bộ dữ liệu là 1421 ảnh bao gồm 3 nhãn tạm ký hiệu B2, B5, B6. Ba nhãn này lần lượt tương ứng với lành tính, nghi ngờ ác tính (SUS) và ác tính trong thang chuẩn Bethesda [9]. Số lượng ảnh tương ứng với các nhãn B2, B5, B6 lần lượt là 103, 541, 777. Định dạng file là JPEG độ phân giải 1224x960 điểm ảnh với nhãn B6 và 1024x768 điểm ảnh với các nhãn B2, B5.

3.3 Cách truy cập và sử dụng

Hiện bộ dữ liệu đang đóng truy cập do tính chất bảo mật. Cách thức truy cập và sử dụng có thể được thay đổi, cập nhật cùng các kết quả mới của nghiên cứu này trong tương lai.

3.4 Cách thức thu thập

Ảnh được thu thập theo phương pháp chọc hút tế bào (FNAB), cụ thể:

- Đầu tiên bệnh nhân sau khi siêu âm được bác sĩ đưa hướng dẫn thực hiện lấy sinh thiết tế bào để xem xét về khả năng phẫu thuật
- Bác sĩ dùng ống kim nhỏ chọc hút liên tục vào vùng nghi ngờ ung thư dưới sự hỗ trợ của việc siêu âm
- Các tế bào sau đó được đặt trên 1 lát kính, loại bỏ máu và vật chất tế bào không quan trọng
- Tế bào trên lát kính sau đó được thực hiện nhuộm màu H&E (Hematoxylin và Eosin). Trong đó:
 - Hematoxylin: Là chất nhuộm có tính kiềm, nhuộm nhân tế bào thành màu xanh tím
 - Eosin: Là chất nhuộm có tính axit, nhuộm chất nền ngoại bào và bào tương thành màu hồng. Các cấu trúc khác có thể nhuộm thành các sắc thái, tông màu và sự kết hợp khác nhau của màu xanh tím và màu hồng
- Sau khi nhuộm màu, lát kính được đặt dưới kính hiển vi và được chụp ảnh

4 Phương pháp nghiên cứu

4.1 Giới thiệu

Để giải quyết các mục tiêu của nghiên cứu, 2 bài toán được đặt ra là:

- Cắt ra vùng quan tâm của ảnh:

- Đầu vào là 1 ảnh kích thước 1024x768 chụp mẫu sinh thiết tế bào dưới kính hiển vi
- Đầu ra là 1 tập hợp các ảnh cắt (kích thước 256x256 hoặc 224x224 hoặc tùy chỉnh) chứa vùng tế bào quan trọng của ảnh đầu vào
- Bài toán được kỳ vọng giúp giảm thiểu vùng nhiễu, ít thông tin liên quan cho mục đích đánh giá loại mẫu sinh thiết. Từ đó giúp mô hình phân loại được xây dựng ở bài toán 2 hoạt động hiệu quả hơn, giúp các bác sĩ tốn ít thời gian hơn khi tập trung vào vùng quan trọng của ảnh
- Phương pháp đề xuất trong khuôn khổ nghiên cứu này là sử dụng 1 mô hình bao gồm các lớp tích chập và tùy chỉnh. Mô hình sẽ xác định hoặc phân đoạn từng tế bào riêng lẻ, kết hợp giải thuật phân cụm để cắt chọn ra được vùng quan tâm chứa các cụm tế bào
- Phân nhãn ảnh sinh thiết:
 - Đầu vào là 1 ảnh, có thể là nguyên ảnh chụp mẫu sinh thiết - tức mức độ trượt (slide level) hoặc ảnh cắt chứa vùng quan tâm - tức mức độ bản vá (patch level)
 - Đầu ra là 1 vector chứa xác suất ảnh thuộc về các nhãn, như trong khuôn khổ của bộ dữ liệu và nghiên cứu sẽ là 3 - B2, B5, B6
 - Bài toán được kỳ vọng giúp giảm gánh nặng đánh giá 1 số lượng lớn ảnh cho các bác sĩ, cung cấp các chẩn đoán gợi ý với độ chính xác cao cho tác vụ phân loại ảnh
 - Phương pháp đề xuất trong khuôn khổ nghiên cứu này là sử dụng mạng tích chập (CNN) để tự động trích xuất đặc trưng phân loại của hình ảnh. Bằng kỹ thuật học chuyển tiếp từ mô hình đã được huấn luyện trên bộ dữ liệu cực lớn (pretrained model) để giảm số lượng trọng số cần cập nhật. Qua đó khắc phục được điểm yếu là dữ liệu không đạt hàng triệu ảnh cho tác vụ học tập với số lượng lớn trọng số. Mà mặt khác, đồng thời vẫn có thể thu được kết quả cao

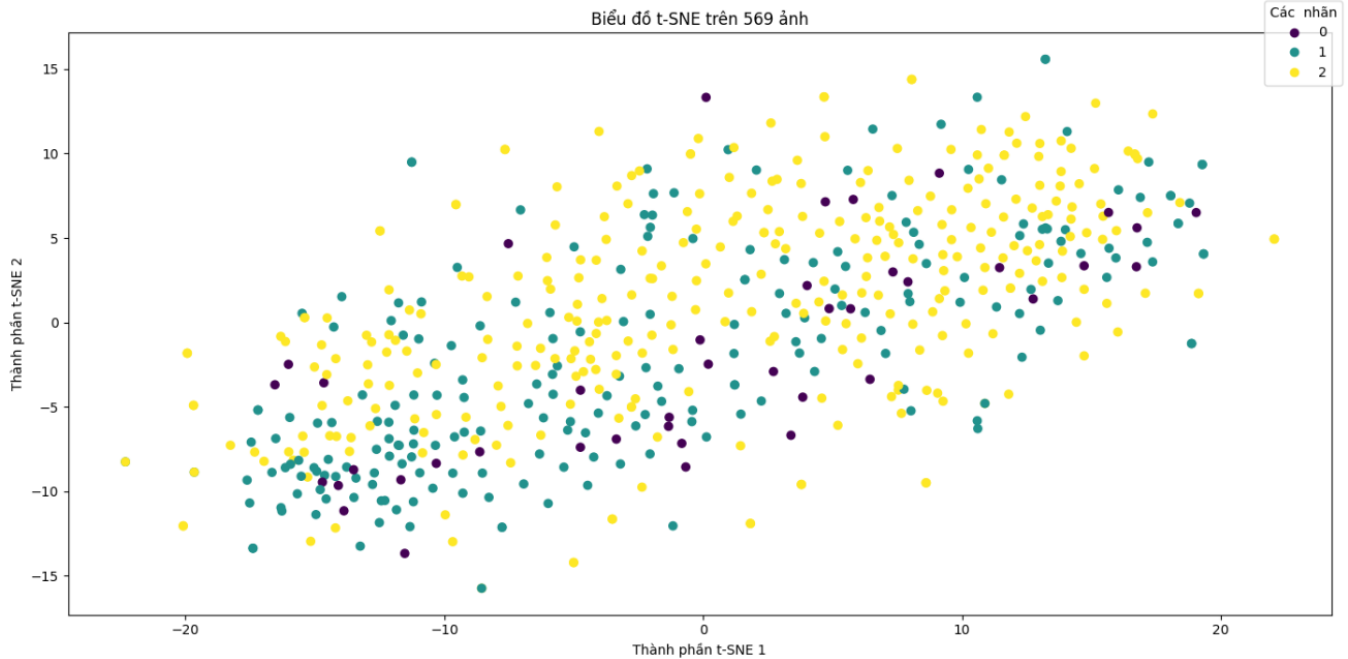
4.2 Bài toán cắt ra vùng quan tâm của ảnh

4.2.1 Giả thuyết đặt ra và ý nghĩa bài toán

Hình 1 giúp trực quan 1 phần bộ dữ liệu với việc giảm chiều tSNE. Cách giảm chiều này giúp giữ nguyên tương quan mối quan hệ (khoảng cách) giữa các ảnh với nhau [15].

Dễ nhận thấy trong hình trên, các nhãn B2 - 0 phân bố gần như được bao trùm bởi các nhãn B5 - 1. Và tương tự các nhãn B5 - 1 nằm trong miền phân phối của nhãn B6 - 2. Điều này đặt ra 1 giả thuyết là các nhiễu (thành phần vật chất tế bào, máu, bụi bẩn, ...) trong ảnh đã giữ các quan hệ giữa các ảnh gần nhau, khó tách biệt rõ ràng.

Để kiểm định giả thuyết nêu trên, nghiên cứu này sẽ tiến hành thử nghiệm so sánh việc sử dụng nguyên dữ liệu gốc với kết quả của bài toán cắt ảnh.



Hình 1: Trực quan hóa bộ dữ liệu trên 569 ảnh

4.2.2 Các bước thuật toán

Dưới đây là đề xuất thuật toán của nghiên cứu này để giúp xác định ra vùng quan trọng trong ảnh:

- Bước 1: Ảnh đầu vào qua 1 mô hình xác định hoặc phân đoạn từng tế bào riêng lẻ, từ đây có được các hình chữ nhật bao quanh tế bào (mô hình xác định tế bào riêng lẻ) hoặc các mặt nạ phân đoạn tế bào (mô hình phân đoạn tế bào)
- Bước 2: Từ tập hợp các điểm tâm (hình chữ nhật bao quanh hoặc các mặt nạ) sẽ tiến hành giải thuật phân cụm để tìm ra n (ví dụ 5) tâm cụm chứa mật độ tế bào cao. Một số giải thuật có thể sử dụng để tìm các tâm cụm mong muốn như là phân cụm K-means, phân cụm trộn Gaussian (GMM), phân cụm theo mật độ (DBSCAN), phân cụm phân cấp. Từ n tâm cụm thu được cắt ra n hình vuông kích thước mong muốn (256x256 hoặc 224x224). Giả sử thuật toán không xác định được tâm cụm có thể trượt trên từng lưới của hình chữ nhật bao quanh cụm để cắt ra từng ảnh
- Bước 3: Từ các ảnh được cắt ra, tiến hành sử dụng xác định lại số lượng tế bào trên 1 ảnh cắt, nếu số lượng là nhỏ hơn 1 ngưỡng cụ thể thì tiến hành loại bỏ ảnh khỏi tập cắt

Sau 3 bước nêu trên sẽ tạo ra 1 bộ dữ liệu đã được tiền xử lý và chọn ra được vùng quan trọng của ảnh. Có thể thấy để thuật toán cho bài toán đầu tiên này được kết quả tốt nhất thì phụ thuộc rất lớn vào:

- Mô hình xác định hoặc phân đoạn tế bào

- Kết quả giải thuật phân cụm để chọn ra vùng chứa cụm tế bào quan tâm

Tiếp sau đây sẽ là phần trình bày chi tiết về cách nghiên cứu này triển khai 2 vấn đề nêu trên.

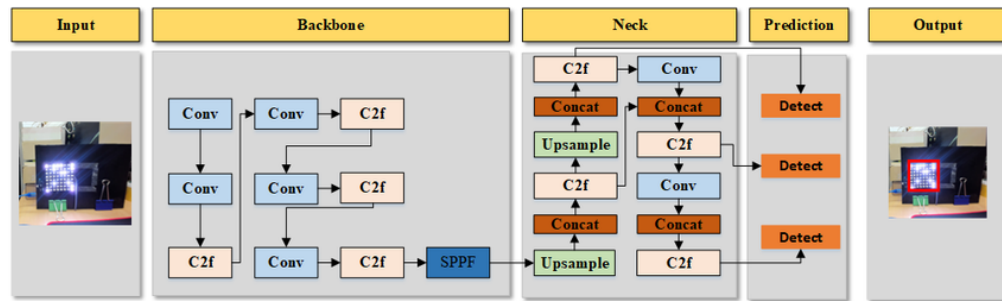
4.2.3 Mô hình xác định/phân đoạn tế bào

Mục tiêu của các mô hình này là giúp xác định tâm từng tế bào riêng lẻ để phục vụ cho thuật toán phân cụm được dùng về sau. Bởi vậy nên các mô hình giúp xác định hình chữ nhật bao quanh hay mặt nạ phân đoạn từng tế bào đều có thể được sử dụng. Dưới đây là 2 cách triển khai được nghiên cứu của em thực hiện:

- Mô hình xác định đối tượng (detect): Sử dụng sử dụng kiến trúc mạng nơ-ron tích chập (CNN) để học tính năng và phân lớp hình ảnh. CNN giúp trích xuất các đặc trưng để phân biệt từ hình ảnh, từ đó phát hiện và nghiệp vụ phân loại đối tượng. Một số mô hình nổi tiếng trong lĩnh vực này bao gồm: YOLO, SSD, Faster R-CNN... Chúng thực hiện phát hiện đa vật thể đồng thời, dự đoán trực tiếp vị trí và loại đối tượng. Trong khuôn khổ nghiên cứu này, em sử dụng YOLO cho tác vụ do sự dễ dàng triển khai, có công cụ gán nhãn nhanh chóng và thuật toán chạy nhanh (cho cả tác vụ huấn luyện và chạy dự báo).
- Mô hình phân đoạn đối tượng (segment): Khác với xác định đối tượng chỉ dự đoán vị trí, mô hình phân đoạn đối tượng sẽ phân lớp từng điểm ảnh trong hình ảnh, xác định chúng thuộc về đối tượng nào. Các mô hình thường được sử dụng bao gồm:
 - U-Net là một trong những mô hình phân đoạn ảnh đầu tiên và phổ biến nhất, sử dụng kiến trúc encoder-decoder
 - Mask R-CNN phát triển từ Faster R-CNN và hỗ trợ tác vụ phát hiện và phân đoạn đồng thời
 - DeepLab sử dụng atrous convolution để tăng độ phân giải không gian
 - YOLACT là sự kết hợp YOLO và Mask R-CNN cho kết quả phân đoạn nhanh và chính xác
 - YOLO-Unet là sự kết hợp YOLO kiến trúc cơ bản của YOLO với cách tiếp cận encoder-decoder của mạng U-Net để thực hiện cả hai nhiệm vụ phát hiện và phân đoạn trong cùng một mô hình

Nhìn chung, các thử nghiệm với các mô hình đã được triển khai sẵn cho tác vụ phân đoạn có hiệu suất tệ và lâu - nhận thấy rõ ràng thông qua kết quả dự đoán trực tiếp với dữ liệu của nghiên cứu này. Mặt khác, em đã tiến hành gán nhãn phân đoạn cho 30 ảnh để huấn luyện, thử nghiệm và kiểm tra. Kết quả cho ra tương đối ổn nhưng việc gán nhãn vẫn quá tốn kém.

Kết luận sự hạn chế về nguồn lực và thời gian của nghiên cứu hiện tại, cùng với sự khó khăn hơn



Hình 2: Kiến trúc mạng YOLO version 8 cho tác vụ xác định đối tượng

về việc gán nhãn dữ liệu (gán nhãn phân đoạn phức tạp và tốn thời gian hơn so với tác vụ xác định đối tượng). **Bởi vậy mô hình được sử dụng ở nghiên cứu này sẽ là mô hình xác định từng tế bào riêng lẻ.**

Hình 2 đến từ bài báo [16]. Kiến trúc YOLOv8 kế thừa cấu trúc cơ bản của thuật toán YOLO ban đầu với các khối CNN để nhận dạng tính năng từ ảnh gốc. Cụ thể:

- Đầu vào là 1 hình ảnh
- Mô-đun trích xuất đặc trưng (Backbone), bao gồm các khối Conv (Convolutional Layer), C2f và SPPF.
 - Các khối Conv được sử dụng để áp dụng các bộ lọc lên hình ảnh đầu vào, giúp trích xuất các đặc trưng như cạnh, góc, và kết cấu
 - Các khối C2f (Cross Stage Partial Networks - CSPNet) giúp giảm bớt thông tin dư thừa và cải thiện hiệu suất bằng cách chia đầu vào thành hai phần, một phần qua mạng con và một phần bỏ qua mạng con, sau đó kết hợp lại
 - SPPF (Spatial Pyramid Pooling - Fast) là một lớp pooling được thêm vào để tăng cường khả năng trích xuất đặc trưng ở nhiều tỉ lệ không gian khác nhau
- Mô-đun Neck tổng hợp các đặc trưng từ Backbone và chuẩn bị chúng cho quá trình dự đoán.
 - Nó bao gồm các lớp C2f, Concat (Concatenation), Upsample (phóng đại kích thước), và Conv
 - Các đặc trưng từ nhiều mức độ khác nhau được kết hợp lại thông qua lớp Concat, sau đó trải qua các lớp C2f và Conv để tiếp tục trích xuất và tinh chỉnh đặc trưng
 - Lớp Upsample giúp tăng kích thước của các đặc trưng để đồng bộ với các đặc trưng từ các tầng khác, giúp kết hợp thông tin từ nhiều độ phân giải khác nhau
- Mô-đun Dự đoán bao gồm các khối Detect, nơi mà các đặc trưng cuối cùng được sử dụng để dự đoán các bounding box và nhãn của các đối tượng trong hình ảnh. Mỗi lớp Detect sẽ dự

đoán các bounding box ở các kích thước khác nhau, giúp phát hiện các đối tượng ở nhiều tỉ lệ khác nhau trong hình ảnh

- Đầu ra là các hình chữ nhật bao quanh đối tượng (bounding box) và nhãn thể hiện các đối tượng đã được phát hiện

Kiến trúc của YOLOv8 kết hợp giữa các phương pháp tiên tiến trong học sâu để đạt được hiệu suất cao trong bài toán phát hiện đối tượng, cân bằng giữa độ chính xác và tốc độ.

4.2.4 Thuật toán phân cụm để chọn ra những cụm tối ưu

Bốn giải thuật phân cụm phổ biến thường là K-means, phân cụm trộn Gaussian (GMM), phân cụm theo mật độ (DBSCAN), phân cụm phân cấp. Đánh giá hiệu quả của 4 giải thuật với dữ liệu cụ thể đã được [17] thực hiện. Theo đó:

- K-means là thuật toán phân cụm phổ biến, dễ triển khai. Tuy nhiên phụ thuộc mạnh vào số lượng cụm k ban đầu
- GMM sử dụng phân bố Gauss để mô tả cụm, phù hợp với dữ liệu có cấu trúc phức tạp hơn. Chịu ảnh hưởng ít từ số lượng cụm
- DBSCAN xác định cụm dựa mật độ điểm, phù hợp với dữ liệu có cụm không đều, dựa trên mật độ điểm xung quanh. Không yêu cầu biết số lượng cụm trước
- Phân cụm phân cấp thì xây dựng cấu trúc phân cấp cụm phức tạp hơn các phương pháp trên

Do tính chất của dữ liệu (tế bào phân bố tập trung vào các vùng theo mật độ), kiểu phân cụm theo mật độ DBSCAN được đề xuất sử dụng. Các phương pháp phân cụm khác sẽ được nghiên cứu để đánh giá về mức độ cải thiện thêm về sau.

Về cơ bản, DBSCAN hoạt động như sau:

- Chọn điểm kiểm tra x , kiểm tra xem nó có đủ n điểm láng giềng cạnh giềng trong bán kính ϵ không
- Nếu có, x là điểm lõi và tất cả các điểm láng giềng cũng thuộc cùng một cụm
- Nếu không, tiếp tục kiểm tra các điểm láng giềng của x nếu chúng có đủ n điểm láng giềng
- Các điểm kết nối với điểm lõi cũng thuộc cùng cụm
- Cụm được mở rộng cho đến khi không còn điểm mới nữa
- Tâm mỗi cụm được tính chính là trung bình của tất cả các điểm trong cụm

Ưu điểm của DBSCAN là có thể phát hiện cụm có hình dạng bất kỳ, không cần biết trước số cụm. Rất phù hợp với bài toán phân cụm tế bào theo mật độ. Sau khi đã có các cụm, ta có thể triển khai 2 hướng cắt ảnh:

- Cách 1 với mỗi cụm ta cắt ra được 1 ảnh con tương ứng với tâm cụm là tâm ảnh
- Cách 2 với mỗi cụm ta lấy ra hình chữ nhật nhỏ nhất (bouding box) bao quanh cả cụm. Thực hiện trượt trên lưới để thu được các ảnh cắt. Như vậy với một cụm có thể cắt ra được nhiều ảnh

Từ các ảnh cắt ra kiểm tra lại ngưỡng tế bào tối thiểu cần có của mỗi ảnh, qua đó loại bỏ những ảnh không đảm bảo điều kiện.

4.3 Bài toán phân nhãn ảnh sinh thiết

Bài toán phân nhãn ảnh sinh thiết là bước quan trọng trong chẩn đoán ung thư bằng cách phân loại tế bào thành các nhóm: ung thư, lành tính, viêm... Ở trường hợp cụ thể của nghiên cứu này, số nhãn cho output sẽ là 3 trong 6 nhãn của thang đo chuẩn Bethesda.

Do tính chất phức tạp và đa dạng của các loại tế bào, phân nhãn ảnh sinh thiết là một nhiệm vụ học sâu. Phương pháp thông dụng hiện nay là sử dụng mạng nơ-ron tích chập (CNN) nhờ khả năng trích xuất các đặc trưng nổi bật từ hình ảnh. Các mạng CNN thường được sử dụng gồm có VGG, ResNet, Inception, EfficientNet, ...

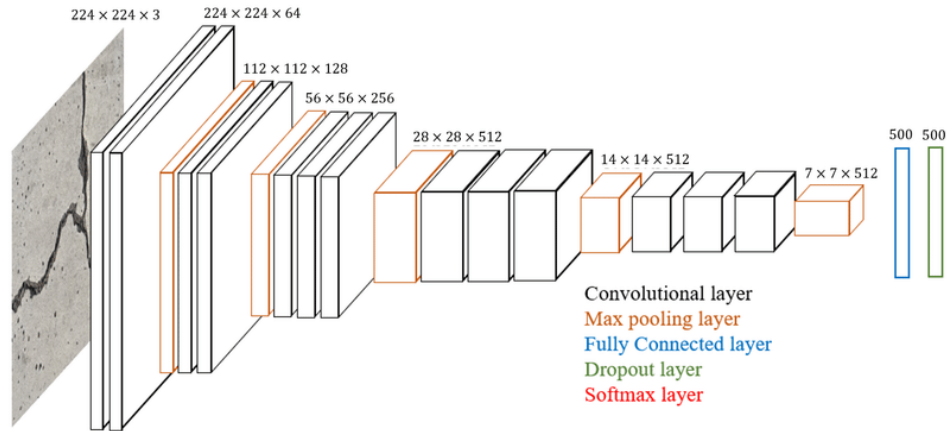
Sau kết quả nghiên cứu đạt được từ giữa kỳ, nhìn chung kết quả 3 mạng được thử nghiệm (VGG19, EfficientNetB0, ResNet50) không khác biệt quá đáng kể, độ chính xác đạt được thời điểm đó cao nhất là 0.66. Lý do cho vấn đề này do tài nguyên tính toán có hạn, việc giảm độ sâu và độ rộng của các tầng kết nối đầy đủ (dense layer) ở cuối của mạng giúp tăng tốc độ tính toán nhưng cũng đồng thời khó ước lượng được siêu phẳng phân loại phù hợp với dữ liệu.

Với vấn đề nêu trên, định hướng nghiên cứu tập trung vào cải thiện kiến trúc 1 mạng là VGG16 - kiến trúc đơn giản, dễ thay đổi trọng số để phù hợp với dữ liệu.

Kiến trúc được mô tả trong hình 3 tham khảo từ [18] là một phiên bản của mô hình VGG16, thường được sử dụng cho bài toán phân loại hình ảnh. Chi tiết kiến trúc:

- Tầng đầu vào kích thước $224 \times 224 \times 3$. Nhận hình ảnh đầu vào với kích thước chuẩn hóa 224×224 pixel và 3 kênh màu (RGB)
- Tầng tích chập gồm các lớp tích chập (Conv) có vai trò trích xuất đặc trưng từ hình ảnh đầu vào. Mỗi lớp tích chập được theo sau bởi một lớp phi tuyến (thường là ReLU) để tăng cường khả năng học của mạng. Các kích thước lần lượt là

– $224 \times 224 \times 64$



Hình 3: Kiến trúc chung của 1 mạng VGG16

- $112 \times 112 \times 128$
 - $56 \times 56 \times 256$
 - $28 \times 28 \times 512$
 - $14 \times 14 \times 512$
 - $7 \times 7 \times 512$
- Tầng hợp nhất (Max Pooling) để giảm kích thước của các đặc trưng, giúp giảm bớt số lượng tham số và kiểm soát hiện tượng overfitting. Các lớp này lấy giá trị lớn nhất trong một cửa sổ (thường là 2×2) và di chuyển cửa sổ này trên toàn bộ đặc trưng. Các lớp này xếp ngay sau các lớp tích chập nêu trên với các kích thước lần lượt là
 - Sau Conv 64: $112 \times 112 \times 64$
 - Sau Conv 128: $56 \times 56 \times 128$
 - Sau Conv 256: $28 \times 28 \times 256$
 - Sau Conv 512: $14 \times 14 \times 512$
 - Sau Conv 512 (lần 2): $7 \times 7 \times 512$
 - Tầng kết nối đầy đủ (Fully Connected Layers) để kết nối toàn bộ các đơn vị đầu vào với các đơn vị đầu ra, giúp mạng học các đặc trưng phi tuyến từ dữ liệu đầu vào. Số lượng nơ-ron ở mỗi tầng lần lượt là (4096, 4096, 1000), với (4096, 4096) là số nơ-ron mặc định ở tầng ẩn, 1000 là số nhãn đầu ra mặc định trong VGG16 gốc
 - Tầng dropout để giảm hiện tượng overfitting bằng cách ngẫu nhiên loại bỏ một số kết nối trong mạng trong quá trình huấn luyện

- Và cuối cùng là tầng softmax sử dụng 1 hàm softmax để đưa đầu ra về giá trị nằm trong khoảng (0, 1) - tương ứng với giá trị xác suất thuộc về 1 trong các nhãn đầu ra.

Như vậy việc cần điều chỉnh ở đây chính là thay đổi số lượng nơ-ron ở các tầng kết nối đầy đủ và 1 số tầng gần cuối của mạng. Hiệu quả của việc này sẽ cao hơn việc huấn luyện lại toàn bộ mạng do số lượng tham số lớn với mạng quá rộng và sâu và dữ liệu không quá nhiều tới hàng triệu.

VGG16 có ưu điểm như sau

- Đơn giản và hiệu quả: Kiến trúc VGG16 khá đơn giản với các lớp tích chập 3x3 và các lớp pooling 2x2, dễ dàng để triển khai và tối ưu hóa
- Khả năng tổng quát tốt: Kiến trúc này đã được chứng minh là có khả năng tổng quát tốt trên nhiều bộ dữ liệu khác nhau
- Sâu và rộng: Với độ sâu của nó, VGG16 có thể trích xuất được nhiều đặc trưng phức tạp từ hình ảnh

Tuy nhiên bên cạnh đó vẫn còn 1 số nhược điểm

- Chi phí tính toán cao: Do có rất nhiều tham số và tầng, VGG16 yêu cầu rất nhiều tài nguyên tính toán, đặc biệt là khi xử lý các hình ảnh lớn
- Overfitting: Dù có các lớp Dropout, VGG16 vẫn có nguy cơ overfitting nếu không có đủ dữ liệu huấn luyện hoặc không được tối ưu hóa đúng cách
- Không linh hoạt trong xử lý kích thước đầu vào: Kiến trúc này yêu cầu kích thước đầu vào cố định, do đó cần phải tiền xử lý hình ảnh để phù hợp với kích thước yêu cầu

Kiến trúc cải tiến của VGG16 có thể bao gồm thêm một số thay đổi nhằm tối ưu hóa hiệu suất và giảm bớt chi phí tính toán, nhưng nhìn chung vẫn giữ nguyên các nguyên tắc cơ bản của VGG16.

5 Kết quả

5.1 Các thang đo đánh giá

5.1.1 Thước đo đánh giá mô hình xác định tế bào

- **Độ nhạy** đánh giá tỷ lệ phát hiện thành công trên tổng số đối tượng thực tế. Đo khả năng bỏ lọt sót mất đối tượng:

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

Trong đó:

- TP là số lượng đối tượng được phát hiện chính xác.
- FN là số lượng đối tượng bị bỏ qua.

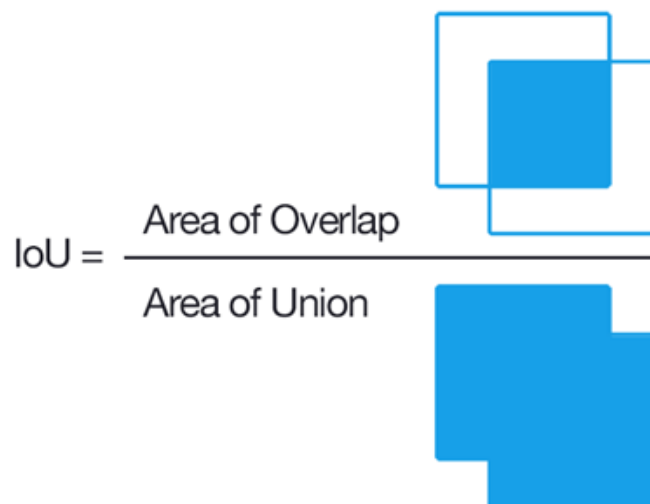
- **Độ chính xác dương (Precision)** đánh giá tỷ lệ phát hiện chính xác trên tổng số phát hiện. Đo lường lỗi sai:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Trong đó:

- TP là số lượng đối tượng được phát hiện chính xác.
 - FP là số lượng phát hiện sai (false positive).
- **Intersection over Union (IoU)** đo tỷ lệ phần giao nhau của hộp dự đoán và thật so với phần hợp. Đánh giá sự chính xác vị trí:

$$IoU = \frac{VungGiao}{VungHop} \quad (3)$$



Hình 4: Cách tính IoU

- **Độ chính xác trung bình - Mean Average Precision (mAP)** đánh giá độ chính xác trung bình, đo lường khả năng phát hiện và xác định vị trí đối tượng.

Như vậy, 4 độ đo này giúp đánh giá khía cạnh của mô hình xác định đối tượng. Tuy nhiên, **nhìn chung thông số đáng quan tâm nhất vẫn là mAP** do đầu ra mô hình cần quan tâm là ảnh có đối tượng gì và đối tượng đó ở đâu trên ảnh.

Nói chi tiết hơn về mAP, mAP được định nghĩa là trung bình của giá trị Average Precision (AP) cho tất cả các lớp. Average Precision (AP) cho một lớp được tính như sau:

$$AP = \sum_{k=1}^n (P(k) \cdot \Delta r(k)) \quad (4)$$

Trong đó:

- n là số lượng tài liệu được truy xuất.
- $P(k)$ là độ chính xác tại vị trí cắt k .
- $\Delta r(k)$ là sự thay đổi của độ phủ tại vị trí cắt k .

Giá trị mAP được tính là trung bình của các giá trị AP cho tất cả Q lớp:

Trong đó:

- Q là số lượng lớp - nhãn đối tượng.
- AP_q là giá trị Average Precision cho lớp đối tượng thứ q .

5.1.2 Thước đo đánh giá giải thuật phân cụm

Để đánh giá hiệu quả phân cụm, độ đo phổ biến là dùng **độ đo Silhouette** có giá trị từ -1 đến 1, đánh giá mức độ phù hợp của mỗi điểm với cụm của nó. Giá trị gần 1 có nghĩa là phù hợp tốt.

Silhouette coefficient của điểm i được tính bằng công thức:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5)$$

Trong đó:

- $a(i)$ là trung bình của khoảng cách từ điểm i đến tất cả các điểm trong cùng cụm.
- $b(i)$ là trung bình thấp nhất của khoảng cách từ i đến tất cả các điểm trong các cụm khác.

Nếu $s(i)$ gần 1 thì điểm i phù hợp với cụm; ngược lại nếu gần -1 thì không phù hợp. Silhouette Coefficient trung bình của tập dữ liệu từ -1 đến 1 đánh giá chất lượng phân cụm của thuật toán.

5.1.3 Thước đo đánh giá mô hình phân loại ảnh

1. Độ chính xác (Accuracy) Độ chính xác là tỷ lệ giữa số lượng dự đoán đúng và tổng số mẫu trong tập kiểm tra.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Trong đó:

- TP (True Positive): Số lượng mẫu thực sự là dương và được dự đoán là dương.
- TN (True Negative): Số lượng mẫu thực sự là âm và được dự đoán là âm.
- FP (False Positive): Số lượng mẫu thực sự là âm nhưng được dự đoán là dương.
- FN (False Negative): Số lượng mẫu thực sự là dương nhưng được dự đoán là âm.

2. Độ chính xác trung bình (Mean Accuracy) Độ chính xác trung bình là trung bình của độ chính xác trên tất cả các lớp.

$$Mean Accuracy = \frac{1}{C} \sum_{i=1}^C Accuracy_i \quad (7)$$

Trong đó C là số lượng lớp và $Accuracy_i$ là độ chính xác của lớp thứ i .

3. Độ nhạy (Recall) Độ nhạy (hay còn gọi là Tỷ lệ dương thật) là tỷ lệ giữa số lượng mẫu dương được dự đoán đúng và tổng số mẫu dương thực sự.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

4. Độ đặc hiệu (Specificity) Độ đặc hiệu là tỷ lệ giữa số lượng mẫu âm được dự đoán đúng và tổng số mẫu âm thực sự.

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

5. Tỷ lệ dự đoán dương (Precision) Độ chính xác (hay còn gọi là Tỷ lệ dương dự đoán) là tỷ lệ giữa số lượng mẫu dương được dự đoán đúng và tổng số mẫu được dự đoán là dương.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

6. F1 Score F1 Score là trung bình điều hòa của độ chính xác và độ nhạy, cung cấp một thước đo cân bằng giữa hai giá trị này.

$$F1 Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (11)$$

7. Đường cong ROC và AUC Đường cong ROC (Receiver Operating Characteristic) là một đồ thị biểu diễn khả năng phân loại của mô hình ở các ngưỡng phân loại khác nhau. AUC (Area Under the Curve) là diện tích dưới đường cong ROC, thể hiện tổng quát khả năng phân loại của mô hình.

8. Ma trận nhầm lẫn (Confusion Matrix) Ma trận nhầm lẫn là một công cụ hữu ích để đánh giá hiệu suất của mô hình phân loại, đặc biệt trong bài toán phân loại đa lớp. Nó cung cấp thông tin chi tiết về các lỗi phân loại mà mô hình mắc phải.

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (12)$$

9. Độ chính xác trung bình (mAP) Độ chính xác trung bình (mAP) là một thước đo đánh giá tổng quát cho các bài toán phân loại đa lớp và truy xuất thông tin. Nó được tính bằng cách lấy trung bình của các giá trị Average Precision (AP) cho tất cả các lớp hoặc truy vấn.

$$mAP = \frac{1}{Q} \sum_{q=1}^Q AP_q \quad (13)$$

Trong đó:

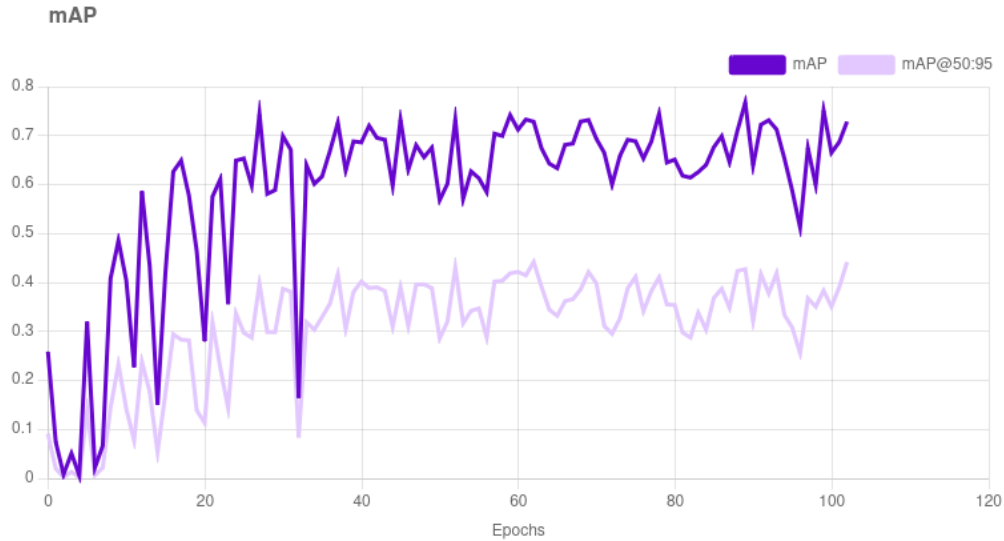
- Q là số lượng truy vấn hoặc lớp.
- AP_q là giá trị Average Precision cho truy vấn hoặc lớp thứ q .

Mặc dù có khá nhiều độ đo, xong trong các bài toán phân loại đa lớp, các độ đo như **độ chính xác (accuracy)** và **F1 Score** thường được ưu tiên vì chúng cung cấp thông tin tổng quát và cân bằng về hiệu suất của mô hình.

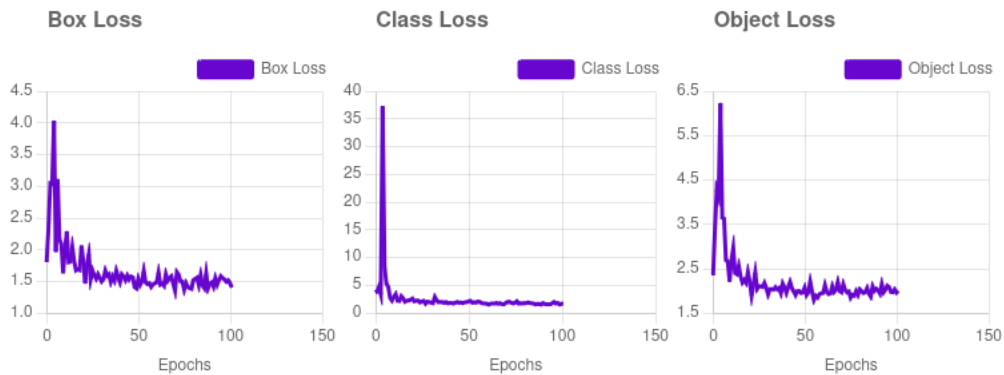
5.2 Kết quả cụ thể

5.2.1 Mô hình xác định tế bào

Kiểm tra trên tập test cho kết quả: mAP: 72.8%, precision: 87.6%, recall: 64.5%



Hình 5: Giá trị mAP của mô hình xác định tế bào trong quá trình huấn luyện



Hình 6: Đánh giá các hàm loss trong quá trình huấn luyện mô hình xác định tế bào

5.2.2 Kết quả giải thuật phân cụm

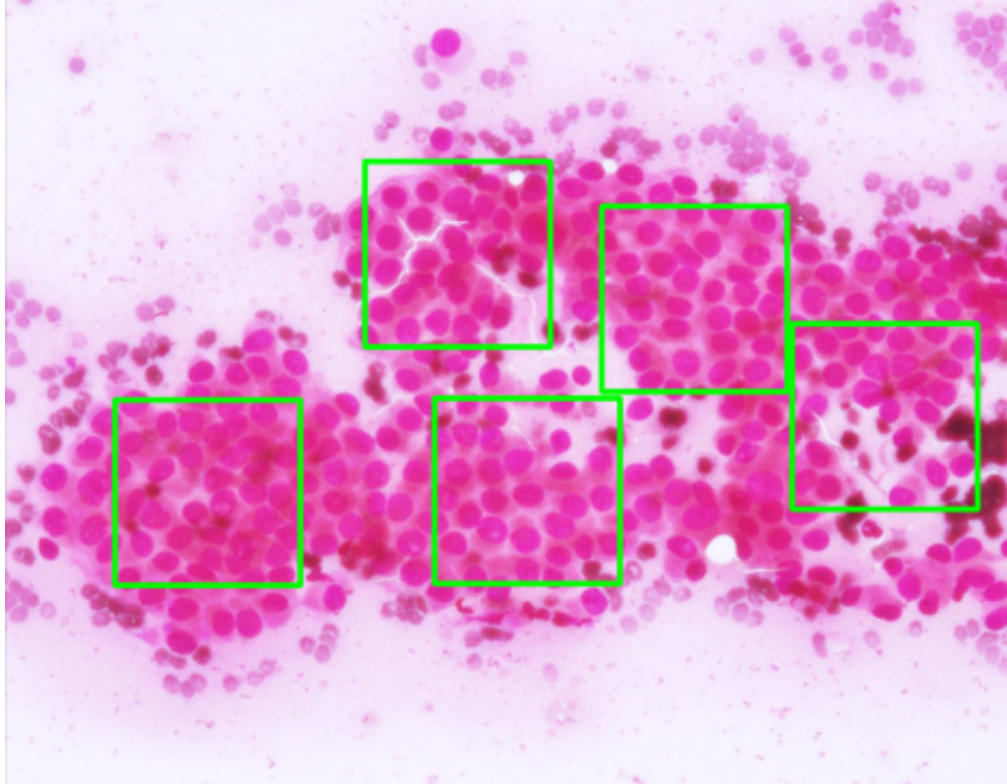
Sau khi tiến hành giải thuật phân cụm theo mật độ, chỉ số Silhouette được tính toán cho từng cụm trong từng ảnh trong tổng số 1421 ảnh.

Sau đó, giá trị Silhouette tính chung cho 1 ảnh là trung bình của các cụm có trong ảnh, thu được 1421 giá trị. Tính trung bình 1421 giá trị này ta thu được độ đo Silhouette khi phân cụm trên toàn bộ tập dữ liệu là: 0.64

5.2.3 Đánh giá kết quả ảnh cắt thu được

Sau khi đã thu được mô hình xác định đối tượng tế bào cùng giải thuật phân cụm, nghiên cứu hiện tại đã tiến hành tạo ra bộ dữ liệu mới bao gồm các ảnh cắt. thiết.

Do chênh lệch bản chất trong dữ liệu nên các nhãn B2, B5, B6 mất cân bằng nên nay lại càng mất



Hình 7: Ví dụ về cách các ảnh được chọn cắt

cân bằng khi các ảnh B5, B6 có nhiều vùng quan tâm hơn (vùng chứa nhiều tế bào). Bởi vậy để tránh mô hình qua thiên vị trong dự đoán về B5, B6; việc cân bằng dữ liệu đã được thực hiện bao gồm các phép quay, lật, phóng to, thêm nhiễu, ...

Kết quả sau khi đã cân bằng lại dữ liệu là: $[B2, B5, B6] = [3381, 3381, 3381]$ tức bao gồm 10143 ảnh. Phân chia tập huấn luyện (train), thử nghiệm (validation), kiểm tra (test) theo tỷ lệ 75:10:15.

5.2.4 Kết quả mô hình phân loại nhãn ảnh sinh thiết

Trước hết để kiểm tra lại giả thuyết đã đặt ra từ đầu về hiệu quả của việc cắt ảnh so với việc giữ nguyên ảnh gốc, nghiên cứu này đã thực hiện huấn luyện với cùng 1 kiến trúc mạng như sau:

Như vậy có thể thấy cùng 1 kiến trúc mạng nhưng ảnh cắt đã cho kết quả tốt hơn so với dùng nguyên ảnh gốc để dự đoán cho mô hình tới 10%, đạt 84% so với 74%. Nói cách khác, có thể khẳng định việc cắt ra vùng quan tâm của ảnh giúp mô hình ít bị ảnh hưởng bởi nhiễu trong dữ liệu hơn. Tiếp tục thử nghiệm với mô hình H2 cho bộ dữ liệu ảnh cắt vùng quan trọng.


```

Model architecture:
Model: "model_1"
-----
Layer (type)                Output Shape                Param #
-----
input_3 (InputLayer)        [(None, 224, 224, 3)]      0
vgg16 (Functional)          (None, 7, 7, 512)          14714688
flatten_1 (Flatten)         (None, 25088)              0
dense_3 (Dense)              (None, 4096)               102764544
dropout_2 (Dropout)         (None, 4096)               0
dense_4 (Dense)              (None, 4096)               16781312
dropout_3 (Dropout)         (None, 4096)               0
dense_5 (Dense)              (None, 3)                  12291
-----
Total params: 134272835 (512.21 MB)
Trainable params: 119558147 (456.08 MB)
Non-trainable params: 14714688 (56.13 MB)

```

Hình 8: Mô hình H1

```

Confusion Matrix
[[ 11   9   1]
 [  6  85  18]
 [  2  39 116]]

```

```

Classification Report:
      precision    recall  f1-score   support

      B2       0.58      0.52      0.55         21
      B5       0.64      0.78      0.70        109
      B6       0.86      0.74      0.79        157

   accuracy          0.74         287
  macro avg       0.69      0.68      0.68         287
 weighted avg     0.76      0.74      0.74         287

```

(a) Kết quả dùng ảnh gốc với mô hình H1

```

Confusion Matrix:
[[388  12   7]
 [ 21 309  77]
 [ 10  69 328]]

```

```

Classification Report:
      precision    recall  f1-score   support

      0       0.93      0.95      0.94         407
      1       0.79      0.76      0.78         407
      2       0.80      0.81      0.80         407

   accuracy          0.84        1221
  macro avg       0.84      0.84      0.84        1221
 weighted avg     0.84      0.84      0.84        1221

```

(b) Kết quả dùng ảnh cắt với mô hình H1

Hình 9: So sánh kết quả dùng ảnh gốc và ảnh cắt với mô hình H1

```

Model architecture:
Model: "model_2"
-----
Layer (type)                Output Shape                Param #
-----
input_7 (InputLayer)        [(None, 224, 224, 3)]      0
vgg16 (Functional)          (None, 7, 7, 512)          14714688
global_average_pooling2d_1 (GlobalAveragePooling2D) (None, 512)              0
dense_9 (Dense)              (None, 1024)               525312
dropout_6 (Dropout)         (None, 1024)               0
dense_10 (Dense)             (None, 1024)               1049600
dropout_7 (Dropout)         (None, 1024)               0
dense_11 (Dense)             (None, 3)                  3075
-----
Total params: 16292675 (62.15 MB)
Trainable params: 1577987 (6.02 MB)
Non-trainable params: 14714688 (56.13 MB)

```

Hình 10: Kiến trúc mô hình H2

Confusion Matrix:						
[[389 18 0]						
[15 338 54]						
[6 60 341]]						
Classification Report:						
			precision	recall	f1-score	support
		0	0.95	0.96	0.95	407
		1	0.81	0.83	0.82	407
		2	0.86	0.84	0.85	407
	accuracy				0.87	1221
	macro avg		0.87	0.87	0.87	1221
	weighted avg		0.87	0.87	0.87	1221

Hình 11: Kết quả mô hình H2 trên tập kiểm tra của bộ dữ liệu ảnh cắt

Kết quả thu được tăng 3%, từ 84% lên 87%.

6 Thảo luận

6.1 Về kết quả đạt được

Mô hình xác định tế bào hoạt động tốt với giá trị mAP cao 72.8% trên tập test, có hiệu quả sử dụng tốt. Với thuật toán phân cụm tính tổng thể toàn tập 1421 ảnh là 0.64, đây là ngưỡng chấp nhận được và kết quả ảnh cắt ra được đánh giá thủ công là tốt.

Với bài toán phân loại ảnh, có thể chứng minh được việc cắt ra các vùng quan trọng của ảnh thay vì dùng nguyên ảnh gốc đem lại kết quả cao hơn 10% khi kiểm tra với mô hình H1, đạt 84% so với 74%. Thực hiện giảm độ phức tạp của mô hình H1 ta thu được mô hình H2. Có thể thấy mô hình H2 cho kết quả trên tập kiểm tra cao hơn 3%, đạt 87% so với 84%. Điều này có thể do sự đơn giản hơn của mô hình đã giúp sự cập nhật trọng số của dữ liệu hiệu quả hơn.

6.2 Về mục tiêu hiện tại và phương hướng tương lai

Có thể nói 3 mục tiêu đặt ra của nghiên cứu đã đạt được, mục tiêu 1 hỗ trợ mục tiêu 2 và mục tiêu 1, 2 hỗ trợ mục tiêu 3. Việc chọn ra được các vùng quan trọng trong ảnh kết hợp với việc phân nhân tương ứng cho ảnh là sự hỗ trợ mạnh mẽ bởi máy móc (CAD) giúp các bác sĩ.

Tuy nhiên kể cả khi đã đạt được 3 mục tiêu đặt ra, nghiên cứu hiện tại vẫn còn hạn chế khi chưa đánh giá được hiệu quả của việc sử dụng nghiên cứu này để cải thiện kết quả chẩn đoán của các bác sĩ - đặc biệt là các bác sĩ thiếu kinh nghiệm và giàu kinh nghiệm.

Như vậy mục tiêu tương lai của nghiên cứu sẽ là:

- Tiếp tục cải thiện kỹ thuật cắt vùng ảnh quan trọng

- Cải thiện hiệu suất của mô hình phân loại
- Đặc biệt là nghiên cứu ảnh hưởng tích cực của kết quả nghiên cứu này đến hiệu quả chẩn đoán của các bác sĩ

7 Kết luận

Kết quả nghiên cứu hiện tại cho thấy tiềm năng cải tiến kỹ thuật học sâu (deep learning) để phân loại ảnh chọc hút tế bào bằng kim nhỏ (FNAB) đa nhân theo thang chuẩn Bethesda. Bằng việc tự động cắt chọn ra vùng quan trọng của ảnh cũng như khả năng phân loại chính xác các nhân sinh thiết là sự giúp sức hiệu quả với các bác sĩ. Đặc biệt trong việc chẩn đoán tình trạng ung thư tuyến giáp và có phương pháp chữa trị phù hợp. Việc nghiên cứu tiếp ảnh hưởng của kết quả nghiên cứu này tới hiệu quả hỗ trợ các bác sĩ sẽ là ứng dụng thực tiễn lớn.

Tài liệu tham khảo

Tài liệu

- [1] YuJiao Deng, HongTao Li, Meng Wang, Na Li, Tian Tian, Ying Wu, Peng Xu, Si Yang, Zhen Zhai, LingHui Zhou, Qian Hao, DingLi Song, TianBo Jin, Jun Lyu, and ZhiJun Dai. Global Burden of Thyroid Cancer From 1990 to 2017. *JAMA Network Open*, 3(6):e208759–e208759, Jun 2020.
- [2] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, Jan 2019.
- [3] Mengmeng Li, Juan P Brito, and Salvatore Vaccarella. Long-term declines of thyroid cancer mortality: An international age-period-cohort analysis. *Thyroid*, 30(6):838–846, Jun 2020.
- [4] Yassine Habchi, Yassine Himeur, Hamza Kheddar, Abdelkrim Boukabou, Shadi Atalla, Ammar Chouchane, Abdelmalik Ouamane, and Wathiq Mansoor. Ai in thyroid cancer diagnosis: Techniques, trends, and future directions. *Systems*, 11(10), 2023.
- [5] Esther Diana Rossi, Liron Pantanowitz, Marco Raffaelli, and Guido Fadda. Overview of the ultrasound classification systems in the field of thyroid cytology. *Cancers*, 13(13), 2021.
- [6] Vijay Vyas Vadhiraaj, Andrew Simpkin, James O’Connell, Naykky Singh Ospina, Spyridoula Maraka, and Derek T. O’Keeffe. Ultrasound image classification of thyroid nodules using machine learning techniques. *Medicina*, 57(6), 2021.
- [7] Yi-Cheng Zhu, Peng-Fei Jin, Jie Bao, Quan Jiang, and Ximing Wang. Thyroid ultrasound image classification using a convolutional neural network. *Annals of Translational Medicine*, 9(20):1526, 2021.
- [8] Aram H AlSaedi, Dalia S Almalki, and Reem M ElKady. Approach to thyroid nodules: Diagnosis and treatment. *Cureus*, 16(1):e52232, 2024.
- [9] B Evranos, S B Polat, H Baser, D Ozdemir, A Kilicarslan, A Yalcin, R Ersoy, and B Cakir. Bethesda classification is a valuable guide for fine needle aspiration reports and highly predictive especially for diagnosing aggressive variants of papillary thyroid carcinoma. *Cytopathology*, 28(4):259–267, Aug 2017.
- [10] Rajiv Savala, Pranab Dey, and Nalini Gupta. Artificial neural network model to distinguish follicular adenoma from follicular carcinoma on fine needle aspiration of thyroid. *Diagnostic Cytopathology*, 46(3):244–249, Dec 2018.

- [11] Parikshit Sanyal, Tanushri Mukherjee, Sanghita Barui, Avinash Das, and Prabaha Gangopadhyay. Artificial intelligence in cytopathology: A neural network to identify papillary carcinoma on thyroid fine-needle aspiration cytology smears. *Journal of Pathology Informatics*, 9(1):43, 2018.
- [12] Qing Guan, Yunjun Wang, Bo Ping, Duanshu Li, Jiajun Du, Qin Yu, Hongtao Lu, Xiaochun Wan, and Jun Xiang. Deep convolutional neural network vgg-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *Journal of Cancer*, 10:4876–4882, Aug 2019.
- [13] Young Lee, Dongmin Ryu, Seungwoo Kim, Juyeon Park, Seog Park, Donghun Ryu, Hayoung Lee, Sungbin Lim, Hyun-seok Min, YongKeun Park, and Eun Lee. Machine-learning-based diagnosis of thyroid fine-needle aspiration biopsy synergistically by papanicolaou staining and refractive index distribution. *Scientific Reports*, 13, Jun 2023.
- [14] Masayuki Tsuneki. Deep learning models in medical image analysis. *Journal of Oral Biosciences*, 64(3):312–320, 2022.
- [15] V. Spiwok and P. Křížová. Time-lagged t-distributed stochastic neighbor embedding (t-sne) of molecular simulation trajectories. *Frontiers in Molecular Biosciences*, 7:132, 2020.
- [16] Herfandi Herfandi, Ones Sitanggang, Muhammad Nasution, Huy Nguyen, and Yeong Min Jang. Real-time patient indoor health monitoring and location tracking with optical camera communications on the internet of medical things. *Applied Sciences*, 14:1153, Jan 2024.
- [17] Junfei Gu. Comparative analysis based on clustering algorithms. *Journal of Physics: Conference Series*, 1994(1):012024, Aug 2021.
- [18] Daegyun Choi, William Bell, Donghoon Kim, and Jichul Kim. Uav-driven structural crack detection and location determination using convolutional neural networks. *Sensors*, 21:2650, Apr 2021.