

BỘ GIÁO DỤC VÀ ĐÀO TẠO BỘ KHOA HỌC VÀ CÔNG NGHỆ
VIỆN NĂNG LƯỢNG NGUYÊN TỬ VIỆT NAM



HỒ QUANG TUẤN

**NGHIÊN CỨU ỨNG DỤNG HỌC MÁY TRONG XỬ LÝ
ẢNH PET/CT LÒNG NGỰC HỖ TRỢ CÔNG TÁC CHẨN
ĐOÁN UNG THƯ PHỔI**

LUẬN ÁN TIẾN SĨ VẬT LÝ NGUYÊN TỬ VÀ HẠT NHÂN

HÀ NỘI – 2026

BỘ GIÁO DỤC VÀ ĐÀO TẠO BỘ KHOA HỌC VÀ CÔNG NGHỆ
VIỆN NĂNG LƯỢNG NGUYÊN TỬ VIỆT NAM

HỒ QUANG TUẤN

**NGHIÊN CỨU ỨNG DỤNG HỌC MÁY TRONG XỬ LÝ
ẢNH PET/CT LỒNG NGỰC HỖ TRỢ CÔNG TÁC CHẨN
ĐOÁN UNG THƯ PHỔI**

LUẬN ÁN TIẾN SĨ

Chuyên ngành: Vật lý nguyên tử và hạt nhân

Mã số: 9.44.01.06

Người hướng dẫn khoa học:

- 1. PGS.TS Nguyễn Văn Thái**
- 2. TS Lê Xuân Chung**

Hà Nội – 2026

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu khoa học độc lập của riêng tôi. Các số liệu sử dụng phân tích trong luận án có nguồn gốc rõ ràng, đã công bố theo đúng quy định. Các kết quả nghiên cứu trong luận án do tôi tự tìm hiểu, phân tích một cách trung thực, khách quan và phù hợp với thực tiễn của Việt Nam. Các kết quả này chưa từng được công bố trong bất kỳ nghiên cứu nào khác.

Tác giả luận án

(ký và ghi rõ họ tên)

LỜI CẢM ƠN

Tôi xin bày tỏ lòng biết ơn sâu sắc đến các thầy hướng dẫn đã tận tình chỉ bảo, hỗ trợ tôi trong suốt quá trình nghiên cứu. Xin trân trọng cảm ơn Ban lãnh đạo cùng các thầy cô tại Trung tâm Đào tạo hạt nhân – Viện Năng lượng nguyên tử Việt Nam đã tạo những điều kiện thuận lợi nhất để tôi thực hiện luận án. Đặc biệt, tôi xin gửi lời cảm ơn chân thành đến Ban lãnh đạo Viện Khoa học và Kỹ thuật hạt nhân đã tạo điều kiện và hỗ trợ nguồn lực quý báu để thực hiện luận án, cùng các đồng nghiệp, bác sĩ tại Bệnh viện Ung bướu Hà Nội, Bệnh viện K và Bệnh viện Ung bướu TP. Hồ Chí Minh đã nhiệt tình cung cấp những thông tin, dữ liệu chuyên môn quan trọng để tôi hoàn thành nghiên cứu này.

Tác giả luận án

(ký và ghi rõ họ tên)

MỤC LỤC

MỤC LỤC	1
DANH MỤC HÌNH VẼ	3
DANH MỤC BẢNG BIỂU	6
DANH MỤC TỪ VIẾT TẮT	7
MỞ ĐẦU	9
CHƯƠNG 1. NGHIÊN CỨU TỔNG QUAN	12
1.1. Vai trò của ảnh chụp PET/CT trong chẩn đoán ung thư phổi	12
1.1.1. Dược chất phóng xạ và cơ chế chuyển hóa	12
1.1.2. Các chỉ dấu hình ảnh PET/CT xác định ung thư phổi	15
1.1.3. Vai trò trong chẩn đoán xác định và phân giai đoạn (Staging)	19
1.1.4. Vai trò trong lập kế hoạch xạ trị, theo dõi và đánh giá đáp ứng ...	21
1.2. Đặc trưng dữ liệu ảnh PET/CT và các tác vụ xử lý cốt lõi	23
1.2.1. Các yếu tố vật lý ảnh hưởng đến chất lượng ảnh PET/CT	23
1.2.2. Các tác vụ xử lý và phân tích hình ảnh PET/CT	26
1.3. Ứng dụng học máy trong xử lý ảnh PET/CT	28
1.3.1. Nguyên lý ứng dụng của học máy và các mô hình học sâu	28
1.3.2. Sự phát triển các mô hình học sâu sử dụng mạng tích chập	30
1.3.3. Xu hướng ứng dụng các mô hình học sâu trong xử lý hình ảnh PET/CT lồng ngực	34
1.4. Đối tượng và mục tiêu nghiên cứu của luận án	37
1.4.1 Đối tượng Radiomics PET/CT: Ưu điểm và hạn chế tồn tại	37
1.4.2 Mục tiêu và bài toán nghiên cứu	43
CHƯƠNG 2. PHƯƠNG PHÁP NGHIÊN CỨU	45
2.1. Xây dựng bộ dữ liệu	45
2.1.1. Tiêu chí lựa chọn và loại trừ	45
2.1.2. Quy trình thu thập dữ liệu	47
2.1.3. Các thông tin thu thập	56
2.1.4. Các chỉ tiêu đánh giá bộ dữ liệu	57
2.2. Xây dựng công cụ đọc và tiền xử lý ảnh	59
2.2.1. Module tiền xử lý ảnh PET/CT từ thiết bị chụp	59

2.2.2.	Mô đun hiển thị thông tin khối u từ ảnh phân đoạn thủ công.....	63
2.2.3.	Mô đun trích xuất dữ liệu của các bệnh nhân	66
2.3.	Phát triển nền tảng tự động phân đoạn ảnh PET/CT ứng dụng mô hình học sâu.....	67
2.3.1.	Cấu trúc chung	67
2.3.2.	Chuẩn bị dữ liệu cho mạng học sâu	68
2.3.3.	Lựa chọn và huấn luyện mô hình học sâu.....	70
2.3.4.	Phương pháp đánh giá.....	81
CHƯƠNG 3.	KẾT QUẢ VÀ BÀN LUẬN	83
3.1.	Đánh giá bộ dữ liệu	83
3.1.1.	Thông tin hiển thị và lưu trữ bộ dữ liệu	83
3.1.2.	Chất lượng bộ dữ liệu PET/CT qua các chỉ số.....	92
3.2.	Đánh giá công cụ đọc và tiền xử lý ảnh PET/CT	101
3.2.1.	Giao diện đồ họa của phần mềm hiển thị.....	101
3.2.2.	Chức năng tiền xử lý của phần mềm.....	102
3.2.3.	Kết quả trích xuất thông tin khối u	105
3.3.	Đánh giá chất lượng nền tảng tự động phân đoạn hình ảnh ứng dụng mô hình học sâu.....	107
3.3.1.	Kiểm định chéo mô hình	107
3.3.2.	Kiểm định chéo dữ liệu	109
KẾT LUẬN		116
DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ		118
TÀI LIỆU THAM KHẢO		120

DANH MỤC HÌNH VẼ

Hình 1.1: Cơ chế tập trung FDG cao trong tế bào ung thư [4].....	14
Hình 1.2: Hình ảnh PET/CT (trái) và PET (phải) trên các mặt cắt ngang (axial), dọc giữa (sagittal) và đứng ngang (coronal) cho thấy sự hấp thu sinh lý ở một bé trai 13 tuổi. Sự hấp thu FDG bình thường được quan sát thấy ở não, tim, gan, lách, đại tràng, bàng quang và tủy xương, trong khi sự tập trung hoạt độ phóng xạ rất mạnh ở não so với phần còn lại của cơ thể [6].	15
Hình 1.3: Mức độ hấp thu của nốt được so sánh với nhu mô phổi xung quanh và hoạt độ nền tại túi máu trung thất (Mediastinal Blood Pool: MBP) [10].	16
Hình 1.4: Biểu đồ Nomogram tích hợp mô hình dự báo lâm sàng và kết quả chụp FDG-PET [12]	16
Hình 1.5: Ảnh hưởng vị trí và góc lệch không đồng trục tới độ phân giải [31]	24
Hình 1.6: Nguồn phát hình tròn (đường kính 10 mm) có hoạt độ đồng đều (100 đơn vị tùy ý) trong nền không phóng xạ tạo ra ảnh đo được trong đó một phần tín hiệu phát ra từ nguồn được nhìn thấy bên ngoài nguồn thực tế. Hoạt độ tối đa trong ảnh đo được giảm xuống còn 85. [36]	26
Hình 1.7: Trí tuệ nhân tạo, học máy và học sâu	28
Hình 1.8: Nguyên lý của mạng tích chập	31
Hình 1.9: Hai nhiệm vụ chính của mạng tích chập	31
Hình 1.10: Quá trình phát triển các mô hình học sâu trong phân đoạn hình ảnh [41].....	33
Hình 1.11 “Đám mây” thuật ngữ.....	34
Hình 1.12 Tần suất sử dụng mô hình.....	36
Hình 1.13 Tần suất các tác vụ của mô hình	37
Hình 1.14 Sự khác biệt về quy trình giữa việc trích xuất các đặc trưng thủ công (handcrafted features) và các đặc trưng sâu (deep features).	39
Hình 2.1: Minh họa máy chụp PET/CT.....	47
Hình 2.2. Giao diện phần mềm 3D-Slicer	54
Hình 2.3. Hình ảnh khối u phổi được đánh dấu trên lát cắt từ máy quét PET.....	54
Hình 2.4. Hình ảnh khối u phổi được phân đoạn trên một lát cắt từ phim chụp CT.....	55
Hình 2.5: Lưu đồ thuật toán chương trình xác định đường bao	64
Hình 2.6. Hình ảnh khối u không được tô kín	65
Hình 2.7. Hình ảnh các khối u phân nhánh.....	66
Hình 2.8. Lưu đồ chương trình ghi thông tin kết quả phân đoạn thủ công.....	67
Hình 2.9. Nền tảng phân đoạn tự động	68

Hình 2.10: Trực quan hóa dữ liệu PET/CT đa phương thức và quy trình khởi tạo nhãn chuẩn. (A) Ảnh CT không cản quang; (B) Ảnh PET ^{18}F -FDG (bản đồ SUV); (C) Ảnh dung hợp PET/CT; (D) Sự khác biệt giữa ba chuyên gia khoanh vùng; (E) Bản đồ xác suất từ thuật toán STAPLE; và (F) Kết quả đồng thuận xác suất cuối cùng được sử dụng làm tiêu chuẩn tham chiếu.	69
Hình 2.11: Cấu trúc mạng U-net	71
Hình 2.12: Quy trình tự động cấu hình phương pháp phân đoạn ảnh sinh y học dựa trên nnU-Net	72
Hình 2.13: Tổng quan về kiến trúc Swin UNETR. Đầu vào của mô hình là khối ảnh MRI đa phương thức 3D gồm 4 kênh. Swin UNETR tạo ra các vùng ảnh (patches) không chồng chập từ dữ liệu đầu vào, rồi sử dụng một lớp phân chia vùng ảnh (patch partition layer) để thiết lập các cửa sổ có kích thước mong muốn phục vụ cho việc tính toán tính năng tự chú ý (self-attention). Các biểu diễn đặc trưng đã được mã hóa trong mạng Swin Transformer sẽ được chuyển đến một bộ giải mã CNN (CNN-decoder) thông qua các kết nối nhảy bậc (skip connections) ở nhiều độ phân giải khác nhau. Kết quả đầu ra phân đoạn cuối cùng.	76
Hình 2.14: Các kịch bản ứng dụng phân đoạn khối với mạng 3D U-Net. (a) Phân đoạn bán tự động: Người dùng gán nhãn một vài lát cắt trên mỗi khối dữ liệu cần phân đoạn, sau đó mạng sẽ dự đoán để đưa ra kết quả phân đoạn dày đặc (dense segmentation). (b) Phân đoạn hoàn toàn tự động: Mạng được huấn luyện dựa trên các lát cắt đã gán nhãn từ một tập huấn luyện mang tính đại diện, sau đó có thể vận hành trực tiếp trên các khối dữ liệu chưa được gán nhãn.	77
Hình 2.15: Kiến trúc mạng 3D U-Net. Các khối màu xanh biểu diễn các bản đồ đặc trưng (feature maps). Số lượng kênh (channels) được ký hiệu ngay phía trên mỗi bản đồ đặc trưng.	79
Hình 3.1. Hình ảnh hiển thị độ phân giải của ảnh CT thu thập trong luận án	88
Hình 3.2. Hình ảnh hiển thị độ phân giải của ảnh PET thu thập trong luận án	88
Hình 3.3. Sơ đồ phân loại và lưu trữ của bộ dữ liệu PET/CT của luận án	89
Hình 3.4. Cấu trúc lưu dữ liệu ảnh PET/CT của nhóm người được gán nhãn là không mắc bệnh ung thư phổi không tế bào nhỏ	90
Hình 3.5. Cấu trúc lưu dữ liệu ảnh PET/CT của nhóm bệnh nhân được gán nhãn là mắc bệnh ung thư phổi không tế bào nhỏ	91
Hình 3.6. Ảnh PET/CT và một ảnh lát cắt của bệnh nhân UBHN27, đã được bác sĩ 01 khoanh vùng và trích trọn thể tích u, xóa các bộ phận khác xung quanh giúp chúng ta có thể quan sát rõ hơn về khối u.	95
Hình 3.7. Ảnh PET/CT và một ảnh lát cắt của bệnh nhân UBHN27, đã được bác sĩ 02 khoanh vùng và trích trọn thể tích u, xóa các bộ phận khác xung quanh giúp chúng ta có thể quan sát rõ hơn về khối u.	96

Hình 3.8. Ảnh PET/CT và một ảnh lát cắt của bệnh nhân UBHN27 ở các góc nhìn khác nhau.	97
Hình 3.9: Ảnh PET/CT của bệnh nhân BVK – 21 do bác sĩ 01 khoanh vùng khối u	98
Hình 3.10: Ảnh PET/CT của bệnh nhân BVK – 21 do bác sĩ 02 khoanh vùng khối u	98
Hình 3.11: Ảnh PET/CT của bệnh nhân UHCM-05 do bác sĩ 01 khoanh vùng khối u	99
Hình 3.12. Ảnh PET/CT của bệnh nhân UHCM-05 do bác sĩ 02 khoanh vùng khối u	99
Hình 3.13. Ảnh PET/CT của bệnh nhân UHCM-94 do bác sĩ 01 khoanh vùng khối u	100
Hình 3.14: Ảnh PET/CT của bệnh nhân UHN-105 do bác sĩ 02 khoanh vùng khối u	100
Hình 3.15. Giao diện chương trình và hiển thị ảnh DICOM	101
Hình 3.16. Hình ảnh PET trước và sau khi thực hiện hiệu chỉnh cường độ	103
Hình 3.17. Ảnh PET/CT trước (A và C) và sau khi sử dụng hàm lọc Gauss (B và D)	103
Hình 3.18. Hình ảnh hợp nhất PET và CT theo trục ngang và theo trục dọc	104
Hình 3.19. Thể tích khối u tại một lát cắt xuất hiện vài điểm không được đánh dấu là khối u phổi	104
Hình 3.20. Hình ảnh của một lát cắt dữ liệu sau khi đã hiệu chỉnh lại	105
Hình 3.21. Một số thông số về các node đã phân loại	106
Hình 3.22. Một số thông số của bộ dữ liệu	106
Hình 3.23. Kết quả định tính tiêu biểu với các kiến trúc huấn luyện khác nhau	108
Hình 3.24. Phân bố DSC trên các tập dữ liệu nội bộ và bên ngoài. Biểu đồ hộp minh họa giá trị trung vị, khoảng tứ phân vị (IQR) và các vòng tròn bao cho các mô hình được đánh giá chuẩn, làm nổi bật tính ổn định vượt trội của khung được tinh chỉnh dưới sự thay đổi miền dữ liệu.	110
Hình 3.25: Phân tích sự tương đồng lâm sàng của MTV được trích xuất bởi mô hình nnU-Net tinh chỉnh so với sự đồng thuận của chuyên gia STAPLE	112
Hình 3.26. Phân tích định tính sự không nhất quán phân đoạn ở các vùng ranh giới giải phẫu phức tạp: (A) phân đoạn quá mức màng phổi; (B) phân đoạn thiếu lõi khối u hoại tử không hấp thụ FDG	113

DANH MỤC BẢNG BIỂU

<i>Bảng 1.1: Phân loại TNM-8</i>	<i>19</i>
<i>Bảng 1.2: Số lượng các nghiên cứu về các loại ung thư và bệnh ác tính khác nhau.....</i>	<i>35</i>
<i>Bảng 1.3: Các nghiên cứu tiêu biểu về ứng dụng mô hình học sâu sử dụng dữ liệu ảnh PET/CT..</i>	<i>39</i>
<i>Bảng 2.1: Một số tiêu chí định lượng của bộ dữ liệu dùng huấn luyện học máy.....</i>	<i>46</i>
<i>Bảng 3.1. Các đặc điểm về tuổi và giới tính của bộ dữ liệu.....</i>	<i>83</i>
<i>Bảng 3.2. Thông tin kích thước khối u.....</i>	<i>83</i>
<i>Bảng 3.3 Mức độ hấp thu ^{18}F-FDG (SUV_{max}) theo kích thước khối u.....</i>	<i>84</i>
<i>Bảng 3.4. Bảng kết quả chụp PET/CT của bệnh nhân</i>	<i>85</i>
<i>Bảng 3.5. Thông tin kết quả chẩn đoán.....</i>	<i>86</i>
<i>Bảng 3.5: Kết quả đánh giá một số tiêu chí định lượng cho bộ dữ liệu.....</i>	<i>87</i>
<i>Bảng 3.6. Một số thông số chi tiết của bộ dữ liệu ảnh PET/CT.....</i>	<i>92</i>
<i>Bảng 3.7. So sánh giá trị độ lệch về thể tích của tập hợp phân đoạn giữa các bác sĩ.....</i>	<i>93</i>
<i>Bảng 3.8. So sánh giá trị chỉ số tương đồng của tập hợp phân đoạn giữa các bác sĩ.....</i>	<i>94</i>
<i>Bảng 3.9: So sánh định lượng hiệu năng phân đoạn giữa các kiến trúc học sâu khác nhau trên bộ dữ liệu nội bộ Việt Nam.....</i>	<i>107</i>
<i>Bảng 3.10. Phân tích DSC về sự phù hợp giữa mô hình đề xuất và ý kiến của từng chuyên gia chẩn đoán hình ảnh.</i>	<i>109</i>
<i>Bảng 3.11. Hiệu suất tổng quát hóa không cần huấn luyện của các mô hình phân đoạn trên các bộ dữ liệu quốc tế độc lập (AutoPET II và LUNG-PET-CT-Dx).</i>	<i>109</i>
<i>Bảng 3.12. Kết quả nghiên cứu loại bỏ đánh giá tác động của học chuyển giao lên độ chính xác phân đoạn và hiệu quả huấn luyện.....</i>	<i>111</i>

DANH MỤC TỪ VIẾT TẮT

STT	Ký hiệu viết tắt	Chữ viết đầy đủ	Tiếng Việt
1.	ACCP	American College of Chest Physicians	Hội Bác sĩ Lồng ngực Hoa Kỳ
2.	AI	Artificial Intelligence	Trí tuệ nhân tạo
3.	AJCC	American Joint Committee on Cancer	Ủy ban Liên hợp về Ung thư Hoa Kỳ
4.	ATP	Adenosine triphosphate	Hợp chất mang năng lượng tế bào
5.	CAD	Computer-Aided Diagnosis	Chẩn đoán có sự hỗ trợ của máy tính
6.	CADe	Computer aided detection	Phát hiện có sự hỗ trợ của máy tính
7.	CAD _x	Computer aided diagnosis	Chẩn đoán có sự hỗ trợ của máy tính
8.	CNN	Convolution neuron network	Mạng nơ-ron tích chập
9.	CT	Computed tomography	Chụp cắt lớp vi tính
10.	CTV	Clinical Target Volume	Thể tích đích lâm sàng
11.	DF	Deep features	Đặc trưng sâu
12.	DICOM	Digital Imaging and Communications in Medicine	Chuẩn lưu trữ và truyền tải hình ảnh y khoa
13.	DSC	Dice similarity coefficient	Hệ số tương đồng Dice
14.	EANM	European Association of Nuclear Medicine (EANM)	Hiệp hội Y học Hạt nhân Châu Âu
15.	FDG	2-deoxy-2-fluoro-deoxyglucose	2-deoxy-2-[18F]fluoro-D-glucose
16.	FOV	Field of view	Trường nhìn
17.	GLUT	Glucose transporter	Protein vận chuyển glucose
18.	GTV	Gross Tumor Volume	Thể tích khối u thô
19.	HF	Handcrafted Features	Đặc trưng thủ công
20.	HGNC	HUGO Gene Nomenclature Committee	Ủy ban Danh pháp Gen HUGO
21.	HU	Hounsfield unit	Đơn vị Hounsfield
22.	LDCT	Low dose computed tomography	Chụp cắt lớp vi tính liều thấp

STT	Ký hiệu viết tắt	Chữ viết đầy đủ	Tiếng Việt
23.	ML	Machine learning	Học máy
24.	MTV	Metabolic Tumor Volume	Thể tích chuyển hóa của khối u
25.	NIFTI	Neuroimaging Informatics Technology Initiative	Sáng kiến Công nghệ Tin học Chẩn đoán Hình ảnh Thần kinh
26.	NPV	Negative Predictive Value	Giá trị dự báo âm tính
27.	NRRD	Nearly Raw Raster Data	Dữ liệu lưới gần thô
28.	NSCLC	Non small cell lung cancer	Ung thư phổi không tế bào nhỏ
29.	PERCIST	Positron Emission Tomography Response Criteria in Solid Tumors	Tiêu chuẩn đánh giá đáp ứng điều trị u rắn theo PET
30.	PET	Positron emission tomography	Ghi hình cắt lớp Positron
31.	PPV	Positive Predictive Value	Giá trị dự báo dương tính
32.	PSNR	Peak Signal-to-Noise Ratio	Tỷ số tín hiệu trên tạp âm đỉnh của hình ảnh
33.	PTV	Planning Target Volume	Thể tích đích lập kế hoạch
34.	QIN	Quantitative Imaging Network	Mạng lưới Định lượng Chẩn đoán Hình ảnh
35.	ROI	Region of interest	Vùng quan tâm
36.	SE	Sensitivity	Độ nhạy
37.	SPECT	Single photon emission computed tomography	Chụp cắt lớp vi tính bằng bức xạ đơn photon
38.	SPN	Solitary Pulmonary Nodule	Nốt đơn độc ở phổi
39.	SUV	Standardized Uptake Value	Giá trị hấp thu chuẩn độ
40.	TLG	Total Lesion Glycolysis	Tổng lượng phân hủy đường của tổn thương
41.	TNM	Classification of Malignant Tumors	Phân loại các khối u ác tính
42.	UICC	Union for International Cancer Control	Hiệp hội Kiểm soát Ung thư Quốc tế
43.	VE	Volume error	Sai số thể tích

MỞ ĐẦU

Chẩn đoán hình ảnh hiện đại đóng vai trò quyết định trong việc cung cấp các bằng chứng định lượng phục vụ xây dựng phác đồ và đánh giá đáp ứng điều trị ung thư. Sự phát triển của các hệ thống ghi hình y tế thế hệ mới đã tối ưu hóa đáng kể độ phân giải giải phẫu và chức năng, song cũng làm phát sinh thách thức về mặt xử lý "dữ liệu lớn" (Big Data). Điển hình là trong tầm soát và quản lý ung thư phổi, kỹ thuật chụp ảnh cắt lớp phát xạ positron kết hợp vi tính (PET/CT) đã trở thành tiêu chuẩn vàng nhờ năng lực hợp nhất thông tin cấu trúc (ảnh CT) và thông tin chuyển hóa sinh học (ảnh PET). Trên cơ sở ảnh hợp nhất, bác sĩ sẽ thực hiện so sánh tương quan giữa độ tương phản (phản ánh mật độ tập trung hoạt độ phóng xạ) và hoạt động chuyển hóa của tổn thương so với các mô nền lân cận hoặc các cơ quan tham chiếu tiêu chuẩn (như gan hoặc túi máu trung thất). Mục tiêu lý tưởng là giá trị của mỗi điểm ảnh (pixel) trên hình ảnh PET phải tỉ lệ thuận với số lượng phân rã hạt nhân xảy ra trong suốt quá trình ghi hình tại đơn vị thể tích (voxel) tương ứng trong cơ thể bệnh nhân.

Chỉ số hấp thụ chuẩn hóa (Standardized Uptake Value - SUV) là một đại lượng tính toán được sử dụng trong ghi hình với ^{18}F -FDG nhằm chuẩn hóa mức độ hấp thụ của dược chất phóng xạ này dựa trên liều lượng tiêm vào, sự phân rã phóng xạ tính từ thời điểm tiêm và khối lượng cơ thể của bệnh nhân. Bác sĩ sẽ sử dụng chỉ số SUV để hỗ trợ đánh giá đặc tính của các mô có sự tích tụ dược chất phóng xạ, đặc biệt là trong việc phân biệt giữa tổn thương lành tính và ác tính, hoặc để theo dõi quá trình điều trị khối u. Các dòng máy chụp ảnh PET/CT tiên tiến hiện nay do các hãng danh tiếng như Siemens Healthineers (Đức), GE HealthCare (Mỹ), Philips Healthcare (Hà Lan) ...v.v đã tập trung phát triển công nghệ đo đạc và xử lý tín hiệu nhằm đo đạc và tính toán một cách chính xác chỉ số SUV để định lượng tốc độ tiêu thụ glucose tại chỗ của mô dựa trên lượng FDG đánh dấu phóng xạ tích tụ, từ đó phân đoạn khối u để đo lường thể tích khối u chuyển hóa (MTV - Metabolic Tumor Volume). Các kết quả xử lý dữ liệu sâu hơn được ứng dụng trong chẩn đoán xác định, phân giai đoạn, lập kế hoạch xạ trị, theo dõi và đánh giá đáp ứng điều trị ung thư phổi.

Mặc dù các hệ thống PET/CT đã tích hợp các phần mềm hỗ trợ xử lý dữ liệu thông minh, việc khai thác hiệu quả khối dữ liệu thể tích 3D tích hợp lên tới hàng ngàn lát cắt này vẫn đang gặp phải những nút thắt lớn do quy trình phân tích lâm sàng truyền thống chủ yếu dựa vào đánh giá định tính của con người. Quá trình đọc và phân tích kết quả ^{18}F -FDG PET/CT được thực hiện độc lập bởi hai bác sĩ y học hạt nhân giàu kinh nghiệm, những người đã nắm vững toàn bộ bệnh sử lâm sàng của bệnh nhân. Trong trường hợp có sự khác biệt về nhận định, một

chuyên gia thứ ba sẽ tham gia đề thảo luận và đưa ra kết luận thống nhất cuối cùng. Thực tế thực hành lâm sàng cho thấy, việc diễn giải thủ công ảnh PET/CT không chỉ tốn thời gian mà còn chịu ảnh hưởng lớn bởi yếu tố chủ quan, sự mệt mỏi của bác sĩ và các nhiễu sinh lý. Điều này dẫn đến sự sai lệch trong việc xác định thể tích khối u thô (GTV), bỏ sót hạch di căn kích thước nhỏ hoặc đánh giá không toàn diện tính bất đồng nhất của khối u.

Nhằm khắc phục các hạn chế này, xu hướng ứng dụng học máy (Machine Learning) và học sâu (Deep Learning) trong xử lý ảnh PET/CT đang trở thành một hướng nghiên cứu mũi nhọn, đặc biệt là xu hướng Radiomics. Về cốt lõi, Radiomics là quá trình trích xuất hiệu năng cao các đặc trưng sâu (Deep Features - DF) từ hình ảnh CT, MRI, PET,..v.v thông qua các mạng thần kinh tầng sâu, đặc biệt là mạng thần kinh tích chập (CNN) và chuyển đổi dữ liệu hình ảnh tại vùng quan tâm (Region of Interest - ROI) thành dữ liệu đa chiều có khả năng khai phá. Các đặc trưng này bao gồm cường độ điểm ảnh, hình dạng, thể tích, đặc trưng cấu trúc, v.v. Các đặc trưng sâu (Deep Features - DF) được trích xuất từ hình ảnh y tế thông qua các mạng thần kinh tầng sâu, đặc biệt là mạng thần kinh tích chập (CNN). So với đặc trưng thủ công, đặc trưng sâu có khả năng nắm bắt được những thuộc tính phức tạp và ở cấp độ cao hơn. Quy trình trích xuất DF hoàn toàn không phụ thuộc vào các thao tác thủ công, từ đó giảm thiểu sai sót do con người và cải thiện đáng kể độ chính xác trong phân tích dữ liệu hình ảnh. Bằng cách tự động học và nhận diện các quy luật phức tạp ẩn trong dữ liệu, DF giúp tăng cường khả năng biểu diễn đặc trưng, đồng thời nâng cao rõ rệt độ chính xác cũng như độ tin cậy của các kết quả chẩn đoán. Cho đến nay, các mô hình mạng tích chập (CNN) và kiến trúc Transformer đa phương thức tích hợp cơ chế chú ý (Attention Mechanism) đã chứng minh tiềm năng vượt trội trong các tác vụ **tự động** đăng ký ảnh, phân đoạn tổn thương và phân loại ác tính một cách trực quan, chính xác.

Hơn nữa, trong nghiên cứu và phát triển các hệ thống trí tuệ nhân tạo (AI) ứng dụng trong chẩn đoán ung thư phổi, dữ liệu đầu vào giữ vai trò tiên quyết, vượt trội hơn vai trò của cấu trúc thuật toán. Cụ thể, dữ liệu ảnh đa phương thức PET/CT đóng vai trò là nền tảng cốt lõi định hình độ chính xác, tính khả thi và độ an toàn sinh học của các mô hình học sâu. Nếu thiếu nguồn dữ liệu PET/CT được chuẩn hóa cấu trúc và gán nhãn chuẩn mực (ground truth), các mô hình AI chẩn đoán sẽ không thể vượt qua ranh giới thử nghiệm phòng thí nghiệm để đạt đủ điều kiện phê duyệt ứng dụng trong thực tiễn lâm sàng. Phần lớn các bộ dữ liệu huấn luyện hiện nay đều có nguồn gốc từ các quần thể phương Tây, dấy lên những lo ngại sâu sắc về tính bền vững của thuật toán khi triển khai trên các bệnh nhân châu Á — những người vốn sở hữu các đặc điểm di truyền, đặc tính bệnh lý và phân phối giá trị hấp thụ chuẩn hóa (SUV) hoàn toàn khác biệt.

Trong bối cảnh đó, luận án với đề tài "**Nghiên cứu ứng dụng học máy trong xử lý ảnh PET/CT lồng ngực hỗ trợ công tác chẩn đoán bệnh ung thư phổi**" được thực hiện tập trung vào ba mục tiêu chính sau:

- (1) Thiết lập cơ sở dữ liệu ảnh chụp PET/CT người Việt bản địa quy mô lớn
- (2) Xây dựng công cụ đọc và tiền xử lý ảnh nhằm trích xuất dữ liệu ảnh DICOM phục vụ huấn luyện mô hình học sâu
- (3) Phát triển nền tảng học sâu tự động phân đoạn ảnh chụp PET/CT hỗ trợ công tác chẩn đoán ung thư phổi

Luận án đi sâu vào giải quyết bài toán kỹ thuật cốt lõi: thiết lập cơ sở dữ liệu ảnh chụp PET/CT người Việt, tiền xử lý khử nhiễu ảnh, hợp nhất đặc trưng PET-CT hiệu quả, đến phát triển nền tảng kiến trúc mạng nơ-ron tối ưu cho việc phân đoạn và phân loại tổn thương phổi dựa trên mô hình mạng tích chập (CNN) và kiến trúc Transformer đa phương thức tích hợp cơ chế chú ý (Attention Mechanism) tiêu biểu như nnUnet, 3D U-Net và Swin UNETR.

Ngoài phần mở đầu, kết luận và tài liệu tham khảo, nội dung chính của luận án được trình bày trong 3 chương như sau:

- **Chương 1** trình bày nghiên cứu tổng quan về vai trò của ảnh chụp PET/CT trong chẩn đoán ung thư phổi, đặc trưng dữ liệu ảnh và các tác vụ xử lý cốt lõi, ứng dụng của các mô hình học máy học sâu trong xử lý ảnh PET/CT, từ đó trình bày đối tượng và mục tiêu nghiên cứu của luận án.
- **Chương 2** trình bày phương pháp nghiên cứu, bao gồm phương pháp cách thức xây dựng bộ dữ liệu người Việt (tiêu chí, quy trình, thông tin và chỉ tiêu đánh giá), phương pháp xây dựng công cụ đọc và tiền xử lý ảnh PET/CT, phương pháp phát triển nền tảng tự động phân đoạn ảnh PET/CT dựa trên mạng học sâu (từ xây dựng cấu trúc, chuẩn bị dữ liệu, lựa chọn mạng học sâu, huấn luyện và đánh giá)
- **Chương 3** trình bày các kết quả đánh giá bộ dữ liệu, công cụ đọc và tiền xử lý ảnh PET/CT và chất lượng nền tảng tự động phân đoạn ảnh PET/CT.

Kết quả nghiên cứu trình bày trong luận án được kỳ vọng sẽ đóng góp một công cụ hỗ trợ chẩn đoán (CAD) có tính chính xác cao, khách quan và khả thi, đáp ứng yêu cầu của nền y tế số hóa hiện nay.

CHƯƠNG 1. NGHIÊN CỨU TỔNG QUAN

1.1. Vai trò của ảnh chụp PET/CT trong chẩn đoán ung thư phổi

1.1.1. Dược chất phóng xạ và cơ chế chuyển hóa

Theo báo cáo của Ủy ban Khoa học Liên Hợp Quốc về Tác dụng của Bức xạ Nguyên tử (UNSCEAR), số lượng ca kiểm tra chẩn đoán hình ảnh và liệu bức xạ hiệu dụng tính theo đầu người từ X-quang can thiệp đang tăng mạnh trên toàn cầu. Sự gia tăng này không chỉ đặt ra yêu cầu cấp bách về vai trò của các nhà vật lý y khoa trong việc đảm bảo an toàn bức xạ, mà còn biến việc ứng dụng học máy thành một xu thế tất yếu nhằm giảm tải áp lực công việc, đồng thời hỗ trợ bác sĩ chẩn đoán nhanh chóng và chính xác hơn. [1]

Xu hướng tối ưu hóa này đặc biệt quan trọng trong các kỹ thuật chuyên sâu như ghi hình PET/CT – phương pháp đang có sự bùng nổ mạnh mẽ những thập kỷ gần đây trong chẩn đoán và quản lý ung thư phổi. Bằng cách tích hợp linh hoạt giữa thông tin chức năng của PET và cấu trúc giải phẫu của CT, PET/CT cho phép đánh giá toàn diện bệnh lý, tối ưu hóa độ chính xác khi phân loại giai đoạn bệnh và định danh các nốt đơn độc ở phổi chưa rõ bản chất.

Đáng chú ý, do phần lớn bệnh lý đều khởi phát từ những biến đổi đặc hiệu trong nội môi và chuyển hóa, kỹ thuật PET sở hữu lợi thế vượt trội so với SPECT nhờ khả năng khai thác đa dạng các đồng vị phóng xạ phát positron. Khi gắn kết với các phân tử sinh học, các đồng vị này hoạt động như những chất đánh dấu phân tử (molecular tracers), giúp phát hiện và định lượng chính xác các chức năng sinh lý cũng như lộ trình chuyển hóa tại các mô biệt hóa. Đây chính là nguồn dữ liệu hình ảnh chuyển hóa chuyên sâu và phức tạp, mở ra không gian lý tưởng cho các thuật toán học máy khai phá, từ đó nâng cao hiệu quả chẩn đoán lâm sàng và an toàn bức xạ cho người bệnh.

a) Cơ chế sử dụng ^{18}F -FDG thăm dò chuyển hoá glucose

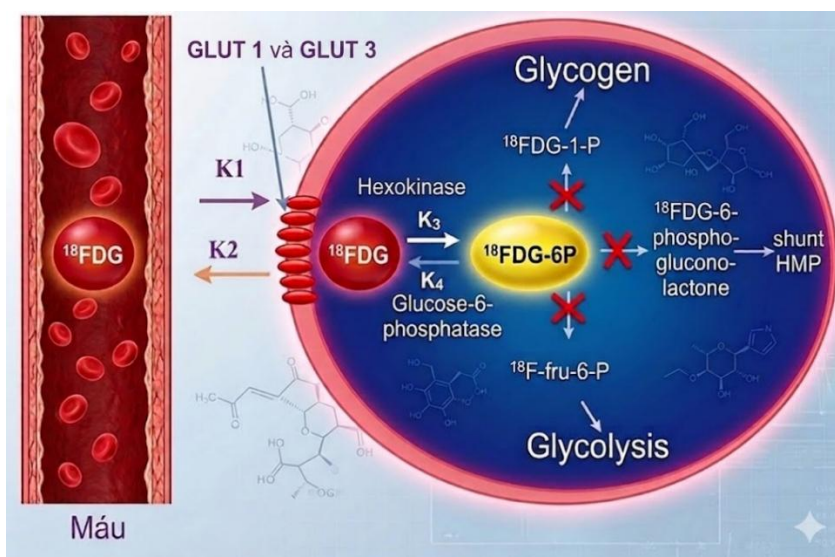
Phân tử quan trọng nhất cung cấp năng lượng cho các phản ứng hoá sinh đa dạng trong cơ thể là adenosin triphosphat (ATP), chất được sinh ra trong ty thể do quá trình chuyển hoá glucose. Ở não chuyển hoá glucose sinh ra tất cả ATP, trong khi ở tim chuyển hoá axit béo là nguồn năng lượng chính để tổng hợp ATP. Chuyển hoá glucose tăng cao thậm chí ngay cả khi sự cung cấp oxy tại đó bị suy giảm. Điều này có thể thấy được trong các mô ung thư. Glucose được vận chuyển vào trong tế bào qua cơ chế khuếch tán thụ động thông qua kênh vận chuyển glucose đặc hiệu nằm trên màng tế bào. Trong bào tương, glucose được phosphoryl hóa nhờ men hexokinase để tạo thành glucose 6-phosphat, chuyển hoá đến sản phẩm cuối cùng là CO_2 và H_2O [2] [3] [4].

Glucose được vận chuyển vào và ra khỏi tế bào thông qua một họ gồm 14 protein vận chuyển khuếch tán (GLUT) có tính đặc hiệu tế bào và chịu sự điều tiết của các yếu tố nội tiết cũng như môi trường. Danh sách các GLUT này có thể tra cứu tại trang web của Ủy ban Gen người HGNC (HUGO Gene Nomenclature Committee). Mặc dù ký hiệu gen chính thức hiện nay là *SLC24x*, nhưng để thuận tiện và gần gũi hơn, ta sẽ sử dụng ký hiệu GLUTx. Sự tăng biểu hiện (upregulation) của các protein GLUT là đặc điểm phổ biến ở hầu hết các loại ung thư và có mối tương quan nghịch với tiên lượng của bệnh nhân. GLUT đóng vai trò là bước giới hạn tốc độ (rate-limiting step) đầu tiên của quá trình chuyển hóa glucose, cho phép vận chuyển glucose qua màng tế bào theo gradient nồng độ mà không tiêu tốn năng lượng. Tình trạng tăng cường chuyển hóa glucose ở khối u có liên quan mật thiết đến sự biểu hiện quá mức mất kiểm soát của các GLUT (chủ yếu là các protein GLUT1 hoặc GLUT3 đáp ứng với tình trạng thiếu oxy). Cơ chế dẫn đến sự biểu hiện quá mức của gen GLUT1 có khả năng xuất phát từ các tương tác phức tạp do khối u điều phối với các yếu tố khởi động (promoter) hoặc tăng cường (enhancer) của GLUT1. Hiện tượng này cũng liên kết chặt chẽ với sự tăng đồng biểu hiện của các dấu ấn xâm lấn tế bào, chẳng hạn như matrix metalloproteinase-2 (MMP-2). Gần đây hơn, việc vận chuyển GLUT12 từ nội bào ra màng sinh chất cũng được cho là có liên quan đến sự tăng tích tụ glucose trong một số loại ung thư, bao gồm ung thư vú. Cần có thêm các nghiên cứu bổ sung để đánh giá loại hình, nồng độ và phạm vi biểu hiện của họ GLUT trong các bệnh lý ung thư khác nhau. Sự gia tăng về nồng độ và hoạt tính của enzyme hexokinase (chủ yếu là hexokinase loại II [HK-II] trong số 4 loại ở mô động vật có vú) cũng được xác định là có liên quan đến nhiều loại ung thư. HK-II liên kết với màng ty thể tại các kênh anion phụ thuộc điện thế (porin) và thực hiện quá trình phosphoryl hóa glucose thành glucose-6-phosphate một cách hiệu quả bằng cách sử dụng năng lượng do ty thể tạo ra [5].

Phân tử 2-deoxy-2-fluoro-deoxyglucose (FDG) được tạo bởi việc đánh dấu nguyên tử C số 2 trong phân tử bằng ^{18}F (thay thế 1 nguyên tử H liên kết với C số 2 bằng ^{18}F). Khi được chất phóng xạ ^{18}F -FDG được tiêm vào cơ thể, nó đóng vai trò như một chất tương đồng với glucose, bắt đầu hành trình từ dòng máu xuyên qua màng tế bào nhờ sự hỗ trợ của các protein vận chuyển GLUT 1 và GLUT 3 (Hình 1.1). Sau khi đã vào bên trong tế bào, FDG gặp enzyme Hexokinase và nhanh chóng bị phosphoryl hóa để trở thành ^{18}F -FDG -6-Phosphate (FDG-6P). Đây là bước ngoặt quan trọng vì cấu trúc hóa học của FDG-6P khiến nó không thể tiếp tục đi sâu vào các con đường chuyển hóa năng lượng thông thường như chu trình đường phân để tạo năng lượng, con đường pentose phosphate hay tổng hợp thành glycogen để dự trữ. Đồng thời, do mang điện tích và cấu trúc biến đổi, FDG-6P cũng không thể thẩm thấu ngược trở lại qua màng tế bào để ra ngoài. Sự

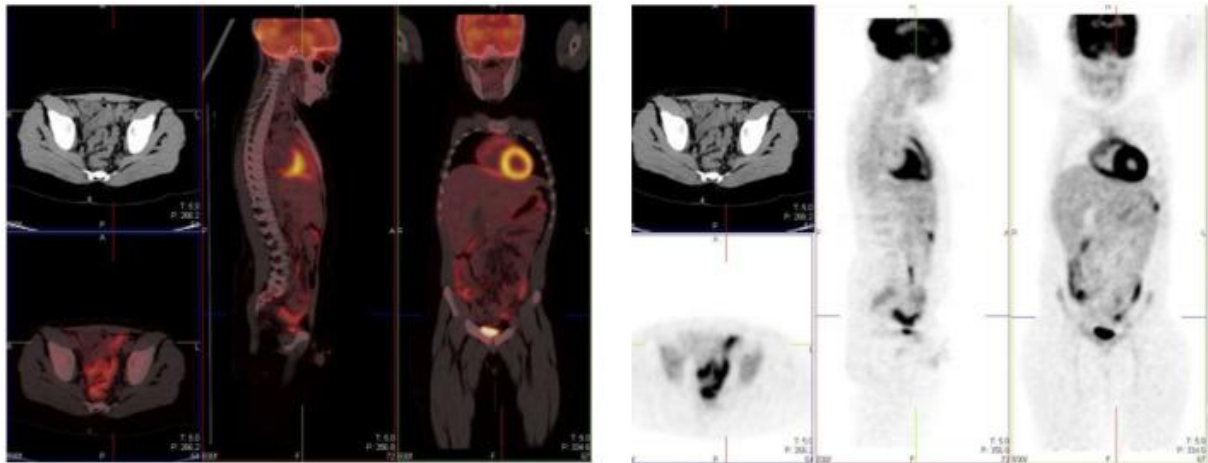
thiếu hụt enzyme Glucose-6-Phosphatase trong phần lớn các loại tế bào, đặc biệt là ở những tế bào có tốc độ tăng sinh cao như khối u, khiến quá trình khử phosphate để giải phóng FDG diễn ra cực kỳ chậm chạp. Kết quả là FDG-6P bị “mắc kẹt” vĩnh viễn bên trong tế bào trong suốt thời gian ghi hình.

Do các tế bào ung thư thường biểu hiện quá mức các kênh vận chuyển GLUT và có hoạt tính enzyme Hexokinase cao hơn hẳn tế bào bình thường để phục vụ nhu cầu năng lượng lớn, chúng sẽ tích tụ một lượng dược chất phóng xạ lớn. Sự chênh lệch về nồng độ phóng xạ này cho phép máy PET phát hiện và tái tạo thành những hình ảnh sáng tương phản, giúp các bác sĩ xác định chính xác vị trí và mức độ hoạt động của các tổn thương bệnh lý trên cơ thể người bệnh. Ở người bình thường thì glucose được lọc và tái hấp thu ở các nephron còn ^{18}F -FDG lại được tái hấp thu rất ít và bài xuất chủ yếu qua nước tiểu. Thời gian bán thải của ^{18}F -FDG khoảng 2 giờ. Nghĩa là khi vào cơ thể, cứ sau khoảng 2 giờ hoạt độ phóng xạ của ^{18}F -FDG trong cơ thể giảm còn một nửa. Sau 10 giờ, hoạt độ phóng xạ tiêm vào cơ thể còn lại chỉ khoảng 3%. **Hình 1.2** cho thấy hình ảnh PET hấp thụ glucose ở người bình thường: tập trung glucose cao ở não, kế tiếp là ở tim, gan và lách, trong khi đó ta có thể nhận thấy sự tăng hấp thụ glucose ở thận, bàng quang.



Hình 1.1: Cơ chế tập trung FDG cao trong tế bào ung thư [4]

Ta có thể tính mức độ hấp thụ glucose ở từng vùng quan tâm trên hình ảnh PET. Giá trị đó được gọi là giá trị hấp thụ chuẩn hóa (Standardized Uptake Value -SUV). Có thể hiểu là nếu ^{18}F -FDG phân bố đều khắp toàn cơ thể thì SUV ở các mô, cơ quan đều có giá trị như nhau và bằng 1. Do đó nếu tính được SUV tương đối của các mô, các cơ quan khác nhau trong cơ thể, so sánh SUV giữa các mô khác nhau hoặc giữa các lần ghi hình khác nhau trên cùng một bệnh nhân, từ đó ta có thể đánh giá được tổn thương, đánh giá đáp ứng với điều trị.



Hình 1.2: Hình ảnh PET/CT (trái) và PET (phải) trên các mặt cắt ngang (axial), dọc giữa (sagittal) và đứng ngang (coronal) cho thấy sự hấp thu sinh lý ở một bé trai 13 tuổi. Sự hấp thu FDG bình thường được quan sát thấy ở não, tim, gan, lách, đại tràng, bàng quang và tủy xương, trong khi sự tập trung hoạt độ phóng xạ rất mạnh ở não so với phần còn lại của cơ thể [6].

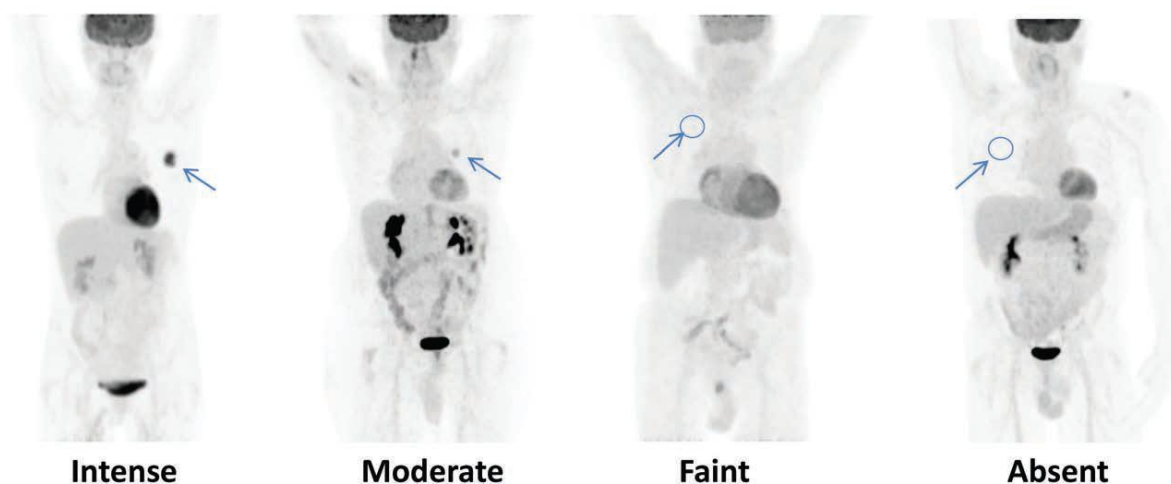
1.1.2. Các chỉ dấu hình ảnh PET/CT xác định ung thư phổi

a) Chỉ dấu định tính

Đánh giá định tính trong ghi hình ^{18}F -FDG PET/CT tập trung vào việc phân tích trực quan mức độ hấp thu dược chất phóng xạ, thay vì dựa vào các con số thuần túy. Phương pháp này đóng vai trò then chốt trong việc phân tách bản chất mô lành – ác, theo dõi đáp ứng điều trị và định vị chính xác tổn thương bệnh lý. Về nguyên tắc, bác sĩ sẽ thực hiện so sánh tương quan giữa độ tương phản (phản ánh mật độ tập trung hoạt độ phóng xạ) và hoạt động chuyển hóa của tổn thương so với các mô nền lân cận hoặc các cơ quan tham chiếu tiêu chuẩn (như gan hoặc túi máu trung thất) [7] [8].

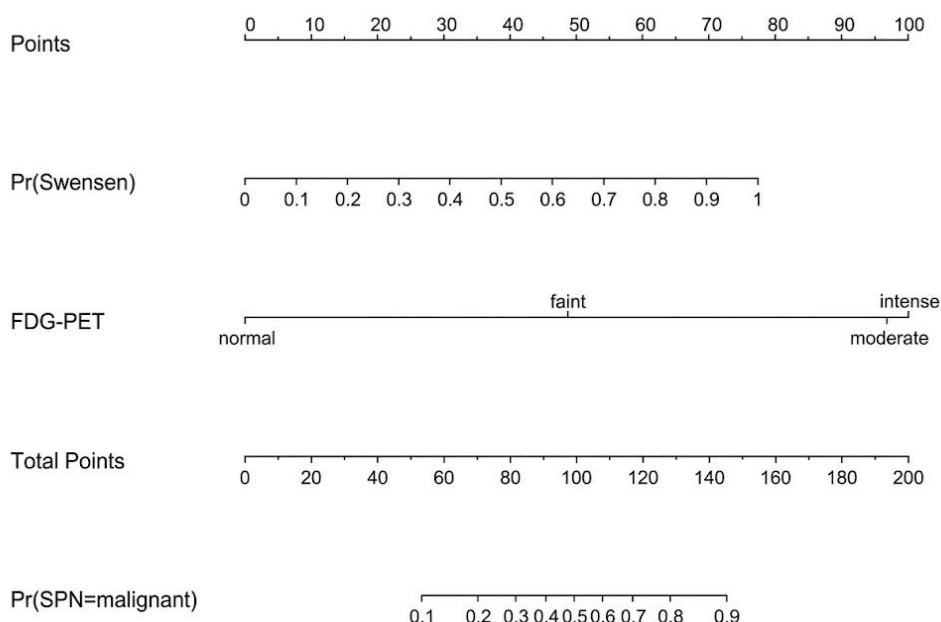
Cường độ hấp thu trực quan (Visual Uptake Intensity): Theo Ủy ban Tiêu chuẩn Chăm sóc Y tế thuộc Hội lồng ngực Anh quốc, mức độ hấp thu FDG tại các nốt đơn độc phổi (SPN) được đánh giá trực quan dựa trên thang điểm Herder 4 mức độ, được minh họa trên Hình 1.3, cụ thể như sau [9]:

- Mức 1 (Mạnh: Intense): Hấp thu rõ rệt, cao hơn hẳn so với túi máu trung thất (Dù không có quy định rạch ròi, nhưng mức hấp thu gấp 3 lần hoạt độ nền túi máu trung thất thường được định nghĩa là hấp thu mạnh, theo <https://radiopaedia.org/>).
- Mức 2 (Vừa: Moderate): Hấp thu cao hơn so với túi máu trung thất.
- Mức 3 (Nhẹ: Faint): Hấp thu thấp hơn hoặc bằng túi máu trung thất.
- Mức 4 (Không hấp thu: Absent. Normal): Hoàn toàn không ghi nhận sự tập trung dược chất.



Hình 1.3: Mức độ hấp thu của nốt được so sánh với nhu mô phổi xung quanh và hoạt độ nền tại túi máu trung thất (Mediastinal Blood Pool: MBP) [10].

Hình 1.4 mô tả biểu đồ Nomogram biểu thị xác suất chẩn đoán ác tính dựa trên sự kết hợp giữa mô hình của Swensen và cộng sự [11] cùng kết quả từ phim chụp FDG-PET. Cách sử dụng biểu đồ được tóm lược theo các bước cụ thể sau: Đầu tiên, xác định vị trí của bệnh nhân trên từng thang đo biến số dự báo. Mỗi vị trí trên các thang đo này tương ứng với một mức điểm tiên lượng cụ thể trên thanh "Điểm" (Points) nằm ở trên cùng. Sau đó ta cộng các giá trị điểm này lại để có được tổng điểm cuối cùng trên trục "Tổng điểm" (Total points). Từ giá trị trên trục tổng điểm, kẻ một đường thẳng dọc xuống trục xác suất để xác định khả năng mắc bệnh ác tính của bệnh nhân.



Hình 1.4: Biểu đồ Nomogram tích hợp mô hình dự báo lâm sàng và kết quả chụp FDG-PET [12]

b) Chỉ dấu định lượng

Chỉ số SUV (Standardized Uptake Value): Chỉ số hấp thụ chuẩn hóa (Standardized Uptake Value - SUV) là một đại lượng tính toán được sử dụng trong ghi hình với ^{18}F -FDG nhằm nỗ lực chuẩn hóa mức độ hấp thụ của dược chất phóng xạ này dựa trên liều lượng tiêm vào, sự phân rã phóng xạ tính từ thời điểm tiêm và khối lượng cơ thể của bệnh nhân [13] [14]. Một số bác sĩ sử dụng chỉ số SUV để hỗ trợ đánh giá đặc tính của các mô có sự tích tụ dược chất phóng xạ, đặc biệt là trong việc phân biệt giữa tổn thương lành tính và ác tính, hoặc để theo dõi quá trình điều trị khối u. Chỉ số này được định nghĩa như sau:

$$SUV = \frac{\text{Nồng độ hoạt độ phóng xạ trong một voxel hoặc một nhóm các voxel}}{\text{Hoạt độ phóng xạ tiêm vào trên khối lượng cơ thể}} \quad (1.1)$$

Trong đó, hoạt độ phóng xạ tiêm vào hoặc nồng độ hoạt độ phóng xạ đã được hiệu chỉnh theo sự phân rã phóng xạ, tính từ thời điểm định lượng liều cho đến khi bắt đầu ghi hình. SUV có đơn vị của khối lượng riêng (ví dụ: g/cm^3). Một số cơ sở y tế sử dụng khối lượng cơ thể không tính mỡ (lean body mass) hoặc diện tích bề mặt cơ thể thay cho tổng khối lượng cơ thể ở mẫu số của phương trình này. Vì chỉ số SUV không dựa trên một mô hình sinh lý học cụ thể (chẳng hạn như mô hình Sokoloff [15]), nên nó được coi là một chỉ số xấp xỉ hoặc chỉ số bán định lượng cho mức độ hấp thụ FDG. Việc đo lường chỉ số SUV chịu ảnh hưởng bởi nhiều yếu tố:

- 1) Độ chính xác khi định lượng hoạt độ tiêm vào: Bao gồm cả việc hiệu chỉnh lượng hoạt độ còn sót lại trong bơm tiêm sau khi tiêm.
- 2) Sự thoát mạch của dược chất phóng xạ: Xảy ra trong một tỉ lệ nhỏ các ca thực hiện tiêm truyền.
- 3) Việc hiệu chuẩn hệ thống PET/CT và độ chính xác của hiệu chỉnh suy giảm: Đảm bảo hệ thống ghi nhận đúng mức năng lượng và mật độ mô.
- 4) Khoảng thời gian từ lúc tiêm đến khi ghi hình: Bao gồm cả việc ghi chép chính xác thời điểm tiêm thuốc.
- 5) Độ chính xác của việc đo khối lượng cơ thể bệnh nhân.
- 6) Trạng thái sinh lý của bệnh nhân tại thời điểm tiêm: Bao gồm công tác chuẩn bị như nhịn ăn trước khi tiêm và có thể là việc sử dụng insulin.
- 7) Thành phần cơ thể bệnh nhân: Ví dụ như tỉ lệ giữa mô mỡ và mô nạc.
- 8) Kích thước tổn thương: Hiệu ứng thể tích bán phần có thể làm giảm giá trị SUV đo được đối với các tổn thương có kích thước rất nhỏ.
- 9) Cử động trong quá trình thu nhận hình ảnh: Đặc biệt là cử động hô hấp, gây ảnh hưởng đáng kể đến các tổn thương nhỏ.

10) Việc lựa chọn vùng quan tâm (ROI) của người sử dụng: Cách khoanh vùng để phân tích dữ liệu trên hình ảnh.

Các bác sĩ lâm sàng khi sử dụng chỉ số SUV để đưa ra các quyết định liên quan đến chẩn đoán, phân giai đoạn, điều trị và lập kế hoạch điều trị cho bệnh nhân nên thảo luận với bác sĩ y học hạt nhân về những sai số tiềm ẩn của chỉ số này tại cơ sở y học hạt nhân hoặc đơn vị cung cấp dịch vụ PET/CT của mình; đồng thời cần hết sức thận trọng khi so sánh hoặc sử dụng các chỉ số SUV đến từ các cơ sở y tế khác nhau.

Thể tích khối u chuyển hóa (MTV - Metabolic Tumor Volume): Định nghĩa của MTV là thể tích nằm trong một vùng quan tâm (ROI) do người dùng hoặc thuật toán xác định nhằm phân tách phần khối u có hoạt động chuyển hóa. Để xác định ranh giới của ROI này, các phương pháp dựa trên ngưỡng (threshold-based) hoặc dựa trên thuật toán (algorithm-based) đã được đề xuất và đánh giá. Rất nhiều thuật toán phân đoạn PET đã được phát triển và ứng dụng cho hình ảnh FDG PET. Việc phân nhóm các thuật toán này thành các lớp riêng biệt là một thách thức do sự đa dạng cực kỳ lớn của chúng. Những thuật toán tiên tiến hơn thường tích hợp kỹ thuật từ nhiều phương pháp khác nhau. Tuy nhiên, các thuật toán phân đoạn PET nhìn chung có thể được phân loại thành phương pháp dựa trên ngưỡng và phương pháp dựa trên thuật toán.

Tổng lượng glycolysis tổn thương (TLG - Total Lesion Glycolysis): là tích số giữa SUV trung bình (mean SUV) và thể tích khối u chuyển hóa (MTV), được Larson và các cộng sự [16] giới thiệu lần đầu tiên nhằm đánh giá đáp ứng điều trị. Chỉ số này là sự kết hợp giữa thông tin về thể tích và thông tin về chuyển hóa trên hình ảnh FDG PET. Nhiều nghiên cứu đã chứng minh tính hữu hiệu của TLG trong việc đánh giá đáp ứng điều trị đối với nhiều loại khối u khác nhau [17] [18] [19] [20] [21] [22] [23]. Các nghiên cứu gần đây còn cho thấy giá trị tiên đoán của chỉ số TLG trước điều trị trong ung thư xương, ung thư biểu mô màng phổi ác tính và ung thư vòm họng. Tuy nhiên, giá trị tiên lượng của TLG trong ung thư phổi không tế bào nhỏ (NSCLC) hiện vẫn còn là một ẩn số [24].

Việc sử dụng PET-FDG trong ung thư học để xác định giai đoạn và chẩn đoán tái phát khối u đang là một trong những lĩnh vực phát triển nhanh chóng nhất của y học hạt nhân nói riêng và chẩn đoán hình ảnh nói chung. Ứng dụng của PET-FDG trong chẩn đoán phân biệt nốt đơn độc ở phổi, cũng như xác định giai đoạn và phát hiện tái phát ung thư phổi, ung thư đại trực tràng đóng vai trò đặc biệt quan trọng trong ung thư học lâm sàng. Giá trị của kỹ thuật PET trong việc phát hiện sự hiện diện của tế bào ung thư trong những tình huống này là kết quả của sự cải thiện vượt bậc về độ nhạy đối với các khối u thể tích nhỏ, dựa trên những tiến bộ gần đây của thiết bị ghi hình PET. Chụp cắt lớp phát xạ positron

(PET) sở hữu tiềm năng đặc biệt trong việc theo dõi đáp ứng điều trị ung thư một cách nhanh chóng và chính xác thông qua hình ảnh chuyển hóa về những thay đổi trong bản chất chuyển hóa của khối u.

1.1.3. Vai trò trong chẩn đoán xác định và phân giai đoạn (Staging)

a) Vai trò trong chẩn đoán xác định

Việc xác định đặc điểm của một nốt đơn ở phổi (SPN) được phát hiện tình cờ hoặc thông qua tầm soát ung thư phổi bằng CT (vốn đang trở nên phổ biến gần đây) là một vấn đề sức khỏe cộng đồng quan trọng. Hiệp hội Bác sĩ lồng ngực Hoa Kỳ (ACCP) khuyến cáo sử dụng chụp CT lồng ngực như một trong những phương thức chính khi tầm soát ung thư phổi ở các nhóm đối tượng nguy cơ cao. ACCP cũng đưa ra lời khuyên nên thực hiện PET kết hợp với CT ngực trong những trường hợp cụ thể [25].

Theo hướng dẫn của ACCP, PET không có chỉ định đối với các nốt đơn độc ở phổi có đường kính dưới 8 mm, hoặc dưới 10 mm theo hướng dẫn của Pháp [25]. Ngưỡng 8–10 mm này được thiết lập để tính đến độ phân giải không gian của hệ thống PET, nhằm hạn chế nguy cơ cao xảy ra kết quả âm tính giả đối với các tổn thương nhỏ. Tuy nhiên, trong thập kỷ qua, độ phân giải không gian của PET đã tăng lên không ngừng giữa phân tích gộp đầu tiên được thực hiện bởi Gould và cộng sự [26] và nghiên cứu gần đây hơn của Cronin và cộng sự [27] – với độ nhạy 95%, độ đặc hiệu 82%, giá trị dự báo dương tính (PPV) 91% và giá trị dự báo âm tính (NPV) 90%. Độ phân giải không gian của các hệ thống mới nhất hiện nay thậm chí còn ưu việt hơn thế.

b) Vai trò trong phân giai đoạn

Dù được phát hiện thông qua tầm soát hay qua quy trình thăm khám nghi ngờ ung thư, **chụp cắt lớp vi tính (CT)** luôn là phương thức đầu tay để chẩn đoán và phân chia giai đoạn ban đầu. Cụ thể, CT liều thấp (LDCT) được dùng trong tầm soát, trong khi CT có tiêm thuốc cản quang tĩnh mạch thường được chỉ định cho các trường hợp nghi nhiễm. Những bệnh nhân được chẩn đoán mắc ung thư phổi sau đó sẽ được chỉ định chụp **¹⁸F-FDG PET/CT** để hoàn tất quá trình phân giai đoạn. Hệ thống phân loại **TNM phiên bản thứ 8 (TNM-8)** (Bảng 1.1) hiện là tiêu chuẩn để phân giai đoạn cho ung thư phổi không tế bào nhỏ (NSCLC).

Bảng 1.1: Phân loại TNM-8

Phân loại	Định nghĩa
T: Khối u nguyên phát (Primary Tumour)	
Tx	Khối u nguyên phát không thể đánh giá được, hoặc có bằng chứng tế bào ung thư qua các phương pháp phi hình ảnh (như trong đờm

Phân loại	Định nghĩa
	hoặc dịch rửa phế quản) nhưng không quan sát thấy qua chẩn đoán hình ảnh (hoặc nội soi phế quản).
T0	Không có bằng chứng của khối u nguyên phát.
Tis	Ung thư biểu mô tại chỗ (biểu mô vảy hoặc biểu mô tuyến).
T1	Khối u có kích thước lớn nhất ≤ 3 cm.
T1a	Khối u ≤ 1 cm.
T1b	Khối u > 1 cm nhưng ≤ 2 cm.
T1c	Khối u > 2 cm nhưng ≤ 3 cm.
T2	Khối u > 3 cm nhưng ≤ 5 cm; hoặc khối u có các đặc điểm: xâm lấn phế quản chính (không tính khoảng cách tới vị trí chia nhánh khí quản – carina), xâm lấn màng phổi tạng, gây xẹp phổi hoặc viêm phổi do tắc nghẽn (một phần hoặc toàn bộ).
T2a	Khối u có kích thước lớn nhất > 3 cm nhưng ≤ 4 cm.
T2b	Khối u có kích thước lớn nhất > 4 cm nhưng ≤ 5 cm.
T3	Khối u > 5 cm nhưng ≤ 7 cm; hoặc khối u xâm lấn trực tiếp vào: màng phổi thành, thành ngực (bao gồm cả khối u rãnh đỉnh phổi/Pancoast), dây thần kinh hoành, màng ngoài tim thành; hoặc có các nốt u riêng biệt nằm cùng thùy với u nguyên phát.
T4	Khối u > 7 cm; hoặc khối u xâm lấn vào: trung thất, cơ hoành, tim, mạch máu lớn, khí quản, dây thần kinh thanh quản quặt ngược, thực quản, thân đốt sống, vị trí chia nhánh khí quản (carina); hoặc có các nốt u riêng biệt nằm ở thùy khác nhau của cùng bên phổi bệnh.
N: Hạch vùng (Regional Lymph Nodes)	
N0	Không có di căn hạch vùng.
N1	Di căn hạch quanh phế quản cùng bên và/hoặc hạch rốn phổi cùng bên, hạch nội phổi, bao gồm cả sự xâm lấn trực tiếp từ khối u.
N2	Di căn hạch trung thất cùng bên và/hoặc hạch dưới vị trí chia nhánh khí quản (hạch dưới carina).
N3	Di căn hạch trung thất đối bên, rốn phổi đối bên, hạch cơ bậc thang hoặc hạch thượng đòn.

Phân loại	Định nghĩa
M: Di căn xa (Distant Metastasis)	
M0	Không có di căn xa.
M1a	Có (các) nốt u riêng biệt ở phổi đối bên; nốt màng phổi/màng ngoài tim hoặc tràn dịch màng phổi/màng ngoài tim ác tính.
M1b	Di căn ngoài lồng ngực đơn ổ tại một cơ quan duy nhất (bao gồm cả di căn hạch không thuộc hệ thống hạch vùng).
M1c	Di căn ngoài lồng ngực đa ổ tại một hoặc nhiều cơ quan.

1.1.4. Vai trò trong lập kế hoạch xạ trị, theo dõi và đánh giá đáp ứng

a) Lập kế hoạch xạ trị

Xạ trị (RT) là phương pháp điều trị mang tính triệt căn dành cho những bệnh nhân ung thư phổi không tế bào nhỏ (NSCLC) giai đoạn sớm (I–II) nhưng không có chỉ định phẫu thuật. Việc ứng dụng ^{18}F -FDG PET mang lại những ý nghĩa quan trọng đối với bác sĩ xạ trị, bởi kỹ thuật này cung cấp các thông tin giá trị giúp tối ưu hóa kỹ thuật xạ, xác định các thể tích bia và kiểm soát mức độ phơi nhiễm phóng xạ của bệnh nhân. Mục tiêu cốt lõi của bác sĩ xạ trị là tối đa hóa hiệu quả điều trị của tia xạ, đồng thời hạn chế tối thiểu liều lượng tác động lên các mô lành xung quanh. Lập kế hoạch xạ trị bao gồm việc xác định chính xác các vùng cần chiếu xạ. Trong đó, xác định thể tích u thô (GTV) là bước đi quyết định, bao gồm các vị trí tổn thương quan sát được trên hình ảnh CT hoặc hình ảnh mô phỏng. Thể tích này sau đó được mở rộng để có được thể tích bia lâm sàng (CTV), bằng cách thêm các lề xung quanh GTV nhằm bao phủ các vùng bệnh lý xâm lấn vi thể (dưới mức lâm sàng).

Tiếp theo, CTV lại được mở rộng thêm một lần nữa để tính tới các sai số về tư thế và sự di động của cơ quan cũng như của bệnh nhân, tạo thành thể tích bia lập kế hoạch (PTV). Mặc dù việc xác định thể tích trên hình ảnh PET đơn thuần có thể gặp khó khăn do độ phân giải thấp và mức độ nhiễu cao, nhưng khi kết hợp với hình ảnh cấu trúc như CT, ^{18}F -FDG PET cung cấp thông tin ưu việt nhất về mức độ lan rộng của khối u. Trên thực tế, PET/CT nên được sử dụng rộng rãi trong lập kế hoạch xạ trị cho NSCLC vì nó cho phép ghi hình ranh giới khối u chính xác hơn đáng kể so với CT đơn thuần [28].

Tầm ảnh hưởng của PET đối với việc lập kế hoạch xạ trị (RT) được thể hiện qua hai khía cạnh chủ chốt: một mặt, nó giúp giảm thiểu liều xạ lên các mô lành xung quanh (trong trường hợp vùng u xác định trên PET nhỏ hơn so với CT); mặt

khác, nó giúp bao phủ cả những vùng chứa tế bào u còn sống nằm ngoài trường chiếu dự kiến (khi PET phát hiện phạm vi u lan rộng hơn so với CT). Nhiều báo cáo đã ghi nhận rằng FDG-PET làm thay đổi đáng kể kết quả phân loại giai đoạn hạch vùng ngực, thường là nhờ khả năng phát hiện được nhiều hạch dương tính hơn so với CT [28].

Thêm vào đó, ghi hình PET/CT còn giúp nâng cao độ chính xác trong việc xác lập ranh giới thể tích bia thông qua kỹ thuật "phác họa cấu trúc sinh học - giải phẫu" (Anatomic Biologic Contour - ABC), được xác định trực tiếp trên các hình ảnh tích hợp PET/CT. Tính ưu việt này đã được minh chứng qua một khối lượng lớn các tài liệu phẫu thuật về độ chính xác của FDG-PET trong phân loại giai đoạn hạch của ung thư phổi không tế bào nhỏ (NSCLC) [28].

b) Theo dõi và đánh giá đáp ứng

Dựa trên nhiều tài liệu hiện nay ủng hộ việc sử dụng ^{18}F -FDG PET để đánh giá đáp ứng điều trị sớm, cũng như nhận thức về những hạn chế vốn có của chẩn đoán hình ảnh giải phẫu, dự thảo các tiêu chuẩn PET cập nhật đã được đề xuất nhằm ứng dụng trong các thử nghiệm lâm sàng và có thể là cả thực hành lâm sàng. Tiền đề của bộ tiêu chuẩn PERCIST 1.0 là đáp ứng của ung thư khi đánh giá qua PET là một biến số liên tục và phụ thuộc vào thời gian. Một khối u có thể được đánh giá vào bất kỳ thời điểm nào trong quá trình điều trị, và mức độ tiêu thụ glucose có thể tăng hoặc giảm so với giá trị nền. Chỉ số SUV nhiều khả năng sẽ biến thiên đối với cùng một khối u và cùng một phác đồ tại các thời điểm khác nhau. Ví dụ, mức độ hấp thu chất đánh dấu của khối u được kỳ vọng sẽ giảm dần theo thời gian nếu điều trị có hiệu quả. Do đó, việc ghi nhận và báo cáo tỷ lệ thay đổi của SUV so với giá trị khởi điểm, cùng với thời điểm thực hiện ghi hình, là những yếu tố tối quan trọng [29].

Số lượng chu kỳ hóa trị tối ưu trước khi thực hiện chụp ^{18}F -FDG PET, cũng như khoảng thời gian lý tưởng từ đợt điều trị cuối cùng đến khi ghi hình, hiện vẫn là những chủ đề gây tranh cãi và có thể thay đổi tùy theo từng phác đồ cụ thể. Qua việc đánh giá các tài liệu, Richard L. Wahl và các cộng sự [29] nhận thấy rằng thời điểm sớm sau điều trị (cụ thể là sau 1 chu kỳ, ngay trước khi bắt đầu chu kỳ kế tiếp) có thể là mốc thời gian hợp lý để theo dõi đáp ứng, nhằm xác định liệu khối u có biểu hiện kháng trị nguyên phát hay không. Trên thực tế, nhiều nghiên cứu — bao gồm nghiên cứu của Avril và cộng sự về ung thư buồng trứng — đã chỉ ra rằng có tới 60%–70% tổng mức giảm chỉ số SUV xảy ra chỉ sau duy nhất một chu kỳ điều trị hiệu quả [30]. Ngược lại, việc chờ đợi cho đến khi kết thúc toàn bộ lộ trình điều trị có thể cung cấp bằng chứng cho thấy tình trạng kháng thuốc đã tồn tại xuyên suốt hoặc mới phát triển trong quá trình điều trị. Việc chụp PET sau kết thúc liệu trình thường được thực hiện như một hình thức đánh giá lại

giai đoạn bệnh để quyết định xem có nên điều trị bổ sung hay can thiệp phẫu thuật hay không.

1.2. Đặc trưng dữ liệu ảnh PET/CT và các tác vụ xử lý cốt lõi

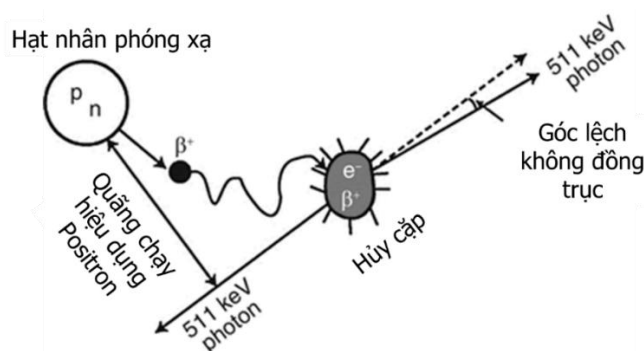
1.2.1. Các yếu tố vật lý ảnh hưởng đến chất lượng ảnh PET/CT

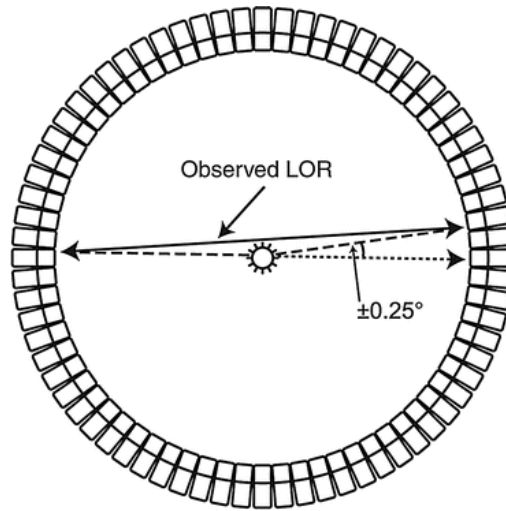
Chất lượng ảnh trong hệ thống chụp ảnh PET/CT phụ thuộc vào rất nhiều yếu tố vật lý phức tạp. Do đây là hệ thống tích hợp phương thức ảnh CT và ảnh PET nên các yếu tố ảnh hưởng sẽ chia thành ba nhóm chính: bản chất vật lý của phân rã phóng xạ, các hiện tượng tương tác bức xạ và các yếu tố kỹ thuật của hệ thống.

a) Yếu tố vật lý liên quan đến dịch chuyển của hạt Positron (β^+)

Bản chất của PET là ghi nhận hai photon hủy cặp (511 keV) phát ra theo hai hướng ngược chiều nhau. Tuy nhiên, trước khi quá trình hủy cặp xảy ra, chuyển động của hạt Positron gây ảnh hưởng đến độ phân giải không gian theo hai cách (Hình 1.5):

- **Quãng chạy của positron (Positron Range):** Sau khi phát ra từ hạt nhân, positron di chuyển một quãng đường ngắn trong mô để tiêu hao động năng trước khi hủy cặp với electron. Vị trí hủy cặp (nơi phát ra photon) lệch một khoảng so với vị trí thực tế, tùy vào hạt nhân phóng xạ. Ví dụ như ^{18}F có năng lượng tối đa thấp (0.63 MeV), quãng chạy trung bình trong nước chỉ khoảng 0.6 mm, cho độ phân giải tốt hơn so với ^{68}Ga (1.90 MeV, quãng chạy ~ 2.9 mm).
- **Độ lệch góc hủy cặp (Non-collinearity: Không đồng trục):** Theo lý thuyết, hai photon phát ra theo hai hướng ngược nhau, tạo góc 180° . Thực tế, do positron vẫn còn một động năng cực nhỏ (mô-men động lượng dư), góc hủy cặp bị lệch khoảng $\pm 0.5^\circ$. Với đường kính vòng detector lớn (ví dụ 80 - 90 cm), độ lệch này có thể làm nhòe ảnh khoảng 1.5 – 2 mm.





Hình 1.5: Ảnh hưởng vị trí và góc lệch không đồng trục tới độ phân giải [31]

b) Các hiện tượng tương tác và nhiễu tia (Coincidence Events)

Hệ thống PET ghi nhận ảnh dựa trên các sự kiện trùng phùng đồng thời (True Coincidences). Tuy nhiên, các tương tác vật lý của photon 511 keV với mô cơ thể sinh ra các loại nhiễu nghiêm trọng:

+ Hiện tượng tán xạ (Scattered Coincidences)

Một hoặc cả hai photon bị tán xạ Compton trong cơ thể, làm thay đổi hướng bay trước khi đập vào detector. Hệ thống vẫn ghi nhận một đường đáp ứng (LOR) nhưng đường này bị sai lệch vị trí, làm giảm độ tương phản của ảnh (Contrast) và tạo ra một vùng mờ (fogging background).

+ Sự trùng phùng ngẫu nhiên (Random Coincidences)

Hai hạt nhân khác nhau hủy cặp gần như cùng lúc. Hai photon không liên quan gì nhau đập vào hai detector đối diện trong cùng một cửa sổ thời gian (Timing Window). Hệ thống ghi nhận nhầm thành một đường đáp ứng LOR giả. Tỷ lệ trùng phùng ngẫu nhiên tăng theo bình phương với liều hoạt độ phóng xạ tiêm vào bệnh nhân (A^2).

Lin và cộng sự [32] đã nghiên cứu mô phỏng Monte Carlo hiệu chỉnh trùng phùng do tán xạ và trùng phùng ngẫu nhiên bằng cách đặt thiết bị Beam Stopper (BS) trên đường đáp ứng (LOR) đi qua vị trí nguồn. Khi đó BS sẽ hấp thụ một tỷ lệ nhất định các sự kiện thực, nhưng hầu như không ảnh hưởng đến các sự kiện tán xạ và ngẫu nhiên. Bằng cách lấy hiệu số giữa hai tập dữ liệu quét – khi có và khi không có BS – ta có thể ước lượng được các sự kiện không thực trên đường LOR. Nghiên cứu sinh cũng đã thực hiện phương pháp tương tự khi mô phỏng

Monte Carlo chuyển động quay của thiết bị xạ trị ART-6000 nhằm xác định chính xác phân bố liều gây bởi nguồn phát tia gamma quay xung quanh trục. [33]

+ Sự đâm xuyên detector (Parallax Effect / Depth of Interaction - DOI)

Khi các photon đi xiên ở vùng rìa của trường nhìn (FOV), chúng có thể xuyên qua cạnh của một thanh tinh thể này và bị hấp thụ ở tinh thể kế cận. Điều này dẫn đến việc xác định sai vị trí LOR, gây ra hiện tượng nhòe ảnh ở rìa FOV (gọi là hiệu ứng Parallax). [34]

c) Sự suy giảm chùm tia (Attenuation)

Khi đi qua cơ thể bệnh nhân, các photon 511 keV bị hấp thụ hoặc tán xạ, dẫn đến việc số lượng photon truyền được ra ngoài giảm theo hàm mũ. Vùng trung tâm cơ thể (hoặc các mô mật độ cao như xương) có sự suy giảm rất lớn, nếu không hiệu chỉnh, ảnh PET sẽ bị tối ở vùng lõi và sáng nhân tạo ở vùng bề mặt da. Giải pháp khắc phục trên hệ thống PET/CT chính hiệu chỉnh suy giảm sử dụng bản đồ hệ số suy giảm góc truyền μ -map) được xây dựng từ chính dữ liệu quét của hệ thống CT. Do năng lượng phát tia của CT thấp hơn (~ 120 kVp), thuật toán phải chuyển đổi ma trận số Hounsfield (HU) sang hệ số suy giảm ở năng lượng 511 keV (Bilinear Scaling). Bất kỳ xảo ảnh (artifact) nào từ CT do kim loại (răng giả, clip phẫu thuật) hoặc chuyển động hô hấp đều sẽ làm sai lệch bản đồ μ -map, dẫn đến hiệu chỉnh quá mức hoặc thiếu hụt trên ảnh PET. [35]

d) Các yếu tố phân cứng của hệ thống Detector

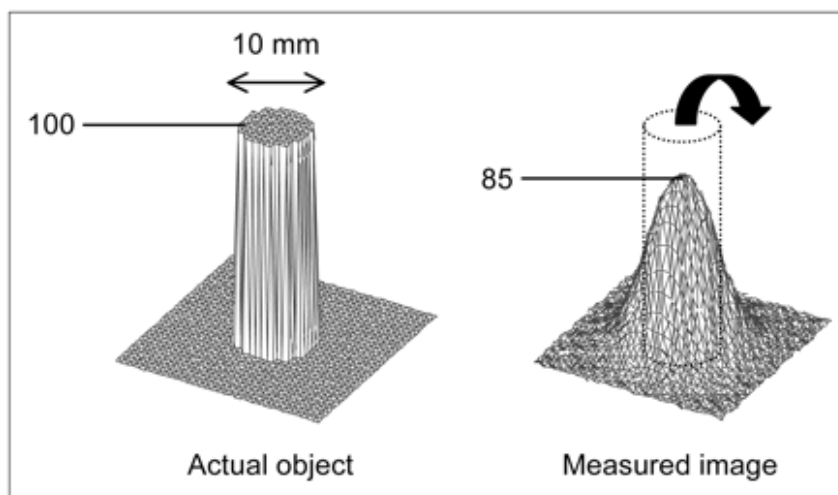
- **Vật liệu tinh thể nhấp nháy (Scintillator):** Các thể hệ máy hiện đại sử dụng tinh thể LSO hoặc LYSO có hiệu suất phát quang cao, thời gian suy giảm (decay time) cực ngắn (~ 40 ns). Điều này cho phép thu hẹp cửa sổ thời gian trùng phùng xuống mức tối đa, giảm nhiễu ngẫu nhiên.
- **Công nghệ ToF (Time-of-Flight):** Khả năng đo chính xác chênh lệch thời gian (Δt) photon đến hai detector đối diện giúp khu trú vị trí hủy cặp dọc theo đường LOR

Cần chú ý độ phân giải thời gian (coincidence time resolution) càng nhỏ (ví dụ < 300 ps), tỷ lệ tín hiệu trên nhiễu (SNR) tăng càng mạnh, đặc biệt có ý nghĩa đối với các bệnh nhân thể trạng lớn (béo phì).

e) Hiệu ứng thể tích bán phần (Partial Volume Effect - PVE)

Do giới hạn về độ phân giải không gian của hệ thống PET (thường từ 4 - 6 mm), các tổn thương có kích thước nhỏ hơn hai lần độ phân giải (FWHM) sẽ chịu ảnh hưởng lớn của PVE:

- **Spill-out:** Hoạt độ phóng xạ từ tổn thương nhỏ bị "tràn" ra vùng nền xung quanh, làm giảm giá trị SUV thực tế (định lượng thấp hơn thực tế) (Hình 1.6).
- **Spill-in:** Hoạt độ từ vùng nền có hoạt độ cao xung quanh tràn ngược vào tổn thương.



Hình 1.6: Nguồn phát hình tròn (đường kính 10 mm) có hoạt độ đồng đều (100 đơn vị tùy ý) trong nền không phóng xạ tạo ra ảnh đo được trong đó một phần tín hiệu phát ra từ nguồn được nhìn thấy bên ngoài nguồn thực tế. Hoạt độ tối đa trong ảnh đo được giảm xuống còn 85. [36]

Để tối ưu hóa chất lượng hình ảnh, các thuật toán tái dựng hình ảnh hiện đại (như OSEM kết hợp mô hình hóa hàm đáp ứng điểm PSF và hiệu chỉnh ToF) luôn phải tích hợp đầy đủ các mô hình toán học giải bài toán ngược cho các hiện tượng vật lý nêu trên.

1.2.2. Các tác vụ xử lý và phân tích hình ảnh PET/CT

Quy trình xử lý ảnh PET/CT (chuyển hóa - giải phẫu) bao gồm một chuỗi các tác vụ tính toán phức tạp. Các tác vụ này trải dài từ giai đoạn tiền xử lý (nhận dạng, chuẩn hóa), tái cấu trúc cho đến phân tích định lượng nâng cao phục vụ lâm sàng và nghiên cứu.

Tác vụ 1. Đăng ký ảnh và hợp nhất ảnh (Image Registration & Fusion)

Đây là tác vụ tiên quyết để kết hợp hai phương thức ảnh có bản chất vật lý khác nhau. Đăng ký ảnh hình học (Rigid & Non-rigid/Deformable Registration): Do PET và CT được chụp trên cùng một hệ thống nhưng không hoàn toàn đồng thời, chuyển động thở hoặc thay đổi tư thế của bệnh nhân có thể gây lệch vị trí. Thuật toán đăng ký ảnh biến dạng (Deformable Image Registration - DIR) giúp nắn chỉnh, chồng khớp các cấu trúc giải phẫu từ CT lên bản đồ chuyển hóa của PET một cách chính xác từng voxel. Dung hợp ảnh (Image Fusion): Trộn thông

tin từ hai kênh màu (ví dụ: CT thang độ xám và PET thang màu giả như Jet hoặc Fire) để tạo ra một ảnh duy nhất hiển thị rõ cả vị trí giải phẫu lẫn mức độ hấp thụ hoạt độ phóng xạ.

Tác vụ 2. Tiền xử lý và Hiệu chỉnh ảnh (Pre-processing & Image Correction)

Mục tiêu của giai đoạn này là đảm bảo chất lượng ảnh PET đạt độ chính xác định lượng cao nhất với một số các tác vụ quan trọng như

- Hiệu chỉnh suy giảm (Attenuation Correction - AC): Tia gamma sinh ra từ quá trình hủy hạt bị hấp thụ một phần bởi các mô bên trong cơ thể (đặc biệt là xương, mô đặc ở lồng ngực). Ảnh CT được chuyển đổi thành bản đồ hệ số suy giảm để hiệu chỉnh lại dữ liệu thô (sinogram) của PET.
- Hiệu chỉnh hiệu ứng thể tích bán phần (Partial Volume Effect Correction - PVEC): Đối với các tổn thương nhỏ (như nốt đơn độc ở phổi), do độ phân giải không gian của PET bị giới hạn, tín hiệu SUV thường bị sụt giảm và lan tỏa ra xung quanh. Các bộ lọc toán học hoặc mô hình học sâu được áp dụng để khôi phục lại giá trị SUV thực tế.
- Khử nhiễu ảnh (Image Denoising): Ảnh PET có bản chất là nhiễu Poisson do số lượng đếm photon bị giới hạn. Các thuật toán như lọc Gaussian, lọc phi tuyến (Non-local Means), hoặc mạng nơ-ron tích chập (CNNs/Autoencoders) được dùng để khử nhiễu mà vẫn giữ được ranh giới tổn thương.

Tác vụ 3. Tái cấu tạo ảnh (Image Reconstruction)

Chuyển đổi dữ liệu thô thu được từ các đầu dò (Sinogram) thành ảnh cắt lớp 3D hiển thị nồng độ phóng xạ. Phương pháp giải tích: Phép chiếu ngược có lọc (FBP - Filtered Back-Projection) – nhanh nhưng nhiễu nhiều. Phương pháp lặp (Iterative Reconstruction): Như MLEM hoặc OSEM (Ordered Subsets Expectation Maximization). Hiện nay, OSEM kết hợp với mô hình hóa hàm đáp ứng điểm (Point Spread Function - PSF) và thời gian bay (Time-of-Flight - TOF) là tiêu chuẩn vàng giúp tăng đáng kể độ phân giải không gian và tỷ lệ tín hiệu trên nhiễu (SNR).

Tác vụ 4. Hiệu chỉnh chuyển động (Motion Correction / Gating)

Đồng bộ nhịp thở/nhịp tim (4D PET/CT Gating): Thu thập tín hiệu hô hấp hoặc điện tâm đồ (ECG) đồng thời với quá trình quét. Dữ liệu PET/CT được phân chia thành các pha (bins) khác nhau của chu kỳ thở để loại bỏ hiện tượng mờ do chuyển động (motion blur), giúp định vị chính xác và không làm phóng đại thể tích u vùng ngực/bụng.

Tác vụ 5. Phân tích định lượng và Trích xuất tính chất (Quantification & Radiomics)

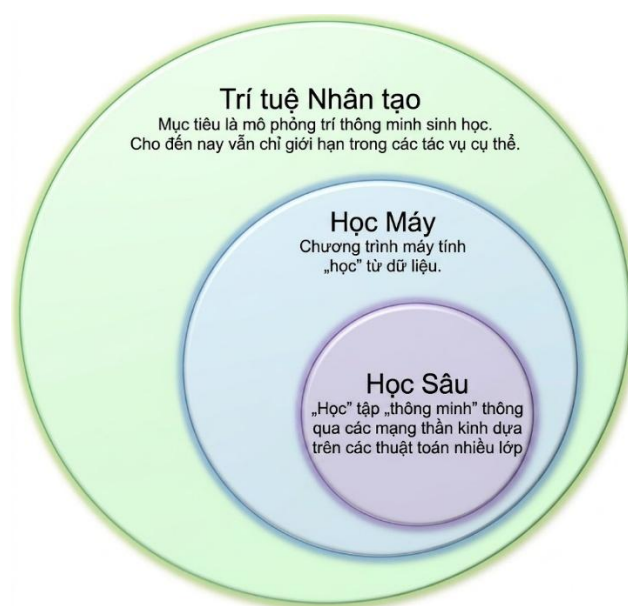
Mục tiêu của tác vụ này là biến đổi dữ liệu hình ảnh thành các chỉ số sinh học có thể đo lường được như:

- Tính toán các chỉ số SUV (Standardized Uptake Value): Tự động tính toán SUV_{max} , SUV_{mean} , SUV_{peak} , cùng các thông số thể tích như: MTV (Metabolic Tumor Volume): Thể tích khối u chuyển hóa. TLG (Total Lesion Glycolysis): Tổng lượng đường phân tổn thương ($TLG = MTV \times SUV_{mean}$).
- Radiomics (Cấu trúc học hình ảnh): Trích xuất hàng trăm đặc trưng toán học từ vùng ảnh PET/CT đã phân đoạn (bao gồm đặc trưng hình dạng, cường độ, và đặc trưng vân ảnh - texture features như GLCM, GLSZM). Các đặc trưng này phản ánh sự không đồng nhất (heterogeneity) của khối u ở cấp độ vi mô, hỗ trợ dự đoán đột biến gene hoặc tiên lượng khả năng sống sót của bệnh nhân.

1.3. Ứng dụng học máy trong xử lý ảnh PET/CT

1.3.1. Nguyên lý ứng dụng của học máy và các mô hình học sâu

Trong khi **Trí tuệ nhân tạo (AI)** về tổng quát dựa trên ý tưởng rằng máy móc có khả năng mô phỏng trí thông minh của con người (ví dụ: giải quyết các vấn đề phức tạp dựa trên logic hoặc các cây quyết định), thì **Học máy (ML)** lại mang tính đặc thù cho từng tác vụ và chỉ tập trung rõ rệt vào quá trình học cho các nhiệm vụ cụ thể đó. Theo đó, quá trình học này chỉ phục vụ duy nhất một mục đích là cải thiện kết quả đạt được cho tác vụ đã thiết kế ([Hình 1.7](#)).



Hình 1.7: Trí tuệ nhân tạo, học máy và học sâu

ML dựa trên các thuật toán nằm ở điểm giao thoa giữa thống kê, khoa học máy tính và AI. Do thiếu đi khía cạnh "trí tuệ" tổng quát, nó chỉ có thể xử lý dữ

liệu có cấu trúc và bán cấu trúc (ngoại trừ học sâu - deep learning). ML tập trung vào hai khía cạnh liên kết chặt chẽ với nhau: thứ nhất là tạo ra các mô hình dựa trên tính toán có khả năng tự động cải thiện thông qua quá trình huấn luyện; thứ hai, ML cũng giải quyết các quy luật thống kê và tính toán nền tảng chi phối các hệ thống học tập. Dưới góc độ toán học, DL có thể được xem là phương pháp xấp xỉ hàm. Nhiệm vụ của chúng là xấp xỉ một ánh xạ thực F — hàm số nhận đầu vào là một hình ảnh x (hoặc nhiều hình ảnh nếu có dữ liệu đa phương thức) và trả về một kết quả đầu ra cụ thể y , thông qua phương trình $y = F(x)$. Định nghĩa về y sẽ thay đổi tùy thuộc vào từng loại công nghệ, vốn được xác định bởi ứng dụng hoặc nhiệm vụ cụ thể.

Ở cấp độ toàn diện, một hệ thống học tập có thể được định nghĩa là một quá trình tìm kiếm giải pháp nhằm cải thiện các thước đo hiệu suất nhất định khi thực hiện một nhiệm vụ cụ thể. Một mô hình học máy có thể được hình dung như một mạng lưới phức tạp của các nút liên kết lẫn nhau. Việc thiết lập một mô hình ML bao gồm hai giai đoạn sử dụng dữ liệu khác nhau: bước đầu tiên, **dữ liệu huấn luyện (training data)** được dùng để đào tạo mô hình. Sau khi mô hình đã thiết lập xong các tham số nội tại, một bộ **dữ liệu kiểm thử (test dataset)** chưa xác định sẽ được sử dụng ở bước thứ hai để kiểm chứng mô hình. Cuối cùng, mô hình mới được đưa vào ứng dụng trên dữ liệu thực tế. Ba phương pháp học máy (ML) được sử dụng rộng rãi nhất là học có giám sát, học không giám sát và học tăng cường. Ngoài ra, còn tồn tại các phương pháp lai như học bán giám sát hoặc học đa thực thể (multi-instance learning).

Học sâu (Deep Learning) là một nhánh chuyên sâu của học máy, được xây dựng dựa trên các mạng thần kinh nhân tạo. Tương tự như các ứng dụng học máy khác, các mạng lưới này bao gồm một tầng đầu vào (input layer) để tiếp nhận dữ liệu và một tầng đầu ra (output layer) để đưa ra kết quả. Điểm khác biệt so với học máy truyền thống nằm ở tính "thông minh" của học sâu: nhiều tầng xử lý như vậy được xếp chồng lên nhau, chứa đựng các mô-đun đơn giản nhưng có tính phi tuyến. Mỗi tầng gồm các nút liên kết sẽ biến đổi dữ liệu từ tầng trước đó thành một biểu diễn ở cấp độ cao hơn, trừu tượng hơn một chút, từ đó hình thành nên những hàm toán học cực kỳ phức tạp [37].

Hơn thế nữa, học sâu sử dụng các thuật toán tích hợp sẵn để điều chỉnh chương trình, giúp nó tự thích nghi các tham số nội tại nhằm tính toán biểu diễn ở mỗi tầng từ biểu diễn của tầng phía trước. Sự kết hợp giữa nhiều tầng xử lý cùng bản chất thích nghi và đệ quy đã cho phép chương trình học được các biểu diễn dữ liệu với nhiều cấp độ trừu tượng khác nhau thông qua quá trình hiệu chỉnh lặp lại [37] — đây chính là yếu tố tạo nên sức mạnh vượt trội của học sâu. Đáng chú ý, lượng dữ liệu đầu vào càng lớn, chương trình càng trở nên tối ưu [38]. Với khối

lượng dữ liệu số không lồ đang ngày càng sẵn có, các mô hình học sâu cũng theo đó mà phát triển mạnh mẽ. Khác với các kỹ thuật học máy khác, học sâu còn có khả năng xử lý dữ liệu phi cấu trúc mà không cần đến các bước tiền xử lý phức tạp thường thấy [39].

Nhờ tập trung vào khả năng "tự học" thay vì xây dựng mô hình thủ công, việc sử dụng DL trong chẩn đoán hình ảnh đánh dấu một bước chuyển mình quan trọng, tách biệt hẳn so với các phương pháp tiếp cận trước đây. Hãy lấy học sâu có giám sát (supervised deep learning) làm ví dụ. Giả sử chúng ta có một bộ dữ liệu huấn luyện $\{(x_n, y_n); n = 1, \dots, N\}$ và một mạng thần kinh nhân tạo được tham số hóa bởi θ - bao gồm các yếu tố như số lượng lớp, số lượng nút (node) trong mỗi lớp, trọng số kết nối, các lựa chọn hàm kích hoạt, v.v. Khi đó, mạng thần kinh được tìm ra để xấp xỉ hàm F có thể được viết dưới dạng $\phi_{\hat{\theta}}(x)$, trong đó $\hat{\theta}$ là các tham số giúp tối thiểu hóa giá trị của cái gọi là hàm mất mát (loss function).

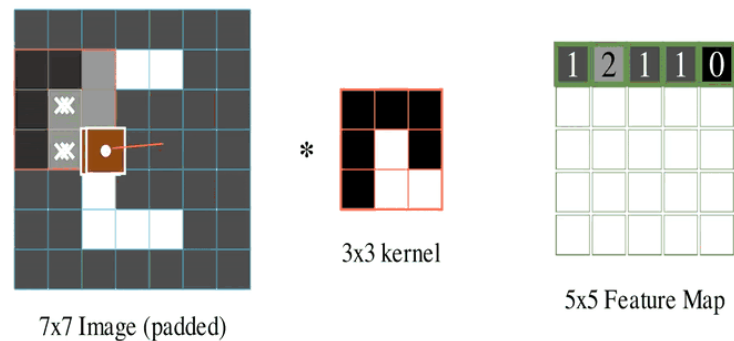
$$L(\theta) = \frac{1}{N} \sum_{n=1}^N l(\phi_{\theta}(x_n), y_n) + R_1(\phi_{\theta}(x_n)) + R_2(\theta) \quad (2)$$

Ở đây, $l(\phi_{\theta}(x_n), y_n)$ là hàm mất mát theo từng phần tử (item-wise loss function) dùng để phạt các sai số dự báo; $R_1(\phi_{\theta}(x_n))$ phản ánh niềm tin tiên nghiệm (prior belief) về kết quả đầu ra, và $R_2(\theta)$ là một số hạng chính quy hóa (regularization term) đối với các tham số của mạng. Mặc dù mạng thần kinh dạng $\phi_{\hat{\theta}}(x)$, đại diện cho một dạng mô hình, nhưng nó thường được xem là một "hộp đen" (black box) vì nó không đại diện cho một mô hình được thiết kế dựa trên các nguyên lý vật lý hoặc toán học chặt chẽ.

Mạng tích chập (Convolutional Neural Network - CNN) là một trong những kiến trúc học sâu (Deep Learning) tiên tiến và phổ biến nhất hiện nay. Nó được thiết kế đặc biệt để xử lý các dữ liệu có cấu trúc lưới hai chiều, điển hình là **hình ảnh**. Thay vì xem một bức ảnh là một dải phẳng các điểm ảnh độc lập, CNN có khả năng hiểu được mối quan hệ không gian giữa các pixel cạnh nhau, từ đó tự động nhận diện các đặc trưng từ đơn giản (cạnh, góc) đến phức tạp (hình dáng, vật thể).

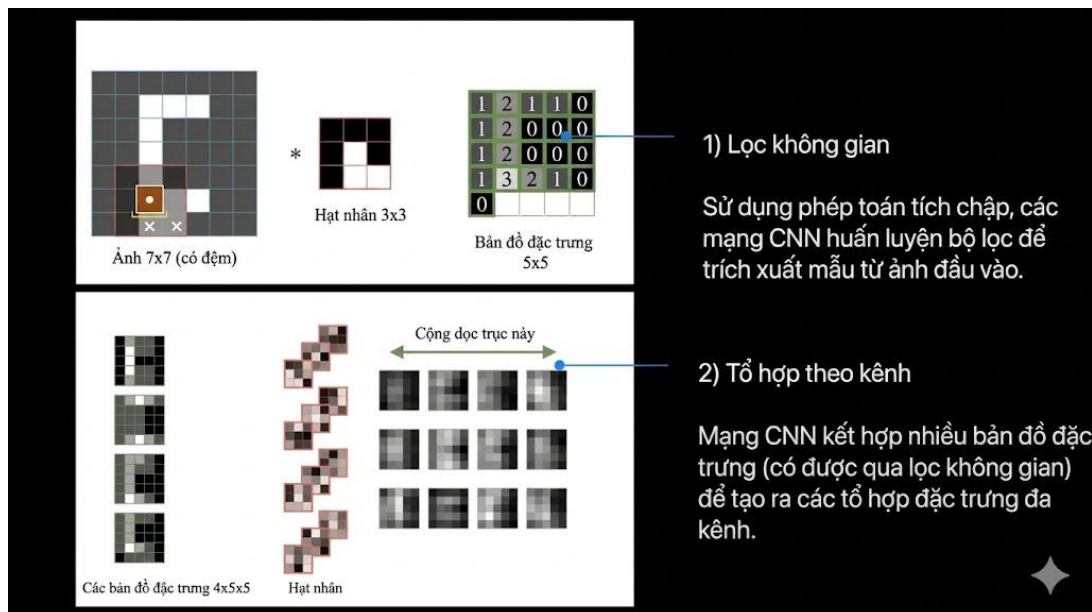
1.3.2. Sự phát triển các mô hình học sâu sử dụng mạng tích chập

Trái tim của một mạng CNN chính là **phép toán tích chập (convolution operation)** thực hiện quét một bộ lọc (filter) trên khắp bức ảnh và tính tích vô hướng của bộ lọc đó với vùng hình ảnh tại mỗi vị trí chồng lấn. Kết quả đầu ra thu được gọi là **bản đồ đặc trưng (feature map)**; nó cho biết mẫu hình của bộ lọc xuất hiện ở đâu và xuất hiện bao nhiêu lần trong bức ảnh (Hình 1.8).



Hình 1.8: Nguyên lý của mạng tích chập

Trong một tầng tích chập, chúng ta huấn luyện nhiều bộ lọc khác nhau để trích xuất các bản đồ đặc trưng đa dạng từ hình ảnh đầu vào. Khi chúng ta xếp chồng nhiều tầng tích chập liên tiếp nhau, kết hợp cùng các hàm phi tuyến, chúng ta sẽ tạo nên một **mạng thần kinh tích chập (CNN)**. Như vậy, mỗi tầng tích chập thực hiện đồng thời hai nhiệm vụ: 1) **Lọc không gian (Spatial filtering)**: Thông qua phép toán tích chập giữa hình ảnh và các hạt nhân (kernels) và 2) **Kết hợp đa kênh (Combining multiple input channels)**: Tổng hợp các kênh đầu vào để tạo ra một tập hợp các kênh đầu ra mới. Có đến 90% các nghiên cứu về CNN tập trung vào việc sửa đổi hoặc cải thiện chính hai khía cạnh này (Hình 1.9).



Hình 1.9: Hai nhiệm vụ chính của mạng tích chập

Việc ứng dụng học sâu (deep learning) vào bài toán phân loại hình ảnh là lĩnh vực khởi nguồn sớm nhất và cũng là mảng đề tài phát triển rực rỡ nhất hiện

nay. Trong số đó, mạng nơ-ron tích chập (CNN) là cấu trúc được sử dụng rộng rãi nhất. Kể từ khi Krizhevsky và các cộng sự đề xuất mô hình AlexNet dựa trên kiến trúc CNN vào năm 2012 và giành ngôi quán quân tại cuộc thi ImageNet năm đó, kỷ nguyên học sâu chính thức bước vào giai đoạn bùng nổ mạnh mẽ. Tiếp nối đà phát triển, năm 2013, Lin và các cộng sự đã giới thiệu cấu trúc Network in Network (NIN), ứng dụng kỹ thuật gộp trung bình toàn cục (global average pooling) nhằm giảm thiểu rủi ro quá khớp (overfitting) [40]. Đến năm 2014, cả GoogLeNet và VGGNet đều thiết lập những kỷ lục mới về độ chính xác trên tập dữ liệu ImageNet. Trong đó, GoogLeNet đã liên tục cải tiến qua các phiên bản v2, v3 và v4 để tối ưu hóa hiệu năng xử lý. Để khắc phục nhược điểm của CNN trong việc yêu cầu kích thước đầu vào cố định, He và các cộng sự đã đề xuất mô hình gộp phân cấp không gian (Spatial Pyramid Pooling - SPP), giúp tăng cường tính bền vững cho dữ liệu đầu vào. Khi các mô hình học sâu ngày càng trở nên phức tạp hơn, để giải quyết hiện tượng suy giảm hiệu năng khi tăng độ sâu của mạng, nghiên cứu sinh của He đã giới thiệu mạng dư (ResNet), tạo ra một bước tiến đột phá đưa công nghệ học sâu vươn tới những tầm cao mới [40].

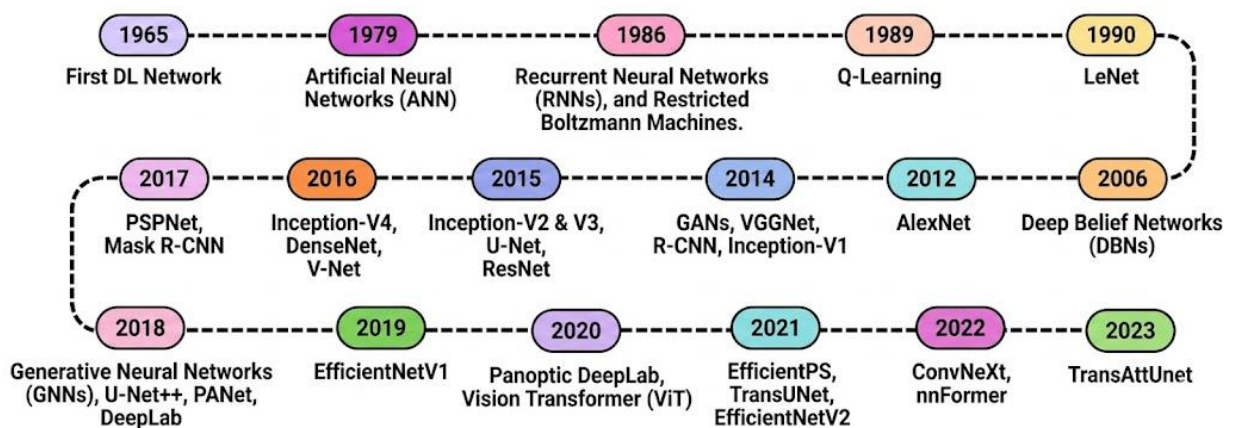
Lấy AlexNet làm ví dụ điển hình. Vào năm 2012, mạng này đã áp dụng cấu trúc 8 lớp, bao gồm 5 lớp tích chập (convolutional layers) và 3 lớp kết nối đầy đủ (fully connected layers). Sau mỗi phép tính tích chập tại 5 lớp đầu, một lớp gộp cực đại (maximum pooling) được thực hiện nhằm giảm thiểu khối lượng dữ liệu. AlexNet tiếp nhận dữ liệu đầu vào có kích thước 227×227 pixel. Trải qua 5 vòng tích chập và gộp, ma trận đặc trưng kích thước $6 \times 6 \times 256$ cuối cùng được chuyển đến các lớp kết nối đầy đủ. Tại lớp thứ sáu, mạng thiết lập 4.096 nhân tích chập và thu được giá trị đặc trưng tuyến tính có kích thước 4.096 thông qua kỹ thuật dropout. Sau hai lớp cuối cùng, chúng ta thu được đầu ra gồm 1.000 dữ liệu kiểu số thực dấu phẩy động (float), tương ứng với kết quả dự báo cuối cùng. Tỷ lệ lỗi của AlexNet tại cuộc thi ImageNet chỉ là 15,3%, vượt xa vị trí thứ hai với 26,2%. Đồng thời, mạng này đã chứng minh tính hiệu quả vượt trội khi sử dụng hàm kích hoạt ReLU thay vì hàm sigmoid truyền thống [40].

Mạng VGG16 lần đầu tiên được đề xuất bởi nghiên cứu sinh VGG thuộc Đại học Oxford. So với AlexNet, mạng này sử dụng liên tiếp nhiều nhân tích chập kích thước nhỏ 3×3 để thay thế cho các nhân lớn như 11×11 hay 5×5 . Đối với một vùng tiếp nhận (receptive field) nhất định, việc sử dụng chuỗi các nhân tích chập nhỏ đem lại hiệu quả tốt hơn so với một nhân lớn; bởi lẽ, cấu trúc đa lớp phi tuyến giúp tăng độ sâu của mạng, đảm bảo học được các quy luật (pattern) phức tạp hơn trong khi chi phí tính toán lại thấp hơn đáng kể. Ra mắt cùng năm với VGGNet, GoogLeNet cũng đạt được những kết quả rất ấn tượng. Khác biệt lớn nhất của GoogLeNet so với VGGNet nằm ở việc thiết kế một mô-đun mang tên Inception.

Đây là một cấu trúc dày đặc (dense structure) kết hợp một lượng nhỏ các nhân tích chập với nhiều kích thước khác nhau, đồng thời ứng dụng lớp tích chập 1×1 để tối ưu hóa và giảm thiểu khối lượng tính toán [40].

Phân đoạn hình ảnh đã trải qua một hành trình tiến hóa ngoạn mục, từ những phương pháp luận Máy học (ML) truyền thống cho đến trạng thái phát triển bùng nổ chưa từng có trong kỷ nguyên Học sâu hiện nay. Ban đầu, các kỹ thuật phân đoạn hình ảnh chủ yếu dựa trên các thuật toán máy học cổ điển, nơi mà các đặc trưng được thiết kế thủ công (hand-crafted features) và các quy tắc dựa trên kinh nghiệm (heuristic) được sử dụng để phân tách hình ảnh thành các vùng khác nhau. Tuy nhiên, sự xuất hiện của Học sâu, đặc biệt là Mạng nơ-ron tích chập (CNN), đã tạo nên một cuộc cách mạng trong phân đoạn hình ảnh bằng cách cho phép tự động trích xuất đặc trưng và học các biểu diễn phân tầng dữ liệu. Sự chuyển đổi mô hình (paradigm shift) này đã mở ra một kỷ nguyên mới cho phân đoạn hình ảnh, cho phép các mô hình nắm bắt được những mối quan hệ không gian phức tạp và thông tin ngữ cảnh, từ đó mang lại kết quả phân đoạn chính xác và linh hoạt hơn trong nhiều ứng dụng khác nhau.

Hình 1.10 minh họa lộ trình phát triển của các mô hình học sâu (DL) trong phân đoạn hình ảnh y tế, phản ánh những cải tiến liên tục về mặt chất lượng cũng như tính cấp thiết của công nghệ này.



Hình 1.10: Quá trình phát triển các mô hình học sâu trong phân đoạn hình ảnh [41]

Theo thời gian, những bước tiến này đã nâng cao đáng kể năng lực của các mô hình học sâu. Khả năng xử lý từ đầu đến cuối (end-to-end), xử lý không gian 3D và học chuyển giao (transfer learning) đã mang lại kết quả phân đoạn chính xác vượt trội; trong khi các phương pháp như giám sát yếu, tự giám sát và mô hình học máy kết hợp (ensemble modeling) đã giải quyết hiệu quả các thách thức về sự khan hiếm dữ liệu có nhãn cũng như sự mất cân bằng dữ liệu giữa các lớp. Khả năng triển khai trong thời gian thực, học liên tục và tính thích ứng cao đã gia tăng giá trị ứng dụng lâm sàng, nơi các mô hình như U-NET, U-NET++ và FCN

Theo số liệu từ [Bảng 1.2](#), các kiến trúc được ưa chuộng nhất trong các nghiên cứu này lần lượt là U-net (36 nghiên cứu), ResNet (20 nghiên cứu) và GAN (16 nghiên cứu). Đáng chú ý, có 12 bài báo đã ứng dụng kỹ thuật học chuyển giao (transfer learning) và 9 bài báo đưa đặc trưng Radiomics vào quy trình nghiên cứu.

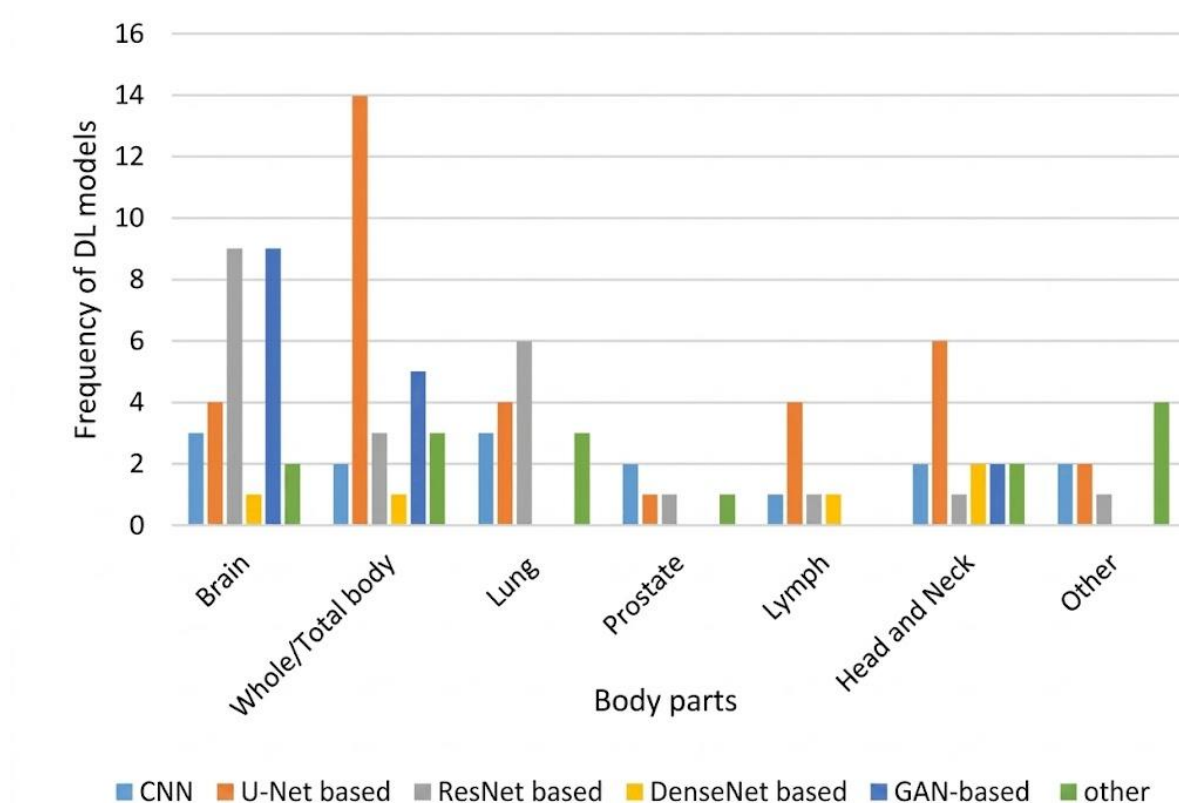
Bảng 1.2: Số lượng các nghiên cứu về các loại ung thư và bệnh ác tính khác nhau

Cơ quan/Bệnh ác tính được nghiên cứu	Số lượng nghiên cứu	Mô hình chiếm ưu thế	Nhiệm vụ được thực hiện nhiều nhất	Loại dữ liệu chính
Phổi (Lung)	15	6 dựa trên ResNet, 4 dựa trên U-net, 3 dựa trên CNN	9 phân loại (dự đoán), 6 phân đoạn	15 nội bộ (private), 1 công khai (public)
Não và các bệnh liên quan (Brain and related disease)	24	9 dựa trên GAN, 9 dựa trên ResNet, 4 dựa trên U-net	10 tạo ảnh, 7 phân loại, 5 phân đoạn, 4 hiệu chỉnh suy giảm	21 nội bộ, 4 công khai
Toàn thân (Whole body/Total body)	26	14 dựa trên U-net, 5 dựa trên GAN, 3 dựa trên ResNet	7 hiệu chỉnh suy giảm, 7 tạo ảnh, 4 khử nhiễu	26 nội bộ, 2 công khai
Vú (Breast)	2	1 dựa trên CNN, 1 dựa trên U-net	1 phân loại, 1 phân đoạn	2 nội bộ, 0 công khai
Đầu và cổ (Head and neck)	12	6 dựa trên U-net, 2 GAN, 2 dựa trên CNN, 2 DenseNet	6 phân đoạn, 3 phân loại, 2 tạo ảnh	12 nội bộ, 1 công khai
U lympho (Lymphoma)	6	4 dựa trên U-net, 1 dựa trên CNN, 1 dựa trên ResNet, 1 DenseNet	3 phân đoạn, 4 phân loại	6 nội bộ, 0 công khai
Tuyến tiền liệt (Prostate)	5	2 dựa trên CNN, 1 ResNet, 1 dựa trên U-net	4 phân loại, 1 phân đoạn	5 nội bộ, 0 công khai
Khác (Others)	9	2 dựa trên U-net, 2 dựa trên CNN	5 phân loại, 3 phân đoạn, 1 khử nhiễu	9 nội bộ, 0 công khai

Đối với các nghiên cứu về não: Kiến trúc ResNet chiếm ưu thế với 9 nghiên cứu, theo sau là dòng U-net với 4 nghiên cứu. Đối với các nghiên cứu sử dụng hình ảnh toàn thân (không phân biệt loại bệnh ác tính): U-net là mô hình phổ biến nhất (14 nghiên cứu), vượt xa các mô hình GAN (5 nghiên cứu) và ResNet (3 nghiên cứu). Trong lĩnh vực nghiên cứu về phổi: ResNet dẫn đầu với 6 nghiên cứu, tiếp đến là U-net (4 nghiên cứu) và CNN (3 nghiên cứu). Số lượng nghiên cứu về các cơ quan khác có phần khiêm tốn hơn. Cụ thể, lĩnh vực ung thư vú chỉ ghi nhận 2 nghiên cứu, mỗi bài sử dụng một mô hình khác nhau (CNN và U-net). Với vùng đầu và cổ, trong tổng số 12 nghiên cứu, U-net vẫn là lựa chọn hàng đầu (6 nghiên

cứ), theo sau là GAN, CNN và DenseNet với cùng 2 nghiên cứu cho mỗi loại. Trong khi đó, với u lympho (6 nghiên cứu), mô hình U-net áp đảo với 4 nghiên cứu, các mô hình CNN, ResNet và DenseNet chỉ xuất hiện rải rác trong 1 nghiên cứu mỗi loại. Đối với tuyến tiền liệt, CNN là kiến trúc được dùng nhiều nhất (2 nghiên cứu), tiếp theo là ResNet và U-net (mỗi loại 1 nghiên cứu). Cuối cùng, với các loại bệnh ác tính khác, các nhà nghiên cứu chủ yếu tin dùng hai dòng kiến trúc U-net và CNN (mỗi dòng 2 nghiên cứu).

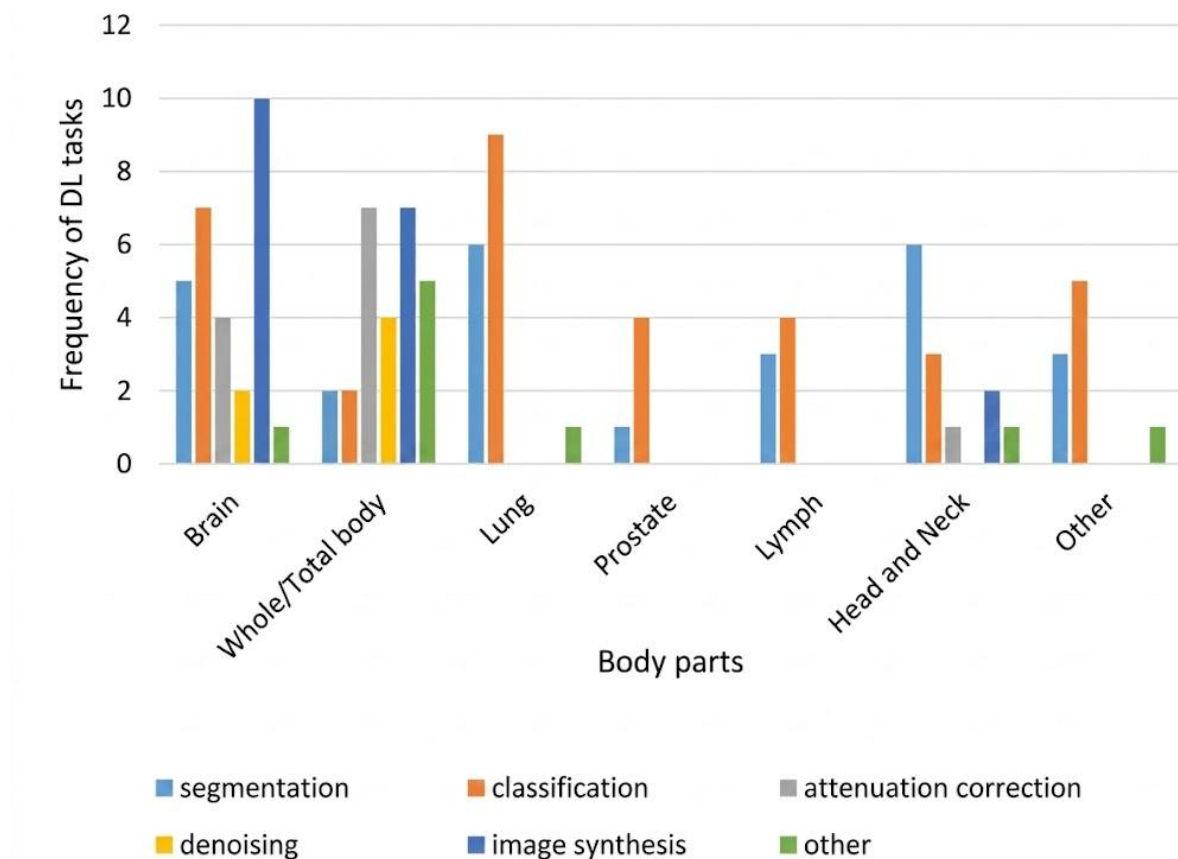
Dưới đây là các biểu đồ thống kê phân loại theo từng bộ phận cơ thể được nghiên cứu. Quan sát các biểu đồ tại [Hình 1.12](#) và [1.13](#), có thể thấy rõ U-net chính là sự lựa chọn ưu tiên cho nhiều cơ quan khác nhau trong phần lớn các nghiên cứu. Bên cạnh đó, các mô hình dựa trên ResNet và CNN cũng được áp dụng rộng rãi, minh chứng cho khả năng thích ứng linh hoạt của chúng trong các ứng dụng đa dạng. Các mô hình dựa trên GAN theo sau với tỷ lệ sử dụng thấp hơn nhóm kể trên nhưng vẫn giữ một vị trí đáng kể. Ngược lại, các loại mô hình học sâu (DL) khác ít được tận dụng hơn và tỷ lệ áp dụng có sự biến thiên tùy theo mục tiêu của từng nghiên cứu cụ thể.



Hình 1.12 Tần suất sử dụng mô hình

Khi so sánh các nhiệm vụ học sâu giữa các cơ quan được nghiên cứu, chúng ta có thể rút ra những nhận định sau: Phân loại (Classification) là nhiệm vụ phổ biến ở tất cả các cơ quan, đạt tỷ lệ cao nhất trong các nghiên cứu về tuyến tiền liệt (80%).

Phân đoạn (Segmentation) cũng là một tác vụ thường gặp, chiếm tỷ trọng lớn nhất ở vùng đầu và cổ (46%) và u lympho (43%). Tổng hợp hình ảnh (Image synthesis) mang tính đặc thù hơn, tập trung chủ yếu ở não (42%) và toàn thân (25%). Hiệu chỉnh suy giảm (Attenuation correction) là nhiệm vụ trọng tâm đối với hình ảnh toàn thân (29%) và vùng đầu cổ (8%). Khử nhiễu (Denoising) tập trung nhiều vào các nghiên cứu toàn thân (14%) và các loại bệnh ác tính khác (11%). Riêng đối với vú và tuyến tiền liệt, mỗi cơ quan này chỉ ghi nhận duy nhất một loại nhiệm vụ nghiên cứu.



Hình 1.13 Tần suất các tác vụ của mô hình

1.4. Đối tượng và mục tiêu nghiên cứu của luận án

1.4.1 Đối tượng Radiomics PET/CT: Ưu điểm và hạn chế tồn tại

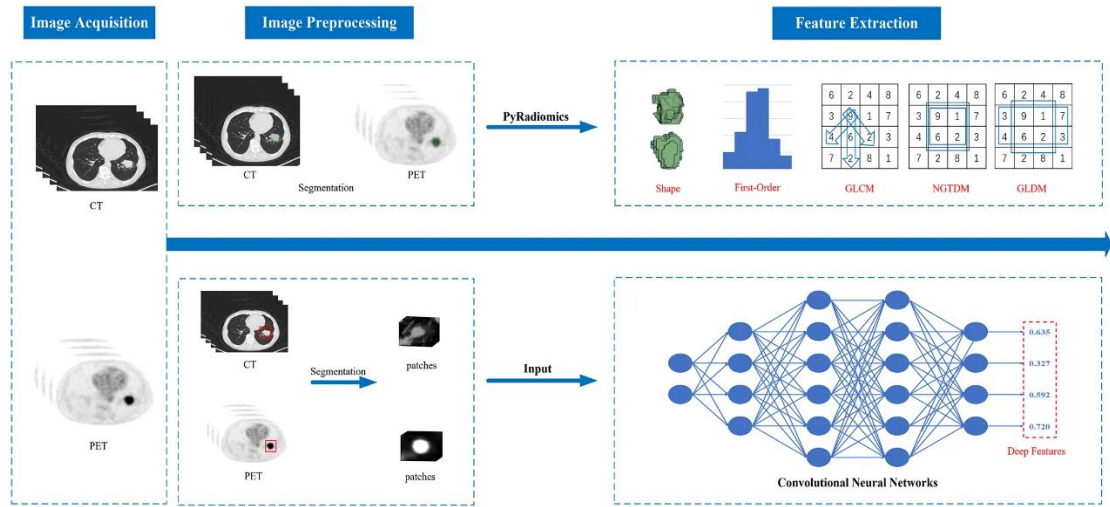
Ung thư phổi không tế bào nhỏ (NSCLC) là nguyên nhân hàng đầu gây tử vong do ung thư trên toàn cầu, với tỷ lệ mắc bệnh đang gia tăng đáng kể tại các quốc gia châu Á, trong đó có Việt Nam. Sàng lọc và chẩn đoán sớm nốt phổi (PNs) có ý nghĩa quyết định trong giảm tỷ lệ tử vong do ung thư phổi. Trong chẩn đoán lâm sàng hiện đại, kỹ thuật chụp cắt lớp phát xạ positron/chụp cắt lớp vi tính ^{18}F -FDG (PET/CT) đóng vai trò quyết định trong việc xác định giai đoạn ung thư và lập kế hoạch xạ trị nhờ sự kết hợp hài hòa giữa các chi tiết giải phẫu từ ảnh CT và các đặc tính chuyển hóa từ ảnh PET. Sự tích hợp đa phương thức này đã được

chứng minh là giúp cải thiện độ chính xác chẩn đoán từ 20% đến 30% so với việc chỉ sử dụng ảnh CT. Dù độ nhạy rất cao ($SEN = 0,89$), phương pháp này có độ đặc hiệu thấp ($SPE = 0,70$), dẫn đến tỷ lệ dương tính giả cao do khó phân biệt nốt ác tính với tổn thương lành tính như u hạt hay viêm phổi [43]. Hơn nữa, việc phân đoạn khối u 3D thủ công vẫn là một thách thức lớn: quy trình này không chỉ tiêu tốn nhiều thời gian, mang nặng tính chủ quan, mà còn gặp phải sự khác biệt đáng kể giữa các người đánh giá khác nhau. Tình trạng bất nhất này ảnh hưởng trực tiếp đến độ tin cậy của các dấu ấn sinh học định lượng như thể tích khối u chuyển hóa (MTV) và tổng lượng phân hủy đường của tổn thương (TLG) — những chỉ số mang tính quyết định trong việc tiên lượng kết quả điều trị của bệnh nhân.

Trong những năm gần đây, **Radiomics** đã nổi lên như một công nghệ phân tích hình ảnh y tế mang tính đột phá. Cùng với sự phát triển thần tốc của công nghệ radiomics và học sâu, vai trò của phương pháp học sâu kết hợp radiomics PET/CT trong thực hành lâm sàng ngày càng trở nên rõ nét. Thông qua việc khai thác chuyên sâu và phân tích chính xác nguồn dữ liệu hình ảnh khổng lồ, công nghệ này có thể thiết lập các chỉ dấu chẩn đoán không xâm lấn mang giá trị cao. Điều này không chỉ giúp bệnh nhân tránh được những đau đớn và rủi ro từ các ca sinh thiết không cần thiết, mà còn kéo giảm đáng kể tỷ lệ xảy ra các biến chứng liên quan đến sinh thiết. Đồng thời, phương pháp này cũng cung cấp một cơ sở vững chắc cho các quyết định điều trị lâm sàng, giúp bệnh nhân có được chẩn đoán chính xác và can thiệp kịp thời ngay từ giai đoạn sớm của bệnh, từ đó nâng cao hiệu quả tỷ lệ chữa khỏi. Ngoài ra, kỹ thuật này còn giảm thiểu các ca kiểm tra lặp lại không đáng có cũng như rút ngắn thời gian theo dõi dài hạn.

Khái niệm này lần đầu tiên được đề xuất vào năm 2012 bởi học giả người Hà Lan – Lambin [44]. Về cốt lõi, **Radiomics** là quá trình trích xuất hiệu năng cao các đặc trưng định lượng từ hình ảnh y tế (như CT, MRI, PET,...) và chuyển đổi dữ liệu hình ảnh tại vùng quan tâm (**Region of Interest - ROI**) thành dữ liệu đa chiều có khả năng khai phá. Các đặc trưng này bao gồm cường độ điểm ảnh, hình dạng, thể tích, đặc trưng cấu trúc,... Radiomics giúp trích xuất đặc trưng định lượng đa chiều từ vùng quan tâm (ROI). Khác với đặc trưng thủ công (HF) dễ sai số do con người, đặc trưng sâu (**Deep Features - DF**) được trích xuất tự động qua các mạng thần kinh tích chập (CNN) (Hình 1.14). Học sâu (**Deep Learning - DL**) vận dụng các mạng thần kinh truyền thẳng đa tầng (**multi-layer feedforward neural networks**) để tiếp nhận trực tiếp dữ liệu hình ảnh đầu vào và thực hiện huấn luyện đầu cuối (**end-to-end**) trong môi trường có giám sát, từ đó tự học các đặc trưng hình ảnh có tính phân loại cao. Quy trình xây dựng mô hình DL bao gồm các giai đoạn: thu nhận hình ảnh, tiền xử lý, huấn luyện mô hình và phân tích. DF có khả năng biểu diễn các thuộc

tính phi tuyến phức tạp ở cấp độ cao, nâng cao độ tin cậy khi phân loại nốt phổi [45] [46] [47] [48] [49].



Hình 1.14 Sự khác biệt về quy trình giữa việc trích xuất các đặc trưng thủ công (handcrafted features) và các đặc trưng sâu (deep features).

Bảng 1.3 tổng hợp năm nghiên cứu tiêu biểu đã ứng dụng DF vào việc chẩn đoán bản chất lành tính và ác tính của các nốt phổi sử dụng ảnh PET/CT.

Bảng 1.3: Các nghiên cứu tiêu biểu về ứng dụng mô hình học sâu sử dụng dữ liệu ảnh PET/CT

Tác giả	Dữ liệu bệnh nhân	Mô hình học sâu
Park và cộng sự [54]	359	ResNet-18
Shao và cộng sự [55]	106	3D-CNN, LR
Lai và cộng sự. [53]	112	HRNet, ResNet
Zhang và cộng sự [56]	174	RF, ResNet, DCNN
Alves và cộng sự [52]	113	3D-CNN

Trong chẩn đoán bản chất lành tính và ác tính của các nốt phổi (PNs), bên cạnh các mô hình 2D truyền thống, các mô hình 2.5D và 3D cũng đã được phát triển mạnh mẽ. Mô hình 2.5D đưa các hình ảnh 2D từ ba mặt phẳng: đứng dọc (**sagittal**), đứng ngang (**coronal**) và trục ngang (**axial**) vào ba kênh để xây dựng mô hình [50]. Trong khi đó, các mô hình 3D tích hợp các lớp cắt liên tục của nốt phổi thành một thực thể ba chiều, giúp tăng cường đáng kể khả năng trích xuất thông tin hình khối [51]. Chẳng hạn, nghiên cứu của Alves và cộng sự [52] đã sử dụng kiến trúc 3D-CNN; dù chỉ sử dụng đầu vào là hình ảnh PET, hiệu suất chẩn đoán của nó vẫn vượt xa mô hình 2D ResNet-50 (AUC: 0,839 so với 0,774). Lai và cộng sự [53] đã đề xuất kiến trúc Mạng độ phân giải cao 3D (**3D HRNet**), một thiết kế loại bỏ các lớp bước nhảy (**stride**) và lớp gộp (**pooling**) để duy trì kích thước nguyên bản của các đặc trưng, cho kết quả chính xác hơn so với ResNet (AUC: 0,781 so với 0,650). Mặc dù các mô hình 3D

thường chiếm ưu thế so với 2D, nhưng trong một số trường hợp, mô hình 2.5D lại cho kết quả chính xác hơn [50], điều này có thể xuất phát từ hiện tượng quá khớp (**overfitting**) thường gặp ở các mô hình 3D.

Phương pháp định lượng phóng xạ dựa trên học sâu (DLR) là sự kết hợp giữa học sâu (DL) và radiomics truyền thống. Trong khi các đặc trưng thủ công (HF) gặp nhiều khó khăn để biểu diễn toàn diện tính dị thể của khối u, làm giới hạn hiệu năng của các mô hình chẩn đoán, thì DLR đã bù đắp được khuyết điểm này bằng cách trích xuất các đặc trưng sâu hơn và có số chiều cao hơn thông qua các lõi tích chập (convolutional kernels), từ đó định lượng tính dị thể của khối u một cách toàn diện hơn [57]. Cùng với sự tiến bộ của công nghệ y học hạt nhân, hàng loạt nghiên cứu đã chứng minh rằng các mô hình radiomics đa phương thức PET/CT cho hiệu năng vượt trội hơn hẳn so với các mô hình chỉ sử dụng đơn độc ảnh CT hoặc PET trong chẩn đoán định tính các nốt ở phổi (PNs) [58] [55]. Bên cạnh đó, các kỹ thuật như ghi hình đa chất đánh dấu (chẳng hạn như kết hợp ^{18}F -FLT và ^{18}F -FDG [59] và ghi hình tại hai thời điểm (DTPI) [46] [60] cũng giúp nâng cao rõ rệt độ chính xác của chẩn đoán.

Ngoài ra, các nghiên cứu cho thấy việc kết hợp giữa HF và các đặc trưng sâu (DF) có thể nâng cao rõ rệt hiệu năng của mô hình [56]. Ví dụ, Zhang và các cộng sự [56] đã bóc tách 100 đặc trưng HF và 2048 đặc trưng DF thu được thông qua mạng ResNet từ hình ảnh PET và CT để phân biệt các nốt lao với ung thư phổi. Kết quả thực nghiệm chỉ ra rằng mô hình mạng thần kinh tích chập sâu nhận đồng thời cả dữ liệu đầu vào HF và DF (Radiomics-DCNN, AUC: 0,820) cho hiệu năng vượt trội hơn hẳn so với mô hình DL chỉ nhận đầu vào DF (AUC: 0,720) và mô hình học máy truyền thống chỉ nhận đầu vào HF (AUC: 0,760). Mặc dù DLR đã chứng minh được tiềm năng to lớn của mình, song các công trình nghiên cứu trong lĩnh vực này hiện vẫn còn khá khan hiếm, và cần có thêm các nghiên cứu chuyên sâu hơn để xác chuẩn tính khả thi cũng như hiệu năng thực tế của phương pháp này.

Hạn chế về dữ liệu: Số lượng và tính đa dạng

Mặc dù HF và DF áp dụng các phương pháp khác nhau trong kỹ thuật radiomics, song mục tiêu và kịch bản ứng dụng của chúng lại tương tự nhau. Với những tiến bộ của công nghệ học sâu (DL), DF thường mang lại hiệu năng chẩn đoán vượt trội hơn HF, với điều kiện là tập dữ liệu huấn luyện phải đủ lớn. Đặc trưng HF rất dễ bị ảnh hưởng bởi những biến thiên trong quá trình phân đoạn thủ công và các tham số máy quét; ngược lại, các mô hình DL giúp cải thiện tính bền vững (robustness) và khả năng tổng quát hóa (generalization) bằng cách tích hợp một cách hữu cơ khâu trích xuất đặc trưng và khâu phân loại thông qua quá trình huấn luyện đầu-cuối (end-to-end) [61]. Tuy nhiên, DF lại có khả năng giải thích

(interpretability) yếu hơn, đòi hỏi khối lượng dữ liệu lớn hơn và tiêu tốn chi phí tính toán cao hơn.

Vấn đề mất cân bằng dữ liệu là một thách thức lớn trong các nghiên cứu radiomics phục vụ chẩn đoán PNs lành tính và ác tính, dễ dẫn đến kết quả chẩn đoán bị thiên lệch (biased). Mặc dù hiện có một vài bộ dữ liệu PET/CT công khai quy mô lớn (ví dụ: AutoPET, LUNG-PET-CT-Dx), song chúng thường thiếu sự xác thực toàn diện về mặt mô bệnh học, hoặc chưa được tối ưu hóa để đánh giá hiện tượng dịch chuyển miền dữ liệu (domain shift) giữa các sắc tộc đa dạng [62]. Một số chiến lược hiệu quả đã được đề xuất áp dụng để giải quyết bài toán này. Thứ nhất, tăng cường dữ liệu (data augmentation) giúp gia tăng cả tính đa dạng lẫn số lượng của các mẫu thử. Thứ hai, tái lấy mẫu hoặc tăng mẫu (oversampling) ở cấp độ đặc trưng: ví dụ, việc sử dụng kỹ thuật SMOTE (Synthetic Minority Over-sampling Technique) [63] để khởi tạo các mẫu tổng hợp có thể giúp mô hình tránh được hiện tượng quá khớp (overfitting). Cuối cùng, có thể ứng dụng các phương pháp học bán giám sát (semi-supervised learning): chẳng hạn như kỹ thuật gán nhãn giả (pseudo-labeling) — một phương thức phân loại dựa trên cơ chế biểu quyết đối với các mẫu chưa được gán nhãn — giúp tăng cường tính bền vững (robustness) cho mô hình.

Mặc dù sở hữu những bước tiến vượt bậc này, một khoảng cách lớn về khả năng tổng quát hóa (generalization gap) mang tính cốt lõi vẫn đang tồn tại. Hầu hết các nghiên cứu radiomics hiện nay vẫn phụ thuộc vào các thiết kế hồi cứu quy mô nhỏ tại đơn trung tâm (single-center), đồng thời còn thiếu vắng các nghiên cứu tiền cứu (prospective studies) cũng như khâu xác chuẩn độc lập từ bên ngoài (external validation). Hạn chế này dẫn đến việc các mô hình có khả năng tổng quát hóa bị giới hạn. Để tối ưu hóa hiệu năng và nâng cao tính khả thi lâm sàng của các mô hình này, việc thiết lập các cơ sở dữ liệu quy mô lớn thông qua sự hợp tác đa trung tâm (multi-center), đồng thời sử dụng các tập xác chuẩn bên ngoài để đánh giá mô hình là bài toán cấp thiết đang được cộng đồng nghiên cứu tập trung giải quyết. Phương pháp tiếp cận này sẽ củng cố mạnh mẽ khả năng tổng quát hóa cùng độ tin cậy của mô hình. Dẫu vậy, phần lớn các bộ dữ liệu huấn luyện hiện nay đều có nguồn gốc từ các quần thể phương Tây, dấy lên những lo ngại sâu sắc về tính bền vững (robustness) của thuật toán khi triển khai trên các bệnh nhân châu Á — những người vốn sở hữu các đặc điểm di truyền, đặc tính bệnh lý và phân phối giá trị hấp thụ chuẩn hóa (SUV) hoàn toàn khác biệt [64].

Hạn chế về cấu trúc mạng: Tự động phân đoạn và kiểm tra chéo

Cấu trúc mạng CNN truyền thống hoạt động dựa trên các bộ lọc tích chập (convolution kernels) quét qua các vùng lân cận cục bộ (local). Mặc dù việc xếp chồng nhiều lớp giúp mở rộng trường thụ cảm, khả năng nắm bắt thông tin ngữ cảnh toàn cục (global context) của CNN vẫn bị giới hạn. Ảnh PET thường có độ

phân giải không gian thấp nhưng cung cấp thông tin chức năng (sự trao đổi chất) quan trọng như vùng hấp thu phóng xạ cao (SUV - Standardized Uptake Value). Nếu thiếu thông tin ngữ cảnh toàn cục để liên kết mối quan hệ giữa các cấu trúc giải phẫu ở xa (từ ảnh CT) với vùng tổn thương (trên ảnh PET), CNN dễ phân đoạn sai các vùng hấp thu sinh lý tự nhiên (như não, tim, bàng quang) thành vùng ung thư di căn.

Để giảm chi phí tính toán và trích xuất các đặc trưng ở cấp độ cao, các mạng CNN (như U-Net, V-Net) sử dụng các lớp Max-pooling hoặc tích chập với bước nhảy (Strided Convolution). Tuy nhiên, quá trình này làm mất đi các chi tiết không gian ở độ phân giải mịn (fine-grained spatial details). Trong phân đoạn hình ảnh PET/CT, đặc biệt là các tổn thương hoặc hạch bạch huyết có kích thước rất nhỏ (vài milimét) trên ảnh CT, việc mất mát thông tin này khiến mô hình gặp khó khăn lớn trong việc xác định chính xác đường biên (boundary) hoặc bỏ sót hoàn toàn tổn thương nhỏ.

Việc kết hợp dữ liệu từ hai nguồn ảnh khác nhau hoàn toàn về bản chất vật lý (CT phản ánh mật độ cấu trúc giải phẫu, PET phản ánh hoạt động chuyển hóa sinh học) là một thách thức lớn. CNN thường chọn giải pháp hợp nhất sớm (Early fusion - ghép kênh đầu vào) hoặc hợp nhất muộn (Late fusion - ghép ở các tầng đặc trưng sâu). Đối với giải pháp Early fusion, mô hình coi các kênh PET và CT như các kênh màu (RGB) thông thường, bỏ qua mối tương quan phi tuyến tính phức tạp giữa cấu trúc giải phẫu và chức năng sinh học. Đối với giải pháp Late fusion, do các lớp tích chập xử lý độc lập ở các giai đoạn đầu, mô hình có thể bỏ lỡ các đặc trưng bổ trợ cho nhau (complementary features) ở mức độ thấp và trung bình, dẫn đến kết quả phân đoạn không tối ưu tại các vùng biên tổn thương có sự bất đối xứng giữa hình ảnh giải phẫu và chuyển hóa.

Trọng số của các bộ lọc tích chập được áp dụng như nhau trên toàn bộ ảnh hoặc bản đồ đặc trưng, không có khả năng tự động tập trung mạnh mẽ vào mối quan hệ giữa các voxel ở khoảng cách xa nhau mà không phụ thuộc vào độ sâu của mạng. Trong ảnh PET/CT, khối u nguyên phát và các hạch vùng hoặc di căn xa có mối quan hệ bệnh lý chặt chẽ. CNN thuần túy không có khả năng mô hình hóa tường minh mối quan hệ phụ thuộc tầm xa (long-range dependencies) này, dẫn đến việc phân đoạn rời rạc và thiếu tính nhất quán mang tính hệ thống của ca bệnh. Hơn nữa, CNN học các mẫu kết cấu (texture) cục bộ rất mạnh mẽ. Ảnh PET vốn có bản chất là độ phân giải thấp, tỷ lệ tín hiệu trên nhiễu (SNR) kém và dễ bị ảnh hưởng bởi chuyển động hô hấp của bệnh nhân (gây nhòe ảnh). Trong khi đó, ảnh CT có thể bị nhiễu do các dị vật kim loại (răng giả, nẹp xương). Các bộ lọc tích chập cục bộ của CNN rất dễ bị đánh lừa bởi các vùng nhiễu này, dẫn đến kết quả phân đoạn xuất hiện nhiều vùng dương tính giả (false positives).

I.4.2 Mục tiêu và bài toán nghiên cứu

Nhằm đề xuất giải pháp phù hợp giải quyết những hạn chế nêu trên, nội dung nghiên cứu của luận án tập trung vào ba mục tiêu chính sau:

- (1) Thiết lập cơ sở dữ liệu ảnh chụp PET/CT người Việt bản địa quy mô lớn
- (2) Xây dựng công cụ đọc và tiền xử lý ảnh nhằm trích xuất dữ liệu ảnh DICOM phục vụ huấn luyện mô hình học sâu
- (3) Phát triển nền tảng học sâu tự động phân đoạn ảnh chụp PET/CT hỗ trợ công tác chẩn đoán ung thư phổi

Để hiện thực hóa mục tiêu nghiên cứu, luận án tiến hành thu thập, xử lý và đánh giá chuẩn đối sánh một bộ dữ liệu quy mô lớn mang tính bản địa, bao gồm 416 ca chụp PET/CT từ một nhóm bệnh nhân Việt Nam [65]. Tính mới của bộ dữ liệu nằm ở ba đóng góp cốt lõi: (1) toàn bộ các ca bệnh NSCLC đều được xác chuẩn một cách chắc chắn thông qua sinh thiết giải phẫu bệnh; (2) quy trình dán nhãn nghiêm ngặt bởi nhiều người đánh giá độc lập với sự tham gia của ba bác sĩ giàu kinh nghiệm, kết hợp cùng thuật toán đồng thuận xác suất STAPLE nhằm giảm thiểu định kiến chủ quan [66] [67]; và (3) việc triển khai các quy trình tiền xử lý chuyên biệt nhằm tối đa hóa tỷ số tín hiệu khối u trên nền (tumor-to-background ratio). Việc tiến hành nghiên cứu trên bộ dữ liệu này không chỉ nâng cao độ tin cậy của mô hình, mà còn đóng góp vào những nỗ lực xóa bỏ sự bất bình đẳng về tính đa dạng sắc tộc trong quá trình phát triển AI y tế toàn cầu. Điều này được hiện thực hóa thông qua việc thiết lập các chuẩn đối sánh cho những thuật toán tiên tiến nhất hiện nay trên bộ dữ liệu PET/CT bản địa của Việt Nam [62] [65] [66] [67] [68], đánh giá hiện tượng dịch chuyển miền dữ liệu và năng lực tổng quát hóa chéo (cross-generalization) thông qua việc thử nghiệm trên các bộ dữ liệu công khai quốc tế [69] [70], và phân tích mối tương quan giữa các chỉ số thể tích do AI dự báo với kết quả lâm sàng thực tế, tuân thủ nghiêm ngặt các hướng dẫn đánh giá sinh y mới nhất [71].

Luận án nghiên cứu lựa chọn các kiến trúc học sâu cho nền tảng tự động phân đoạn hình ảnh PET/CT không chỉ nhằm mục đích đạt được hiệu năng cao nhất, mà còn để thiết lập một hệ thống đánh giá chuẩn đối sánh (benchmarking) toàn diện. Hệ thống này giúp đánh giá khả năng thích ứng của các thuật toán tiên tiến đối với các đặc điểm nhân trắc học đặc thù và phân phối hình ảnh riêng biệt của bệnh nhân Việt Nam. Trong đó, nnU-Net v2 được sử dụng làm mô hình cơ sở (baseline) nhờ khả năng tự cấu hình vượt trội dựa trên đặc trưng dữ liệu, cho phép tự động xác định các tham số huấn luyện mà không cần đến sự can thiệp thủ công. Với kích thước khối dữ liệu đầu vào (input patch size) là $128 \times 128 \times 128$ voxel, mô hình này tiếp nhận đồng thời dữ liệu đa kênh PET và CT, mang lại độ ổn định cao trong việc xác định các ranh giới giải phẫu [62]. Nhằm khắc phục những hạn

chế của các mạng CNN truyền thống vốn bị ràng buộc bởi trường thụ cảm cục bộ (local receptive fields), kiến trúc Swin UNETR đã được tích hợp để tận dụng cơ chế tự chú ý (self-attention) của mạng Transformer. Bằng cách phân rã dữ liệu 3D thành các token dạng khối nhỏ (patches) và xử lý chúng qua các khối Swin Transformer phân tầng với kích thước cửa sổ $7 \times 7 \times 7$, mô hình này có khả năng mô hình hóa các mối phụ thuộc tầm xa (long-range dependencies) của các tín hiệu y sinh. Đặc tính này đặc biệt phù hợp để nhận diện các khối u NSCLC có kích thước lớn hoặc có hình thái xâm lấn phức tạp trong bộ dữ liệu, từ đó góp phần đạt được độ nhạy (sensitivity) tổng thể cao nhất. Đồng thời, việc triển khai mạng 3D U-Net [72] dựa trên khung làm việc MONAI (phiên bản 1.3 trở lên) [73] kết hợp với mạng xương sống (backbone) ResNet đóng vai trò như một kiến trúc kiểm chứng (architectural control) cho hiệu quả của việc dung hợp đặc trưng ở cấp độ bộ mã hóa (encoder). Sự tương phản về hiệu năng giữa mô hình này và phiên bản nnU-Net tinh chỉnh — vốn đã đạt được bước đột phá với chỉ số Dice là 83,44% — khẳng định rằng các kiến trúc tiêu chuẩn, nếu không trải qua quá trình học chuyển giao (transfer learning) và tinh chỉnh (fine-tuning) trên miền dữ liệu mục tiêu, sẽ rất khó vượt qua được hiện tượng dịch chuyển miền dữ liệu (domain shift).

CHƯƠNG 2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Xây dựng bộ dữ liệu

Mục tiêu chính của việc thu thập dữ liệu ảnh PET/CT là tạo ra một bộ dữ liệu đại diện, chất lượng cao, có nhãn chính xác để huấn luyện, hiệu chỉnh và đánh giá các mô hình trí tuệ nhân tạo (AI), đặc biệt trong chẩn đoán và hỗ trợ ra quyết định lâm sàng đối với ung thư phổi không tế bào nhỏ (NSCLC) và các bệnh lý liên quan. Dữ liệu sẽ được sử dụng để phát triển các mô hình như: Phân đoạn tổn thương (tumor segmentation), Phân loại tổn thương (classification: ác tính/lành tính), Dự đoán giai đoạn bệnh (staging), Ước tính khả năng sống còn hoặc đáp ứng điều trị. Phạm vi dữ liệu bao gồm: Ảnh PET/CT của bệnh nhân đã được chẩn đoán xác định (confirmed diagnosis), ảnh từ hệ thống máy PET/CT các bệnh viện tại các khu vực khác nhau có thể bao gồm cả hình ảnh của bệnh nhân không có bệnh lý phổi ác tính để làm mẫu đối chứng (negative control). Các nguồn dữ liệu được khai thác hợp pháp và đảm bảo tính đại diện: dữ liệu lưu trữ lịch sử của các Bệnh viện, có văn bản đồng ý tham gia nghiên cứu của các Bệnh viện.

Để đảm bảo tính ứng dụng và độ tin cậy của mô hình AI, dữ liệu PET/CT cần đáp ứng các đặc điểm sau: đa dạng về bệnh lý (bệnh nhân ung thư phổi không tế bào nhỏ giai đoạn khác nhau, có di căn xa hay di căn hạch, trường hợp nốt phổi đơn độc, ...), đa dạng về thiết bị (dữ liệu từ các đơn vị khác nhau), đa dạng dân số (phân bố hợp lý độ tuổi, giới tính, hút thuốc hay không hút thuốc, nhóm nguy cơ cao), độ phân giải ảnh PET không thấp hơn 128x128, ảnh CT từ 512x512 (bề dày lớp cắt $\leq 5\text{mm}$), không bị nhiễu mạnh, không có lỗi về trường nhìn hay cắt cụt hình ảnh. Để đảm bảo bảo mật các trường PatientName, PatientID, BirthDate được xóa hoặc mã hóa, dữ liệu được lưu trữ nội bộ, có phân quyền truy cập.

2.1.1. Tiêu chí lựa chọn và loại trừ

Các tiêu chí lựa chọn được xây dựng dựa trên hướng dẫn của các tổ chức như EANM [74], AJCC Cancer Staging Manual [75] và Fleischner Society [76] để đảm bảo tính chính xác và phù hợp với mục tiêu phân loại NSCLC hoặc không phải NSCLC.

- Nhóm NSCLC:
 - Bệnh nhân được chẩn đoán xác định NSCLC qua sinh thiết hoặc giải phẫu bệnh sau mổ, bao gồm các loại mô học như ung thư biểu mô tuyến, ung thư tế bào vảy, ung thư tế bào lớn.
 - Có bệnh sử đầy đủ và bộ dữ liệu hình ảnh PET/CT đầy đủ, chất lượng cao, sử dụng chất đánh dấu phóng xạ ^{18}F -FDG, với độ phân giải tối thiểu 128 x 128 pixel cho PET và 512 x 512 pixel cho CT.
 - Độ tuổi ≥ 18 .

- Nhóm không phải NSCLC:
 - Bệnh nhân có tổn thương phổi nghi ngờ ung thư nhưng được xác nhận không phải NSCLC qua sinh thiết hoặc theo dõi lâm sàng (ví dụ, nốt phổi lành tính, viêm phổi, lao phổi, hoặc u hạt).
 - Có hình ảnh PET/CT tương đương về chất lượng với nhóm NSCLC.
- Tiêu chí loại trừ:
 - Hình ảnh PET/CT có nhiều nghiêm trọng, độ phân giải thấp, hoặc không đồng bộ giữa PET và CT.
 - Thiếu dữ liệu lâm sàng hoặc kết quả giải phẫu bệnh.
 - Bệnh nhân có bệnh lý nền nghiêm trọng (như suy tim, suy thận nặng) hoặc nhiễm trùng cấp tính làm sai lệch kết quả PET/CT.
 - Bệnh nhân đã được điều trị đặc hiệu (bao gồm phẫu thuật, hóa trị, xạ trị, hoặc điều trị bằng thuốc nhắm mục tiêu đối với ung thư biểu mô tuyến phổi trước khi khám PET/CT);
 - Bệnh nhân có nhiều nốt khối u ở phổi hoặc nhiều khối u ở các bộ phận khác của cơ thể;
 - Tổn thương khối u gần trung tâm và không thể tách rời khỏi vùng giải phẫu lân cận;

Theo kết quả giải phẫu bệnh cuối cùng, các trường hợp dữ liệu thu thập được chia thành nhóm dương tính và nhóm âm tính với ung thư không tế bào nhỏ. Nghiên cứu hồi cứu tuân theo một quy trình đã được Hội đồng Đánh giá Tổ chức tại Bệnh viện phê duyệt và không cần sự đồng ý của bệnh nhân. Đây cũng là một trong các tiêu chí chung để lựa chọn bộ dữ liệu đang được sử dụng trong các nghiên cứu hiện nay. Nghiên cứu sinh đã thực hiện việc thu thập bộ dữ liệu có sẵn, công khai về các nghiên cứu Chụp cắt lớp phát xạ Positron/Chụp cắt lớp điện toán (PET/CT) từ năm 2019 đến năm 2023. Tất cả dữ liệu ảnh đều được thực hiện trên máy quét PET/CT hiện đại của Bệnh viện Ung Bướu Hà Nội, Bệnh viện K và Bệnh viện Ung Bướu TP. Hồ Chí Minh. Việc thu thập dữ liệu PET/CT này đã được sự đồng ý bằng văn bản của 03 bệnh viện trên.

Bảng 2.1: Một số tiêu chí định lượng của bộ dữ liệu dùng huấn luyện học máy

STT	Tiêu chí	Mức yêu cầu định lượng
1.	Quy mô dữ liệu	100 – 200 Bệnh nhân cho nghiên cứu ban đầu
2.	Chất lượng ảnh	Độ phân giải ảnh PET không thấp hơn 128x128, ảnh CT từ 512x512 (bề dày lớp cắt $\leq 5\text{mm}$)
3.	Đa dạng thiết bị	Tối thiểu từ 2 hệ thiết bị chụp khác nhau

STT	Tiêu chí	Mức yêu cầu định lượng
4.	Chất lượng gắn nhãn	Có xác nhận từ bác sĩ và thông tin giải phẫu bệnh
5.	Tính đầy đủ của metadata	Đảm bảo $\geq 90\%$ dữ liệu có metadata đầy đủ
6.	Định dạng ảnh	DICOM là định dạng chính, có thể xuất thêm NIfTI/PNG nếu cần.
7.	Đa dạng dân số học	Có ≥ 3 nhóm tuổi, cả nam và nữ.
8.	Tính cập nhật	Dữ liệu trong vòng 3-5 năm tại thời điểm nghiên cứu
9.	Phiếu đồng thuận bệnh nhân	Có sự thông qua của hội đồng y đức hoặc xác nhận đồng thuận bệnh nhân

2.1.2. Quy trình thu thập dữ liệu

2.1.2.1 Quy trình chụp PET/CT

Các hệ thống chụp PET/CT của 3 bệnh viện đều có model là Discovery IQ của GE. Quy trình chụp dựa theo hướng dẫn của Bộ Y tế tại Quyết định 705/QĐ-BYT ngày 28/02/2014



Hình 2.1: Minh họa máy chụp PET/CT

Dữ liệu ảnh PET/CT thu thập tại các bệnh viện được sử dụng trong luận án đều được thực hiện theo quy trình như sau:

Bước 1: Chuẩn bị bệnh nhân

- Bệnh nhân được hướng dẫn nhịn ăn 4-6h trước khi tiêm ^{18}F -FDG để giảm mức đường sinh lý và giảm insulin về ngưỡng cơ bản. Kiểm tra glucose máu $< 11,1 \text{ mmol/l}$. Bệnh nhân sẽ được đặt lại lịch tiêm nếu lượng glucose máu $> 8,3\text{-}11,1 \text{ mmol/l}$. Có thể sử dụng insulin để kiểm soát đường huyết. Đánh giá tính chính xác của người bệnh và các xét nghiệm: đúng người bệnh, đúng chẩn đoán, đúng chỉ định.
- Bác sĩ khám lâm sàng bệnh nhân trước khi chụp.
- **Tiêm được chất phóng xạ:** Bệnh nhân được tiêm tĩnh mạch ^{18}F -FDG với liều $0,10\text{-}0,15 \text{ mCi/kg}$ cân nặng. Sau tiêm, người bệnh nằm nghỉ trong phòng yên tĩnh, hạn chế ánh sáng trong 60 phút. Thuốc phóng xạ nên được tiêm ở vùng đối bên với vị trí tiếp giáp với vùng tổn thương.

Bước 2: Chụp PET/CT

- PET/CT được chụp sau 50- 60 phút sau tiêm thuốc phóng xạ. Bệnh nhân được đưa sang phòng chụp. Đặt người bệnh nằm ngửa thoải mái trên bàn máy PET/CT, tay giơ cao ngang đầu. Hướng dẫn người bệnh không cử động trong quá trình chụp. Vùng khảo sát từ nền sọ đến 1/2 giữa đùi.
- Chụp CT toàn thân liều thấp với các thông số chụp: 120 kVp, 100 mAs, độ dày lát cắt $2,5 - 3,75 \text{ mm}$, FOV 70 cm, bước chuyển bàn (pitch) 1,0 và 0,6 giây/ vòng quay
- Chụp PET/CT toàn thân với tốc độ 1,5 phút mỗi vị trí giường (bed position), 7 – 8 vị trí, FOV là 50 cm, kích thước ma trận 256×256
- Hình ảnh PET/CT được tái tạo ba chiều và hiệu chỉnh sự suy giảm bằng hình ảnh CT, sau đó fusion với CT để tạo thành hình ảnh PET/CT.

Bước 3: Đánh giá kết quả:

- Hình ảnh phân bố sinh lý của ^{18}F -FDG trong cơ thể như hấp thu mạnh ở não, thận, bể thận, niệu quản, bàng quang; hấp thu trung bình ở gan, lách, tủy xương; hấp thu thay đổi ở một số cơ quan như tuyến nước bọt, amidan, tim, ruột, tử cung, buồng trứng.
- Hình ảnh bệnh lý : Các tổn thương tăng hấp thu phóng xạ bất thường tại các mô và cơ quan.

2.1.2.2 Quy trình gán nhãn và phân đoạn dữ liệu

Gán nhãn dữ liệu là một bước quan trọng trong việc xây dựng cơ sở dữ liệu PET/CT để hỗ trợ ứng dụng học máy trong chẩn đoán ung thư phổi, đặc biệt để phân loại tổn thương phổi thành ung thư phổi không tế bào nhỏ (NSCLC) hoặc không phải NSCLC (bao gồm các tổn thương lành tính, viêm phổi, lao phổi, hoặc u hạt). Quá trình gán nhãn được thực hiện dựa trên kết quả giải phẫu bệnh, theo dõi lâm sàng dài hạn, và các thông tin bổ sung như đặc điểm hình ảnh học và xét

nghiệm gen. Nhân dữ liệu cần đảm bảo tính chính xác, nhất quán, và phù hợp với các tiêu chuẩn y tế quốc tế để hỗ trợ huấn luyện các mô hình học máy hiệu quả. Quy trình này được thực hiện bởi một nhóm chuyên gia bao gồm bác sĩ giải phẫu bệnh, bác sĩ lâm sàng chuyên khoa ung thư, và bác sĩ chẩn đoán hình ảnh, với sự đồng thuận để giảm thiểu sai số chủ quan. Các nhân được lưu trữ trong cơ sở dữ liệu có cấu trúc, liên kết với hình ảnh PET/CT và vùng quan tâm (ROI) được phân đoạn bằng phần mềm 3D Slicer.

a) Tiêu chí lựa chọn bác sĩ thực hiện công việc gắn nhãn và phân đoạn dữ liệu

- Tiêu chí lựa chọn bác sĩ để thực hiện việc gắn nhãn (phân loại) dữ liệu và phân đoạn khối u trong nghiên cứu này được thiết kế để đảm bảo tính chính xác, chuyên môn, và nhất quán của quá trình, phù hợp với các tiêu chuẩn y tế quốc tế. Dưới đây là các tiêu chí chi tiết để lựa chọn bác sĩ mà luận án đã sử dụng:
- Bác sĩ phải có ít nhất 5 năm kinh nghiệm trong lĩnh vực y học hạt nhân hoặc chẩn đoán hình ảnh, đã làm việc tại các trung tâm y tế lớn và uy tín, với trọng tâm là chẩn đoán ung thư phổi sử dụng hình ảnh PET/CT. Kinh nghiệm này đảm bảo khả năng diễn giải chính xác các đặc điểm hình ảnh học, như giá trị SUV trên PET và các đặc điểm giải phẫu trên CT.
- Đối với gắn nhãn, bác sĩ cần có kinh nghiệm trong việc đánh giá kết quả giải phẫu bệnh, đặc biệt là phân loại các loại mô học của ung thư phổi không tế bào nhỏ (NSCLC) như ung thư biểu mô tuyến, ung thư biểu mô vảy, hoặc ung thư biểu mô tế bào lớn, theo phân loại của Tổ chức Y tế Thế giới (WHO). Đối với phân đoạn, bác sĩ cần có kinh nghiệm thực hiện phân đoạn khối u trên hình ảnh PET/CT, bao gồm việc nhận diện ranh giới khối u và các dấu hiệu như vôi hóa hoặc dấu hiệu cây cối (tree-in-bud) trên CT [76].
- Bác sĩ tham gia phân đoạn phải được đào tạo chuyên sâu về phần mềm 3D Slicer, bao gồm các module như Segment Editor, GrowCut, DeepGrow, và PET/CT Fusion (nhóm nghiên cứu đào tạo). Đào tạo này cần bao gồm kỹ năng phân đoạn thủ công và bán tự động, cũng như cách sử dụng các công cụ để căn chỉnh hình ảnh PET/CT [77]. Họ cũng cần biết cách lưu trữ và quản lý dữ liệu ở định dạng DICOM, NRRD, hoặc NIFTI. Ưu tiên bác sĩ đã tham gia các khóa đào tạo chính thức về 3D Slicer, ví dụ, thông qua các hội thảo của QIN (Quantitative Imaging Network) hoặc các chương trình đào tạo trực tuyến từ cộng đồng 3D Slicer.
- Bác sĩ cần được đào tạo về radiomics, bao gồm cách trích xuất đặc trưng hình học, cường độ, và texture từ vùng quan tâm (ROI). Điều này đảm bảo rằng họ hiểu tác động của phân đoạn đến chất lượng đặc trưng radiomics, như được mô tả trong nghiên cứu [78] [79].

- Các bác sĩ cần có khả năng làm việc nhóm để tham gia các buổi họp đánh giá đồng thuận, đặc biệt khi có mâu thuẫn trong phân đoạn hoặc gắn nhãn. Kỹ năng giao tiếp và thảo luận đa ngành là cần thiết để giải quyết các trường hợp phức tạp, ví dụ, khi kết quả giải phẫu bệnh không rõ ràng hoặc hình ảnh PET/CT có đặc điểm không điển hình.
- Bác sĩ phải hiểu và tuân thủ các quy định đạo đức y khoa, bao gồm tuyên bố Helsinki và các quy định bảo mật dữ liệu như HIPAA hoặc GDPR, đặc biệt khi xử lý dữ liệu bệnh nhân nhạy cảm [80]. Họ cần có kinh nghiệm làm việc với dữ liệu ẩn danh và đảm bảo rằng các nhãn và ROI không chứa thông tin nhận dạng cá nhân.
- Số lượng bác sĩ được chọn trong luận án là: ba bác sĩ được chọn để thực hiện phân đoạn khối u nhằm đảm bảo sự đồng thuận và giảm sai số chủ quan, như khuyến nghị trong nghiên cứu [81]. Đối với gắn nhãn, ít nhất hai bác sĩ (một bác sĩ giải phẫu bệnh và một bác sĩ lâm sàng) tham gia, với một bác sĩ thứ ba để kiểm tra chéo khi cần.

b) Phân loại tổn thương

Phân loại tổn thương phổi thành NSCLC hoặc không phải NSCLC dựa trên các căn cứ khoa học và lâm sàng, tuân thủ các hướng dẫn từ Hiệp hội Ung thư Quốc tế (UICC), Hiệp hội Y học Hạt nhân Châu Âu (EANM), và Fleischner Society [74] [75] [76]. Các căn cứ chính bao gồm:

- Kết quả giải phẫu bệnh:

- Đối với nhóm **NSCLC**, nhãn được xác định dựa trên kết quả sinh thiết hoặc mẫu phẫu thuật, xác nhận sự hiện diện của các loại mô học đặc trưng của NSCLC, bao gồm ung thư biểu mô tuyến, ung thư biểu mô vảy, hoặc ung thư biểu mô tế bào lớn. Kết quả giải phẫu bệnh được báo cáo theo phân loại mô học của Tổ chức Y tế Thế giới (WHO) về u phổi. Các đặc điểm mô học như mức độ biệt hóa, sự hiện diện của hoại tử, hoặc xâm lấn mạch máu cũng được ghi nhận để hỗ trợ phân loại phụ nhóm NSCLC.
- Đối với nhóm **không phải NSCLC**, kết quả giải phẫu bệnh xác nhận tổn thương là lành tính, viêm hoặc bệnh lý khác (như lao phổi). Sinh thiết được thực hiện trong các trường hợp nghi ngờ. Nếu không có sinh thiết, tổn thương được đánh giá dựa trên theo dõi lâm sàng dài hạn, như mô tả dưới đây.

- Theo dõi lâm sàng (đối với nhóm không phải NSCLC):

- Các tổn thương nghi ngờ ung thư nhưng không có sinh thiết xác nhận được theo dõi từ 6 đến 24 tháng để đánh giá sự tiến triển. Theo hướng dẫn của Fleischner Society, các nốt phổi < 6 mm ở bệnh nhân nguy cơ

thấp (không hút thuốc, tuổi < 40) thường được coi là lành tính và không cần sinh thiết ngay lập tức ([76], 2017). Các nốt lớn hơn (> 6 mm) hoặc có đặc điểm nghi ngờ (như ranh giới mờ, tăng kích thước nhanh) được theo dõi bằng CT định kỳ (3, 6, 12, 24 tháng). Nếu tổn thương không tăng kích thước hoặc có dấu hiệu thoái triển (như vôi hóa, giảm mật độ), nó được phân loại là không phải NSCLC.

- **Đặc điểm hình ảnh học trên PET/CT:**

- o Hình ảnh PET/CT được sử dụng như một yếu tố bổ sung để hỗ trợ phân loại, đặc biệt khi kết quả giải phẫu bệnh hoặc theo dõi lâm sàng chưa rõ ràng. Các tổn thương NSCLC thường có giá trị SUVmax cao (> 2,5) trên ảnh PET, phản ánh hoạt động chuyển hóa mạnh, trong khi các tổn thương lành tính hoặc viêm có thể có SUV thấp hơn [74]. Trên ảnh CT, NSCLC thường biểu hiện ranh giới gai hoặc không đều, trong khi tổn thương lành tính như hamartoma có thể có ranh giới trơn (smooth) hoặc vôi hóa. Dấu hiệu cây cối (tree-in-bud) trên CT gợi ý viêm phế quản hoặc lao phổi, hỗ trợ phân loại không phải NSCLC [76] [82]. Tuy nhiên, đặc điểm hình ảnh chỉ được sử dụng như thông tin tham khảo, không thay thế kết quả giải phẫu bệnh hoặc theo dõi lâm sàng.

c) Quy trình gắn nhãn

Quy trình gắn nhãn được thực hiện một cách có hệ thống để đảm bảo tính chính xác và nhất quán, với sự tham gia của các chuyên gia đa ngành:

- **Bước 1: Thu thập và xem xét dữ liệu:**

- o Nhóm chuyên gia (bác sĩ giải phẫu bệnh, bác sĩ lâm sàng, và bác sĩ chẩn đoán hình ảnh) xem xét kết quả giải phẫu bệnh, báo cáo lâm sàng, và hình ảnh PET/CT của từng bệnh nhân. Dữ liệu được truy xuất từ bệnh án điện tử hoặc hồ sơ y tế, đảm bảo đầy đủ thông tin về sinh thiết, theo dõi lâm sàng, và xét nghiệm bổ sung.
- o Đối với nhóm NSCLC, báo cáo giải phẫu bệnh được kiểm tra để xác nhận loại mô học, phân loại giai đoạn ung thư theo hệ thống AJCC phiên bản 8, và các đặc điểm mô học khác [77].
- o Đối với nhóm không phải NSCLC, báo cáo theo dõi lâm sàng và kết quả xét nghiệm (vi sinh, hình ảnh học) được xem xét để xác nhận tính chất lành tính hoặc không ác tính.

- **Bước 2: Gắn nhãn chính:**

- o Mỗi bệnh nhân được gắn nhãn chính dựa trên phân loại nhị phân: NSCLC hoặc **không phải NSCLC**. Nhãn này được xác định dựa

trên kết quả giải phẫu bệnh (ưu tiên hàng đầu) hoặc theo dõi lâm sàng (khi không có sinh thiết).

- **Bước 3: Đánh giá sự đồng thuận:**

- Để giảm sai số chủ quan, các nhãn được xem xét bởi ít nhất hai chuyên gia độc lập (một bác sĩ giải phẫu bệnh và một bác sĩ lâm sàng). Nếu có mâu thuẫn (ví dụ, phân loại không rõ ràng do thiếu dữ liệu sinh thiết), nhóm chuyên gia họp để thảo luận, sử dụng thêm thông tin từ hình ảnh PET/CT hoặc xét nghiệm bổ sung.
- Sự đồng thuận được đánh giá bằng chỉ số **Cohen's Kappa** để đo mức độ thống nhất giữa các chuyên gia, với ngưỡng chấp nhận $Kappa \geq 0,7$ [83].

- **Bước 4: Kiểm tra chéo và xác nhận:**

- Nhãn được kiểm tra chéo với dữ liệu lâm sàng (bệnh án, báo cáo sinh thiết, kết quả theo dõi) để đảm bảo không có sai sót. Một mẫu ngẫu nhiên 10-15% trường hợp được kiểm tra bởi một chuyên gia thứ ba để xác nhận tính chính xác.
- Các trường hợp không rõ ràng (ví dụ, thiếu sinh thiết và theo dõi < 6 tháng) được loại bỏ khỏi cơ sở dữ liệu để đảm bảo chất lượng.

- **Bước 5: Lưu trữ nhãn:**

- Nhãn được lưu trong cơ sở dữ liệu SQL hoặc NoSQL, liên kết với hình ảnh PET/CT, ROI (được phân đoạn bằng 3D Slicer), và đặc trưng radiomics thông qua mã định danh bệnh nhân ẩn danh (UUID).
- Định dạng nhãn tuân thủ chuẩn như CSV hoặc JSON, hỗ trợ tích hợp với các công cụ học máy như TensorFlow hoặc scikit-learn.

d) Căn cứ xác định vị trí và thể tích khối u

Thể tích khối u NSCLC được xác định thông qua quá trình phân đoạn vùng quan tâm (ROI) trên hình ảnh PET/CT, sử dụng các công cụ như 3D Slicer. Các căn cứ chính bao gồm ngưỡng chuyển hóa, ranh giới giải phẫu, và các thuật toán phân đoạn, với các chi tiết như sau:

- **Ngưỡng chuyển hóa trên ảnh PET:**

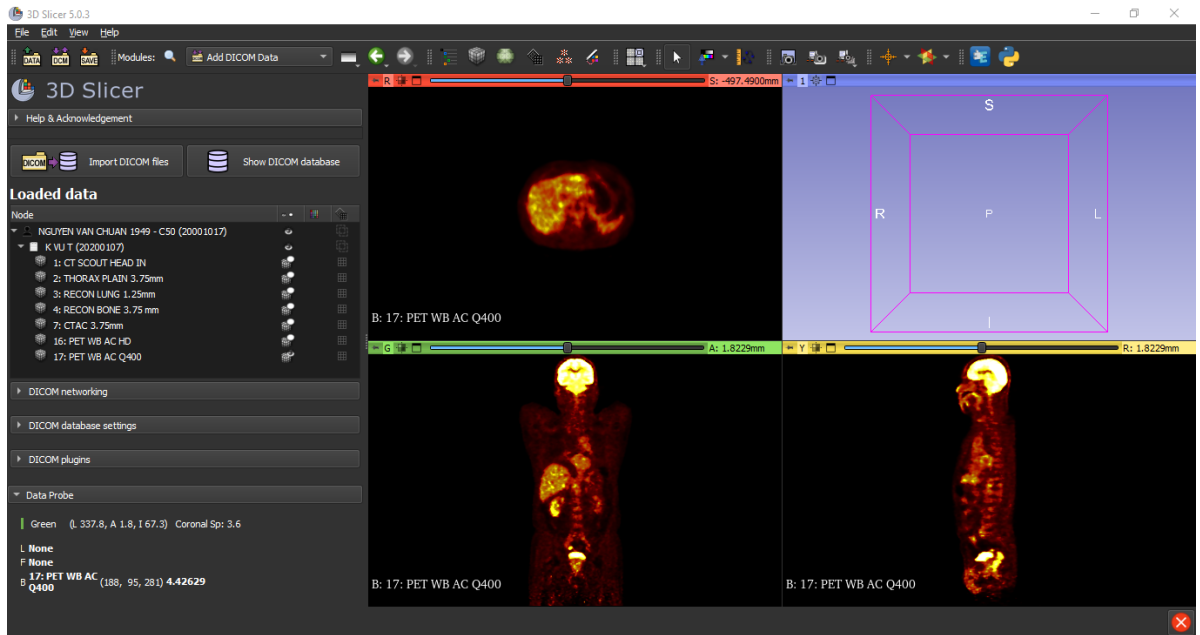
- **Ngưỡng SUV:** Theo hướng dẫn EANM, khối u được phân đoạn dựa trên ngưỡng $SUV \geq 2,5$ hoặc $40\% SUV_{max}$ để xác định ranh giới khối u ác tính. Ngưỡng này giúp loại bỏ mô lành hoặc vùng viêm có SUV thấp hơn, đảm bảo thể tích khối u phản ánh đúng vùng ung thư hoạt động [74].

- **SUVmax và SUVmean:** SUVmax được sử dụng để xác định điểm chuyển hóa cao nhất trong khối u, từ đó thiết lập ngưỡng phân đoạn. SUVmean cung cấp thông tin về mức độ chuyển hóa trung bình, hỗ trợ đánh giá thể tích khối u đồng nhất hay không đồng nhất [81].
- **Hiệu chỉnh nhiễu:** PET sử dụng thuật toán như tối ưu hóa kỳ vọng với tập con có thứ tự (Ordered Subset Expectation Maximization-OSEM) để hiệu chỉnh nhiễu nền, đảm bảo rằng thể tích khối u không bị phóng đại do tín hiệu từ mô xung quanh [74].
- **Ranh giới giải phẫu trên ảnh CT:**
 - CT cung cấp hình ảnh độ phân giải cao để xác định ranh giới khối u dựa trên mật độ HU và đặc điểm giải phẫu. Khối u NSCLC thường có ranh giới không đều, với HU > 30, khác với mô phổi lành (HU khoảng -700 đến -900). Các đặc điểm như xâm lấn màng phổi hoặc thành ngực được sử dụng để điều chỉnh ranh giới phân đoạn [76].
 - Kích thước khối u trên CT (đường kính tối đa, thể tích) được đo để hỗ trợ phân loại T theo hệ thống AJCC phiên bản 8, ví dụ, T1a (≤ 1 cm), T2a (> 3 cm nhưng ≤ 4 cm), hoặc T3 (> 5 cm nhưng ≤ 7 cm) [75].

e) Quy trình phân đoạn khối u thủ công

Phân đoạn khối u (segmentation) là bước quan trọng để xác định vùng quan tâm (ROI) phục vụ cho việc trích xuất đặc trưng radiomics và huấn luyện mô hình học máy. Quy trình này cần đảm bảo tính chính xác và đồng thuận giữa các bác sĩ, dựa trên các tiêu chuẩn quốc tế. Trong nghiên cứu này, phần mềm 3D Slicer được chọn do tính mã nguồn mở, khả năng tích hợp các module phân đoạn tiên tiến, và sự phổ biến trong nghiên cứu y học [77].

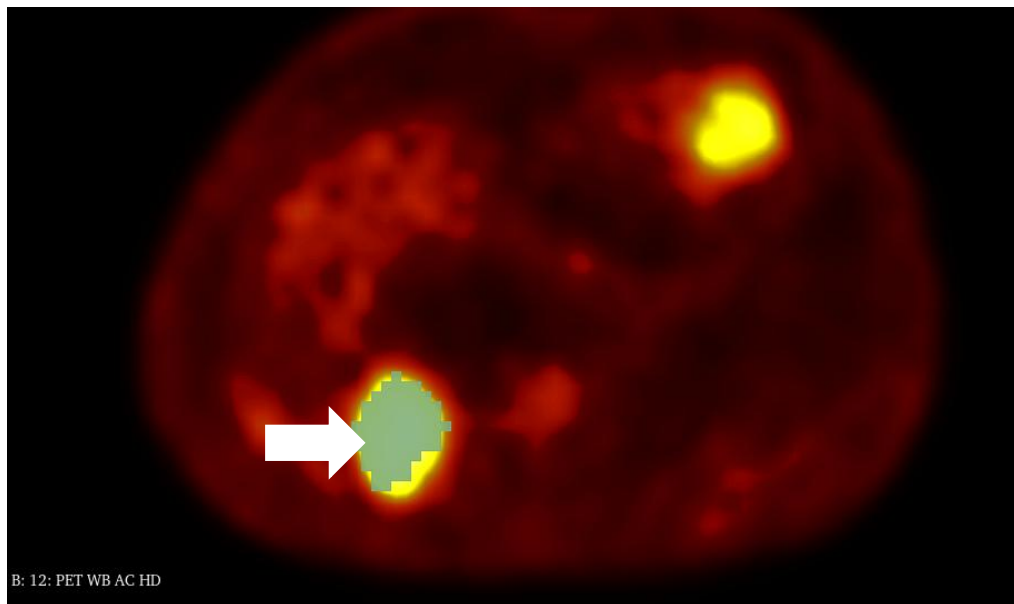
- **Công cụ và module trong 3D Slicer:**
 - Segment Editor: Hỗ trợ phân đoạn thủ công và bán tự động, với các công cụ như brush, eraser, và thresholding.
 - GrowCut: Thuật toán phân đoạn bán tự động dựa trên cạnh tranh vùng, phù hợp với khối u có ranh giới rõ trên CT [84].
 - DeepGrow: Module học sâu tích hợp để hỗ trợ phân đoạn nhanh, đặc biệt với khối u có ranh giới phức tạp trên PET.
 - PET/CT Fusion: Module căn chỉnh và trực quan hóa đồng thời PET và CT để hỗ trợ khoanh vùng chính xác.



Hình 2.2. Giao diện phần mềm 3D-Slicer

Bước 1: Phân đoạn độc lập từ thông tin trên ảnh PET và CT:

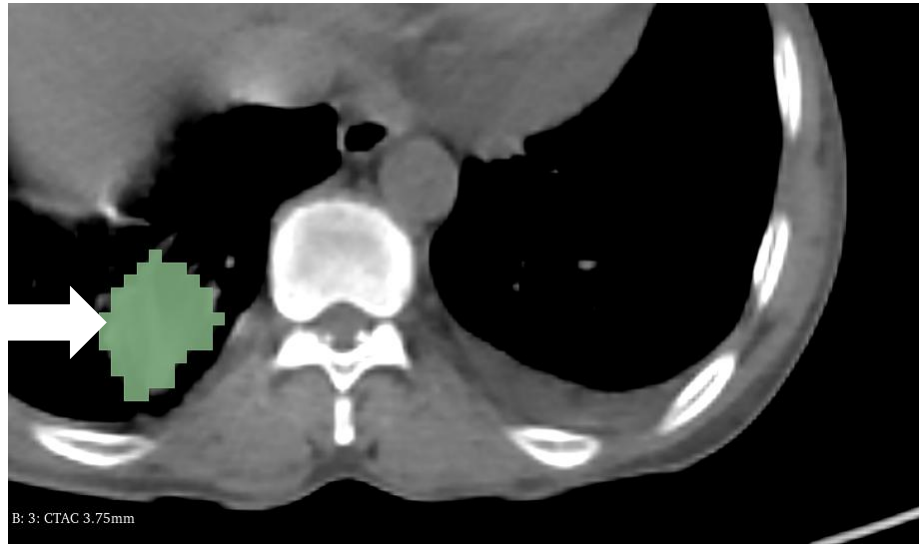
- Mỗi bác sĩ sử dụng 3D Slicer để khoanh vùng tổn thương trên cả PET và CT một cách độc lập.



Hình 2.3. Hình ảnh khối u phổi được đánh dấu trên lát cắt từ máy quét PET

- Trên ảnh PET: Khối u thường xuất hiện dưới dạng vùng có mức hấp thu glucose cao, thể hiện qua các điểm sáng. Áp dụng ngưỡng $SUV \geq 2,5$ hoặc 40% SUV_{max} , theo hướng dẫn EANM [74]. Các tổn thương có SUV thấp hơn nhưng nghi ngờ ác tính trên CT vẫn được phân đoạn. Hình 2.3 chỉ ra một khối u có mức hấp thu glucose cao đã được bác sĩ phân đoạn (mũi tên chỉ vào khối u).

- Trên ảnh CT: Khối u được xác định qua cấu trúc bất thường (kích thước, hình dạng, mật độ). Việc phân đoạn dựa trên ranh giới giải phẫu, bao gồm kích thước, hình dạng (spiculated, lobulated, hoặc smooth), và mật độ HU (thường > 30 cho khối u ác tính) (xem khối u được chỉ bởi mũi tên trong [Hình 2.4](#)).
- Bác sĩ sẽ chồng ghép hai loại ảnh này để khoanh vùng chính xác, tránh nhầm lẫn với các vùng lành tính có hoạt động trao đổi chất cao (ví dụ: mô viêm).



Hình 2.4. Hình ảnh khối u phổi được phân đoạn trên một lát cắt từ phim chụp CT

- Đối với nhóm không phải NSCLC, khoanh vùng các nốt phổi nghi ngờ dựa trên đặc điểm hình ảnh như vôi hóa, ranh giới rõ, hoặc dấu hiệu cây cối (tree-in-bud) [76].
- **Bước 2: Đánh giá sự đồng thuận:**
 - Tính chỉ số tương đồng - Dice Similarity Coefficient (DSC) để đo mức độ trùng lặp giữa các vùng quan tâm mà ba bác sĩ đã đánh dấu. Ngưỡng chấp nhận theo các nghiên cứu gần đây là $DSC \geq 0,7$.
 - Tính chỉ số độ chênh lệch về thể tích - Volume error (VE) với ngưỡng chấp nhận: $VE \leq 0,2$.
 - Sử dụng module Segment Comparison trong 3D Slicer để tự động tính DSC và các chỉ số khác.
- **Bước 3: Giải quyết mâu thuẫn:**
 - Nếu $DSC < 0,7$ hoặc $VE > 0,2$, ba bác sĩ tổ chức họp trực tuyến hoặc trực tiếp, sử dụng 3D Slicer để xem lại hình ảnh và dữ liệu lâm sàng (kết quả sinh thiết, tiền sử bệnh).

- Kết quả cuối cùng là một ROI được đồng thuật, được lưu dưới định dạng NRRD hoặc NIFTI để tích hợp với các công cụ phân tích radiomics.
- **Bước 4: Phân đoạn bán tự động:**
 - Sử dụng module GrowCut hoặc DeepGrow trong 3D Slicer để khởi tạo ROI ban đầu, sau đó bác sĩ hiệu chỉnh thủ công để đảm bảo độ chính xác.
 - Theo nghiên cứu [85], sự kết hợp giữa phân đoạn bán tự động và thủ công giúp giảm thời gian và tăng tính nhất quán.
- **Bước 5: Kiểm tra chất lượng:**
 - Kiểm tra ngẫu nhiên 15-20% trường hợp bằng cách so sánh ROI với kết quả giải phẫu bệnh hoặc hình ảnh theo dõi.
 - Lưu nhật ký phân đoạn (segmentation log) trong 3D Slicer, ghi lại các quyết định, mâu thuẫn, và thông số kỹ thuật (ngưỡng SUV, công cụ sử dụng).

2.1.3. Các thông tin thu thập

Quá trình thu thập thông tin được thực hiện một cách có hệ thống, bao gồm hình ảnh PET/CT và dữ liệu lâm sàng. Hình ảnh PET/CT là thành phần cốt lõi của cơ sở dữ liệu, bao gồm dữ liệu từ hai phương thức hình ảnh bổ sung cho nhau.

- Ảnh PET: Cung cấp thông tin chuyển hóa với giá trị SUV để đánh giá mức độ hấp thu glucose của khối u. Độ phân giải không gian tối thiểu của hình ảnh PET là 128x128 pixel, với kích thước voxel khoảng 4-6 mm, đảm bảo khả năng phát hiện các tổn thương nhỏ.
- Ảnh CT: Cung cấp thông tin giải phẫu với độ phân giải không gian tối thiểu 512x512 pixel và độ dày lát cắt từ 1-3 mm. Giá trị Hounsfield Units (HU) được sử dụng để phân biệt các đặc điểm giải phẫu như mật độ khối u (thường > 30 HU cho khối u ác tính), ranh giới, và sự hiện diện của vôi hóa hoặc dấu hiệu chồi cây trong các tổn thương không phải NSCLC [76].
- Định dạng: DICOM, hỗ trợ tích hợp với 3D Slicer và các công cụ phân tích khác.

Dữ liệu lâm sàng được thu thập để bổ sung cho hình ảnh PET/CT, cung cấp thông tin toàn diện về bệnh nhân và hỗ trợ gắn nhãn chính xác cho cơ sở dữ liệu.

- Thông tin nhân khẩu học bao gồm tuổi, giới tính, tiền sử hút thuốc.. vì những yếu tố này có liên quan đến nguy cơ ung thư phổi và đặc điểm mô học [79].

- Kết quả giải phẫu bệnh là yếu tố quan trọng, xác định loại mô học của NSCLC (ung thư biểu mô tuyến, ung thư biểu mô vảy, hoặc ung thư biểu mô tế bào lớn) và giai đoạn bệnh theo hệ thống AJCC phiên bản thứ 8 [75]. Các thông số giải phẫu bệnh khác như kích thước khối u, mức độ biệt hóa, và sự hiện diện của hoại tử cũng được ghi nhận.
- Các chỉ số lâm sàng khác bổ sung bao gồm: các bệnh lý đồng mắc và tiền sử điều trị trước đó (hóa trị, xạ trị, hoặc phẫu thuật). Những dữ liệu này được tổ chức trong cơ sở dữ liệu có cấu trúc, liên kết với hình ảnh PET/CT thông qua mã định danh bệnh nhân ẩn danh.

Tất cả dữ liệu thu thập được kiểm tra chất lượng để đảm bảo tính đầy đủ và nhất quán. Hình ảnh PET/CT được đánh giá để xác nhận không có nhiều nghiêm trọng hoặc lỗi kỹ thuật, trong khi dữ liệu lâm sàng được kiểm tra chéo với bệnh án và báo cáo giải phẫu bệnh. Quá trình thu thập dữ liệu tuân thủ các quy định đạo đức y khoa và ẩn danh hóa thông tin nhận dạng.

2.1.4. Các chỉ tiêu đánh giá bộ dữ liệu

Để đánh giá chất lượng của bộ dữ liệu PET/CT phục vụ ứng dụng học máy trong chẩn đoán ung thư phổi, các nhà nghiên cứu thường sử dụng một số tiêu chí quan trọng nhằm đảm bảo dữ liệu phù hợp, đáng tin cậy và hỗ trợ hiệu quả cho việc huấn luyện mô hình. Dưới đây là các tiêu chí chính được áp dụng trong nghiên cứu này:

- **Độ chính xác và độ tin cậy của nhãn:** Nhãn (labels) gắn trên dữ liệu (ví dụ: ung thư/lành tính, vị trí khối u) phải phản ánh chính xác chẩn đoán lâm sàng. Cách đánh giá được áp dụng là: So sánh nhãn với "tiêu chuẩn vàng" (gold standard) như kết quả sinh thiết hoặc chẩn đoán cuối cùng của bác sĩ; Tính độ đồng thuận giữa các chuyên gia, sử dụng chỉ số tương đồng và độ sai lệch về thể tích. Đối với độ chính xác nhãn cao lớn hơn 90% và mức độ đồng thuận giữa các bác sĩ cao thường có giá trị lớn hơn 0,7 [86].
- **Chất lượng hình ảnh:** Ảnh PET/CT cần có độ phân giải, độ tương phản tốt và ít nhiễu để đảm bảo thông tin hình ảnh đáng tin cậy. Kiểm tra độ phân giải (resolution) của ảnh CT (thường từ 1-5 mm) và PET (thường từ 4-6 mm). Mức độ nhiễu (noise) thông qua tỷ lệ tín hiệu trên nhiễu cao, ảnh rõ nét, không có hiện tượng mờ hoặc lỗi kỹ thuật. Đảm bảo SUV trên ảnh PET được chuẩn hóa chính xác với thiết bị và quy trình chụp [74] [79].
- **Tính đa dạng và đại diện:** Bộ dữ liệu cần bao quát các trường hợp lâm sàng khác nhau để mô hình học máy có khả năng khái quát hóa tốt. Số lượng bệnh nhân và phân bố theo độ tuổi, giới tính, loại ung thư (tế bào nhỏ, tế bào không nhỏ), giai đoạn bệnh (I-IV). Bao gồm cả các trường hợp

- khó (ví dụ: tổn thương nhỏ, không điển hình) và nhóm đối chứng (lành tính). Bộ dữ liệu có ít nhất vài trăm bệnh nhân, phân bố cân đối giữa các nhóm, tránh thiên lệch [75].
- **Kích thước bộ dữ liệu:** Số lượng mẫu dữ liệu phải đủ lớn để huấn luyện mô hình học máy, đặc biệt là các mô hình học sâu (deep learning). Đếm số lượng ảnh PET/CT (hoặc số bệnh nhân) và số lượng lát cắt (slices) có gắn nhãn. So sánh với yêu cầu tối thiểu của mô hình (thường cần hàng nghìn đến hàng chục nghìn mẫu cho học sâu). Thông thường số lượng tối thiểu được coi là tốt nếu bộ dữ liệu có tối thiểu 100-200 bệnh nhân cho nghiên cứu ban đầu, lý tưởng là >500 bệnh nhân với hàng nghìn lát cắt để đạt hiệu suất cao [87].
 - **Tính nhất quán và chuẩn hóa:** Dữ liệu cần được thu thập và xử lý theo cách đồng nhất để tránh sai lệch do khác biệt kỹ thuật. Thông thường chúng ta sẽ kiểm tra xem ảnh có cùng định dạng (DICOM), độ phân giải, và thang đo SUV không. Quy trình chụp (liều tiêm FDG, thời gian chờ) được chuẩn hóa giữa các mẫu. Tiêu chí tốt ở đây là dữ liệu đồng nhất về định dạng và thông số kỹ thuật, ít biến thiên không cần thiết [85] [88].
 - **Tính đầy đủ của thông tin bổ trợ:** Ngoài ảnh, thông tin lâm sàng bổ trợ (metadata) như tiền sử bệnh, kết quả sinh thiết, hoặc triệu chứng cần được cung cấp đầy đủ, dùng để huấn luyện hoặc cho con người đánh giá, phát triển bộ dữ liệu. Cách đánh giá thông qua việc kiểm tra tỷ lệ mẫu có đầy đủ thông tin (ví dụ: % mẫu có kết quả sinh thiết); đánh giá mức độ chi tiết của metadata (tuổi, giới tính, giai đoạn bệnh, loại ung thư). Thông thường tiêu chí tốt cần đạt được lớn hơn >90% mẫu có metadata đầy đủ và liên quan [86].
 - **Tính pháp lý và đạo đức:** Bộ dữ liệu phải tuân thủ các quy định về bảo mật và đạo đức nghiên cứu. Cách đánh giá: cần có xác nhận dữ liệu đã được ẩn danh (anonymized) để bảo vệ thông tin cá nhân; kiểm tra sự đồng ý của bệnh nhân và phê duyệt từ hội đồng đạo đức. Trong bộ dữ liệu này, do nghiên cứu hồi cứu, nên không thể nhận được ý kiến đồng ý của bệnh nhân. Tuy nhiên, toàn bộ bộ dữ liệu đã được hội đồng đạo đức phê duyệt, thông tin về danh tính của bệnh nhân được ẩn đi, ký hiệu theo mã bệnh nhân. Các bác sĩ tham gia phân đoạn thủ công dữ liệu phải có kinh nghiệm trên 5 năm, phải biết sử dụng thành thạo các công cụ phân đoạn cơ bản [87].
 - **Khả năng sử dụng cho học máy:** Dữ liệu cần được tổ chức và định dạng phù hợp để tích hợp vào pipeline học máy. Cách đánh giá: Kiểm tra xem nhãn đã được lưu dưới dạng mặt nạ (mask) hoặc tọa độ, và dữ liệu có thể chuyển đổi sang định dạng như NIfTI, HDF5 không. Đánh giá khả năng chia tập dữ liệu thành huấn luyện (training), kiểm tra (validation), và thử nghiệm (test) mà không bị chòng lún (data leakage). Tiêu chí tốt là: Dữ liệu

có cấu trúc rõ ràng, dễ tích hợp, và phân chia hợp lý (ví dụ: 70% huấn luyện, 15% kiểm tra, 15% thử nghiệm) [77] [85].

Việc đánh giá bộ dữ liệu PET/CT cần dựa trên sự cân bằng giữa chất lượng, số lượng và tính thực tiễn cho học máy.

2.2. Xây dựng công cụ đọc và tiền xử lý ảnh

2.2.1. Module tiền xử lý ảnh PET/CT từ thiết bị chụp

Tiền xử lý ảnh PET/CT là một bước quan trọng trong quá trình phân tích và đánh giá hình ảnh y tế. Dưới đây là một số lý do quan trọng về tại sao cần tiền xử lý ảnh PET/CT:

- Loại bỏ nhiễu: Hình ảnh PET/CT thường chứa nhiễu do nhiều nguyên nhân như nhiễu điện từ, nhiễu do chuyển động của bệnh nhân hoặc thiết bị chụp. Tiền xử lý giúp loại bỏ hoặc giảm thiểu nhiễu này để tăng độ chính xác của phân tích và đánh giá.
- Cải thiện chất lượng: Tiền xử lý có thể cải thiện chất lượng hình ảnh bằng cách cân bằng độ sáng, làm đồng nhất cường độ, loại bỏ nhiễu ảnh giả và tăng cường độ tương phản.
- Hợp nhất ảnh: Trong hình ảnh PET/CT, việc hợp nhất hình ảnh PET với hình ảnh CT là cần thiết để ghép nối thông tin chức năng và cấu trúc của cơ thể. Tiền xử lý giúp chuẩn bị dữ liệu để thực hiện quá trình hợp nhất ảnh một cách chính xác.
- Phân đoạn và phân loại: Tiền xử lý làm cho việc phân đoạn và phân loại hình ảnh dễ dàng hơn. Phân đoạn giúp phân biệt các cấu trúc cụ thể hoặc mô bệnh trong hình ảnh, trong khi phân loại giúp xác định vị trí, kích thước và tính chất của các khối u, tổn thương hoặc cấu trúc quan trọng khác.
- Chuẩn hóa và đồng nhất: Tiền xử lý giúp chuẩn hóa cường độ của hình ảnh, làm cho các hình ảnh từ các bệnh nhân khác nhau hoặc từ các thiết bị chụp khác nhau có thể so sánh được với nhau.
- Giảm kích thước và tăng tốc độ xử lý: Trong một số trường hợp, tiền xử lý có thể giúp giảm kích thước của hình ảnh và tăng tốc độ xử lý, đặc biệt quan trọng khi làm việc với các bộ dữ liệu lớn.

Tóm lại, tiền xử lý ảnh PET/CT là cần thiết để chuẩn bị dữ liệu hình ảnh cho quá trình phân tích và đánh giá trong lĩnh vực y tế, đảm bảo tính chính xác và đáng tin cậy của các kết quả. Có một số phương pháp tiền xử lý ảnh PET/CT từ thiết bị chụp để chuẩn bị dữ liệu hình ảnh cho việc phân tích và đánh giá. Dưới

đây là một số kỹ thuật đã được thực hiện để nâng cao chất lượng của hình ảnh PET/CT trong luận án.

Cân bằng độ sáng: Điều chỉnh độ sáng và độ tương phản của hình ảnh để loại bỏ nhiễu và làm nổi bật các chi tiết cần thiết. Cân bằng độ sáng giúp làm đồng nhất độ sáng của hình ảnh trên toàn bộ vùng quan tâm. Xét một ảnh xám rời rạc X có n_i điểm ảnh có giá trị mức xám i . Mật độ xác suất giá trị điểm ảnh được chọn được tính bằng công thức:

$$p_X(i) = \frac{n_i}{n}; 0 \leq i \leq L \quad (2.1)$$

Với L là giá trị cực đại của thang xám, n là tổng số điểm ảnh có trong hình ảnh đang quan tâm. Hàm số phân bố tích lũy của điểm ảnh có giá trị i trong ảnh X được xác định như sau:

$$cdf_X(i) = \sum_{j=0}^i p_X(j) \quad (2.2)$$

Ta muốn xây dựng một phép biến đổi để khiến cho histogram trở nên phẳng hơn trong khoảng thang xám từ 0 đến $L-1$. Như vậy hàm phân bố tích lũy của ảnh mới sẽ có dạng:

$$cdf_X(i) = (i + 1)K \quad (2.3)$$

Với K là hằng số. Tính chất của hàm phân bố tích lũy cho phép hàm biến đổi T được định nghĩa như sau:

$$T(i) = cdf_X(i) \quad (2.4)$$

Hàm T ánh xạ mức xám thành giá trị số thực nằm trong khoảng từ 0 đến 1.

Lọc: Sử dụng các phương pháp lọc như lọc Gaussian hoặc lọc median để loại bỏ nhiễu khỏi hình ảnh PET/CT. Việc này giúp cải thiện chất lượng của hình ảnh và làm tăng độ chính xác của các phép đo và phân tích.

Hàm lọc Gaussian hai chiều được xây dựng như sau:

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \quad (2.5)$$

Trong đó: x là khoảng cách từ gốc tọa độ đến trục hoành, y là khoảng cách từ gốc tọa độ đến trục tung và σ là độ lệch chuẩn của phân bố Gauss.

Hàm lọc Median được thực thi bằng cách cộng trung bình các điểm ảnh trong cửa sổ lọc với kích thước nhất định, giá trị cộng trung bình sẽ được điền vào tâm của vị trí của cửa sổ lọc. Cửa sổ lọc sau đó sẽ trượt trên toàn bộ ảnh để thực hiện học. Mã giả của bộ lọc Median được thể hiện như sau:

1. allocate outputPixelValue[image width][image height]
2. allocate window[window width × window height]

```

3. edgex := (window width / 2) rounded down
4. edgey := (window height / 2) rounded down
for x from edgex to image width- edgex do
  for y from edgey to image height- edgey do
    i = 0
    for fx from 0 to window width do
      for fy from 0 to window height do
        window[i] := inputPixelValue[x + fx- edgex][y + fy- edgey]
        i := i + 1
      sort entries in window[]
      outputPixelValue[x][y] := window[window width * window height / 2]

```

Đăng ký ảnh: Đăng ký (registration) là quá trình đưa các hình ảnh từ các dòng thời gian hoặc từ các góc chụp khác nhau về cùng một không gian tọa độ. Trong hình ảnh PET/CT, việc đăng ký giữa ảnh PET và CT là quan trọng để ghép nối thông tin chức năng và cấu trúc của cơ thể. Hình ảnh PET và CT thu được từ thiết bị PET CT là hai hình ảnh độc lập, mặc dù hình ảnh này có cùng trục tọa độ bệnh nhân, tuy nhiên, vẫn tồn tại một số khác biệt nhỏ. Độ phân giải hình ảnh khác nhau: thông thường hình ảnh CT có độ phân giải cỡ 1 mm, trong khi ảnh PET có độ phân giải cỡ 3.5 mm. Sự khác biệt đôi chút về kích thước và vị trí lát cắt: cả ảnh PET và CT đều sử dụng topogram để xác định vị trí (theo trục Z) và kích thước của lát cắt. Về mặt lý thuyết vị trí và kích thước lát cắt của hai loại ảnh sẽ bằng nhau, tuy nhiên, do phương pháp tái tạo, sai lệch cơ khí dẫn tới có sự sai khác về vị trí và kích thước lát cắt giữa ảnh PET và ảnh CT. Quá trình chồng chập FOV cũng gây ra sai số của việc hợp nhất.

Sau khi đã hiệu chỉnh được hình học của PET và CT vẫn có một vấn đề ta cần phải quan tâm đó là sự khác biệt giữa giá trị điểm ảnh của ảnh PET và ảnh CT. Hình ảnh CT thường là hình ảnh 12 bit, các điểm ảnh thể hiện hệ số HU có giá trị từ -1024 đến 3071. Trong khi hình ảnh PET thường là số 16 bit không âm (từ 0 đến $2^{16}-1$). Khi hợp nhất hình ảnh PET/CT cần thiết phải điều chỉnh (lấy tỷ lệ) lại giá trị điểm ảnh của hai loại ảnh để ảnh hợp nhất được thể hiện trên cùng giải cường độ (cùng dải màu). Thông thường cả ảnh PET và CT đều được lưu trữ dưới dạng DICOM, đây là một siêu tệp có chứa dữ liệu hình ảnh và các thông tin để hiệu chỉnh hình ảnh.

Sau khi đã hiệu chỉnh, xử lý, hình ảnh PET và CT sẽ được hợp nhất bằng công thức như sau:

$$R_{\alpha} = \alpha R_{PET} + (1 - \alpha) R_{CT} \quad (2.6)$$

Với α là hệ số điều chỉnh hiển thị nằm trong khoảng từ 0 đến 1; R_α , R_{PET} và R_{CT} lần lượt là các ảnh hợp nhất, ảnh PET và ảnh CT tương ứng.

Chuẩn hóa cường độ: Đôi khi, việc chuẩn hóa cường độ của hình ảnh là cần thiết để làm cho các hình ảnh từ các bệnh nhân khác nhau hoặc từ các thiết bị chụp khác nhau có thể so sánh được với nhau. Áp dụng các phương pháp chuẩn hóa cường độ để đảm bảo tính đồng nhất của hình ảnh, cho toàn bộ dữ liệu. Quá trình chuẩn hóa cường độ sẽ được thực hiện bằng cách lấy tỷ lệ giá trị của mỗi điểm ảnh với giá trị cực đại của điểm ảnh trong bộ dữ liệu hoặc nhân với hệ số tỷ lệ của hình ảnh có chứa trong tiêu đề của tập tin DICOM.

$$I_N(i, j) = I(i, j)k \quad (2.7)$$

Với k là hệ số tỷ lệ, $I_N(i, j)$ và $I(i, j)$ lần lượt là giá trị điểm ảnh sau và trước khi thực hiện lấy chuẩn hóa.

Xử lý chuyển đổi: Áp dụng các phương pháp xử lý hình ảnh như xoay, cắt, thu phóng để điều chỉnh hình dạng và vị trí của ảnh PET/CT. Phép biến đổi xoay ảnh được thực hiện thông qua phép biến đổi Euler. Giả sử mỗi hình ảnh lát cắt được xếp theo trục Z, như vậy hai trục tọa độ còn lại của hình ảnh bao gồm trục X và trục Y. Ta có thể xây dựng được ma trận quay Euler quanh trục Z cho hình ảnh như sau:

$$R_z(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \quad (2.8)$$

Với ϕ là góc cần xoay ảnh, như vậy mỗi điểm ảnh mới sẽ thu được khi thực hiện phép quay Euler với các điểm ảnh cũ thu được như sau:

$$I_N(i, j) = I(i, j)R_z \quad (2.9)$$

Quá trình cắt hình ảnh được thực hiện thông qua kỹ thuật trượt chỉ số ma trận, thay vì truy xuất toàn bộ hình ảnh thì ta có thể truy xuất từ các vị trí bất kỳ của chỉ số điểm ảnh, từ đó tạo ra các hình ảnh cắt theo khung mong muốn. Quá trình thu nhỏ ảnh được thực hiện bằng cách cộng trung bình nhiều điểm ảnh hoặc loại bỏ các điểm ảnh theo lân cận. Quá trình phóng to điểm ảnh được thực hiện bằng cách lặp lại các điểm ảnh nhiều lần. Đây là các phương pháp đơn giản, đều được hỗ trợ bởi các chương trình xử lý hình ảnh cơ bản.

Loại bỏ ảnh giả: Các ảnh giả của PET/CT chủ yếu là do cấy ghép kim loại, chuyển động hô hấp, sử dụng chất cản quang và cắt bớt hình ảnh. Điều này có thể ảnh hưởng đến chất lượng hình ảnh PET/CT. Việc loại bỏ các ảnh giả này giúp cải thiện chất lượng của hình ảnh và đảm bảo tính chính xác của các phân tích. Loại

bỏ các ảnh giả khỏi hình ảnh PET/CT bằng cách sử dụng các kỹ thuật như interpolation hoặc thresholding.

- ✓ Interpolation (Nội suy): Trong trường hợp này, chúng ta sẽ sử dụng các phương pháp nội suy để tạo ra các giá trị mới dựa trên các giá trị lân cận trong hình ảnh. Ví dụ, chúng ta có thể sử dụng phương pháp nội suy tuyến tính hoặc spline để tạo ra các giá trị mới cho các pixel bị thiếu hoặc bị loại bỏ.

$$I_N(i, k) = \frac{I(i, j) + I(i, j+1)}{2} \quad (2.10)$$

Nếu nội suy để tạo ra hình ảnh mới, giá trị cực đại của k phải gấp 2 lần giá trị cực đại của j. Trong trường hợp thực hiện nội suy để loại bỏ giá trị bất thường của một điểm ảnh, ta sẽ sử dụng phương pháp nội suy lân cận 8 như sau:

$$I(i, j) = \frac{I(i-1, j-1) + I(i-1, j) + I(i+1, j-1) + I(i, j-1) + I(i, j+1) + I(i+1, j-1) + I(i+1, j) + I(i+1, j+1)}{8} \quad (2.11)$$

- ✓ Thresholding (Ngưỡng): Trong kỹ thuật này, chúng ta sẽ thiết lập một ngưỡng và loại bỏ các pixel có giá trị thấp hơn hoặc cao hơn ngưỡng đã chọn. Điều này có thể giúp loại bỏ nhiễu hoặc các cấu trúc không mong muốn khỏi hình ảnh. Kết quả của quá trình định ngưỡng sẽ tạo ra hình ảnh nhị phân với các cấu trúc cần quan tâm đã được làm nổi bật. Mã giả của kỹ thuật lấy ngưỡng được thể hiện như sau:

```

Thresh_val = val
for x from edgex to image width- edgex do
    for y from edgey to image height- edgey do
        if I(x, y) ≤ Thresh_val do
            I(x, y) = 0
        else do
            I(x, y) = 1
        
```

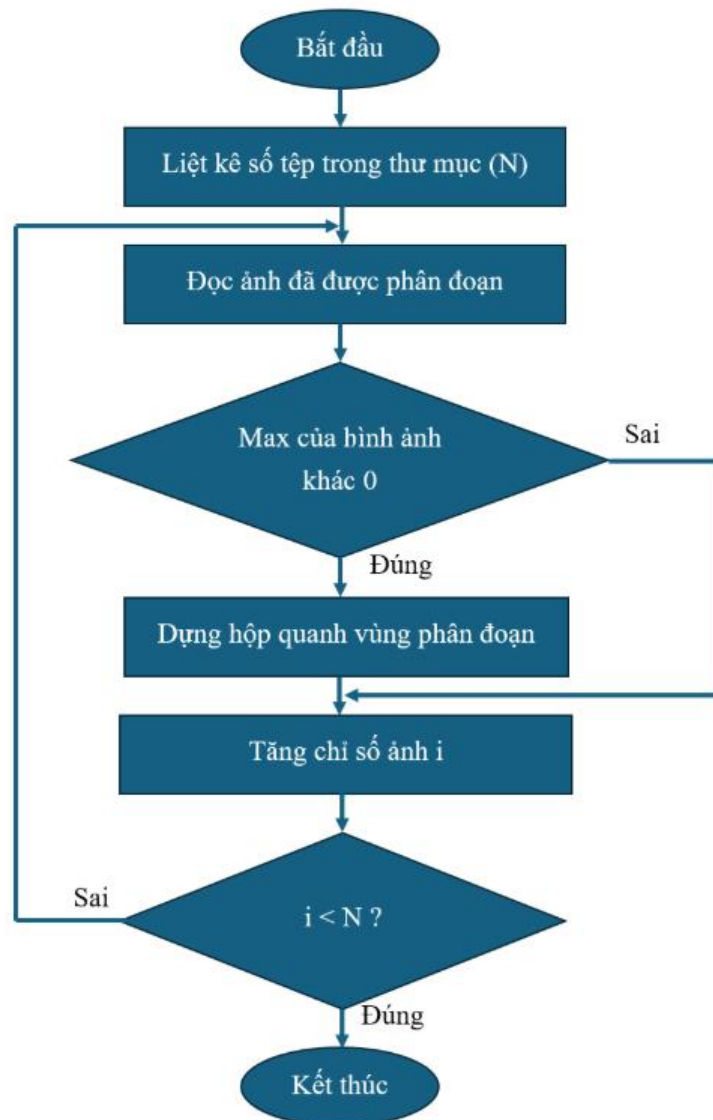
Đối với phương pháp xác định ngưỡng, ta chỉ cần xác định một giá trị ngưỡng (trong trường hợp này chọn ngưỡng theo giá trị SUV).

Người dùng có thể áp dụng một cách độc lập hoặc kết hợp với nhau các phương pháp tiền xử lý này trong chương trình để tạo ra một hình ảnh PET/CT tốt nhất cho mục đích phân tích và đánh giá trong y tế.

2.2.2. Mô đun hiển thị thông tin khối u từ ảnh phân đoạn thủ công

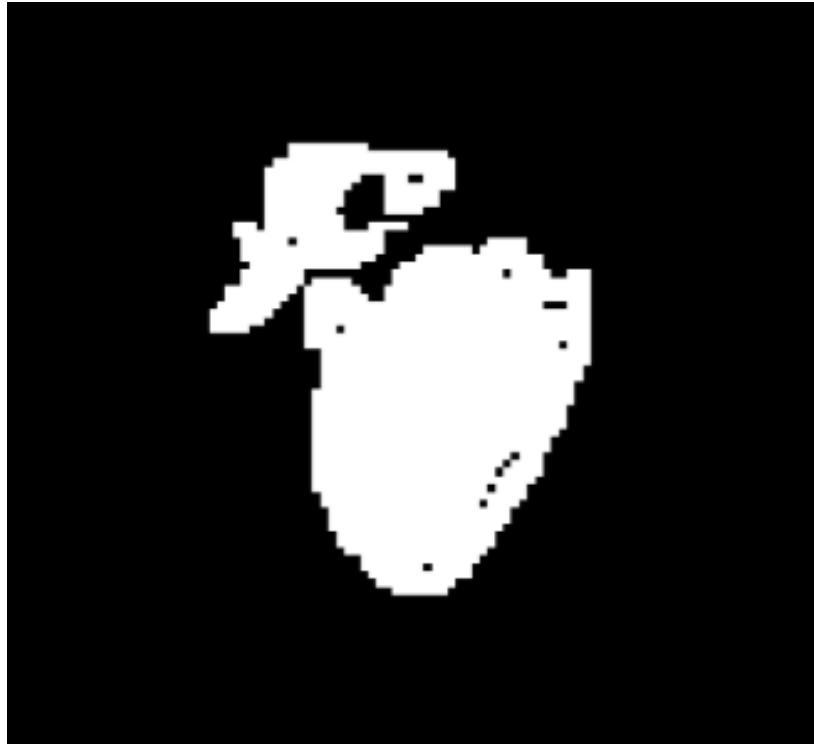
Các khối u có trong dữ liệu hình ảnh PET/CT của các bệnh nhân được bác sĩ đánh dấu thủ công bằng cách khoanh và tô kín diện tích khối u. Trong quá trình tô, các bác sĩ có thể lựa chọn gán thẻ cho từng vùng khoanh, phần mềm phân đoạn thủ công sẽ gán giá trị tương ứng cho các điểm ảnh mà bác sĩ đã gán thẻ. Trong luận án, các điểm ảnh của khối u có giá trị bằng 1, phần mềm đọc và hiển thị khối u sẽ đọc và xác định khu vực có chứa các điểm ảnh có kích thước bằng 1 và bao bọc quanh khối u bằng một hình chữ nhật. Việc xác định đường biên dạng chính xác của khối u là không cần thiết vì máy tính chỉ có thể lấy hình ảnh là hình vuông hoặc chữ nhật để đưa vào mô hình huấn luyện.

Để xác định biên dạng của khối u trong từng lát cắt, tác giả đã xây dựng hàm *search_boundingbox* có chức năng xác định các khối u có trong lát cắt và bao khối u bằng một hình chữ nhật tương ứng, đồng thời xác định tâm của khối u chính là tâm của hình chữ nhật vừa tìm được. Lưu đồ thuật toán của chương trình được thể hiện trên [Hình 2.5](#).



Hình 2.5: Lưu đồ thuật toán chương trình xác định đường bao

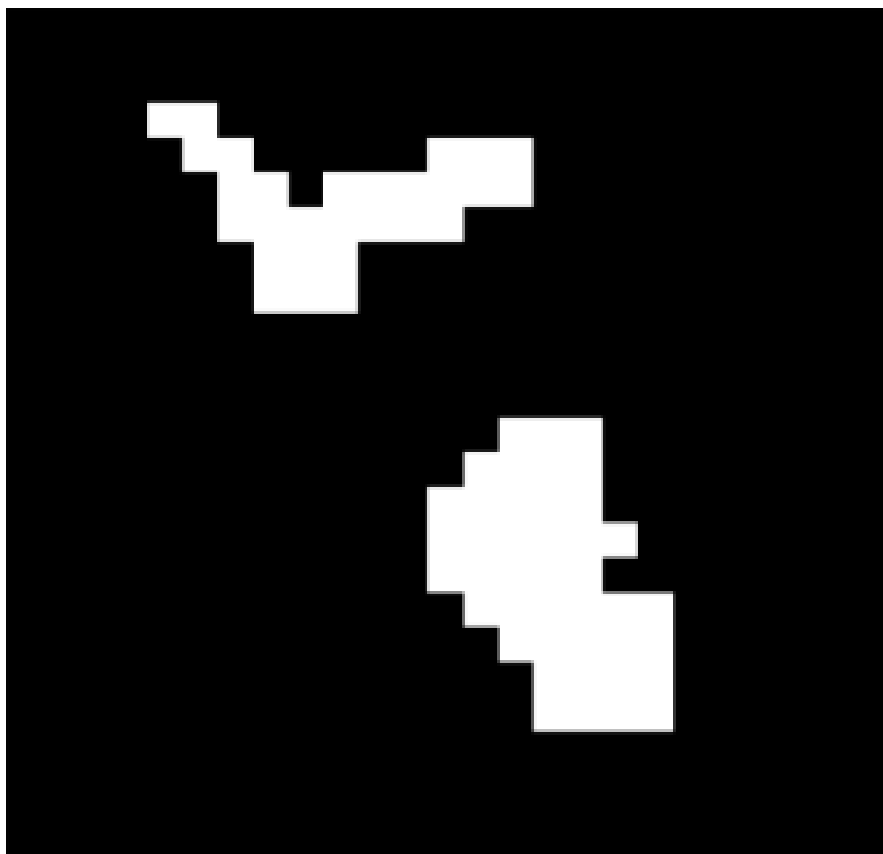
Để xác định đường bao của khối u, chương trình sẽ thực hiện tính vi phân theo từng hàng, cột của hình ảnh phân đoạn. Kết quả, ta sẽ thu được các điểm mà tại đó cường độ thay đổi một cách đột ngột. Biên dạng của khối u có thể được xác định bằng cách lấy các điểm vi phân có vị trí liên kề. Với thuật toán này, ta sẽ xác định được rất nhiều đường biên dạng của khối u vì các khối u có thể được tô không kín hoặc khối u có hình dạng phân nhánh (kết quả trên một lát cắt ta sẽ nhìn thấy hai vùng được tô kín, [Hình 2.6](#))



Hình 2.6. Hình ảnh khối u không được tô kín

Đối với các khối u không được tô kín, chương trình sẽ phát hiện được rất nhiều biên dạng nằm bên trong một biên dạng lớn. Để xử lý được trường hợp này, nghiên cứu sinh tiến hành loại bỏ các đường bao mà có thể tích nhỏ hơn 5 mm, do mục tiêu của luận án chỉ phát hiện các nốt có kích thước lớn cỡ centi mét. Các đường bao quanh vùng diện tích lớn hơn 5 mm sẽ được tiếp tục xử lý dựa trên tọa độ của tâm đường bao, nếu tọa độ tâm của đường bao nằm bên trong một đường bao lớn hoặc không cách đường bao lớn quá 5 mm thì sẽ được coi là các điểm khoanh lỗi hoặc thuộc về đường bao lớn. Các đường bao sau khi được xử lý sẽ hợp nhất lại thành các đường bao thỏa mãn điều kiện của luận án.

Đối với các khối u được khoanh thành nhiều phần kín, liên tục, chương trình sẽ tiến hành xác định diện tích và tâm của khối u, nếu khoảng cách tâm của các đường bao khối u không lớn hơn kích thước của các đường bao đó, thì các đường bao sẽ được hợp nhất và coi là một khối u duy nhất ([Hình 2.7](#)).

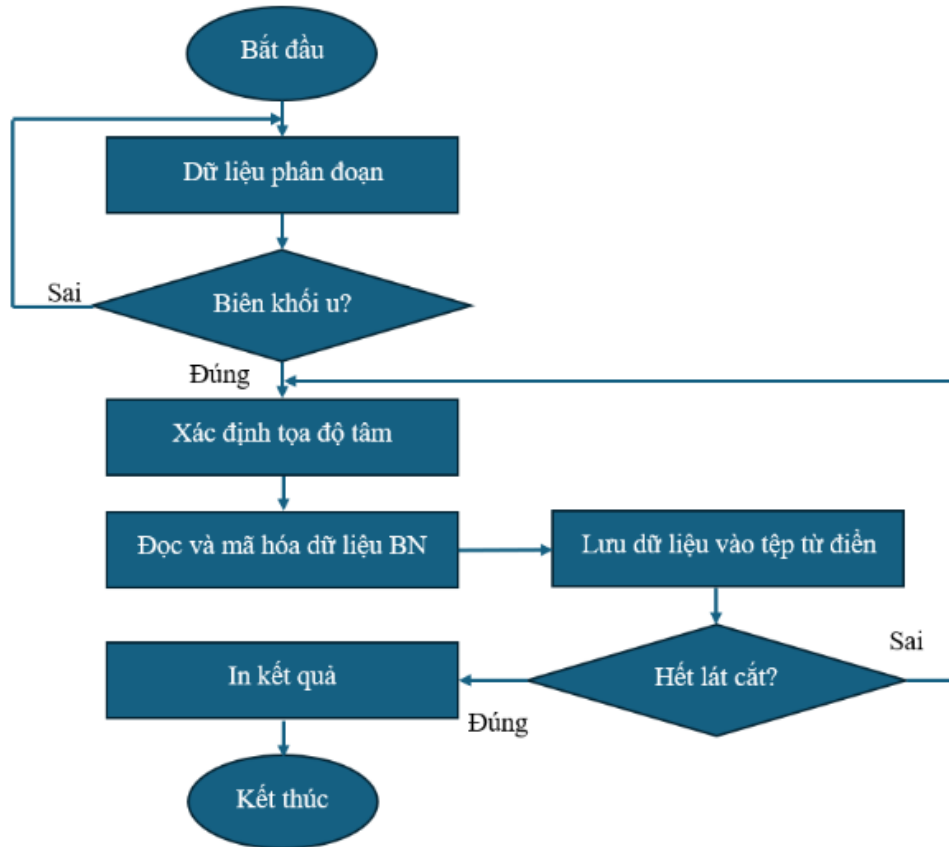


Hình 2.7. Hình ảnh các khối u phân nhánh

2.2.3. Mô đun trích xuất dữ liệu của các bệnh nhân

Các đường bao sau khi đã được tìm kiếm và xử lý sẽ được xếp vào một biến từ điển có tên là **NodCandidate**. Biến từ điển này chứa thông tin của bệnh nhân đã được phân đoạn và các thông tin của khối u. Trường thông tin của bệnh nhân được đọc trực tiếp từ lớp tiêu đề của tệp tin DICOM. Trường thông tin khối u sẽ bao gồm diện tích của khối u được tính bằng diện tích bên trong đường bao quanh khối u đã được xác định. Tâm của khối u có tọa độ (x, y, z) được xác định là tâm của phần diện tích trong đường bao khối u. Mỗi lát cắt có thể có một hoặc nhiều khối u được đánh dấu, chương trình sẽ ghi hết thông tin của các khối u có trong mỗi lát cắt ra file kết quả.

Việc lưu giữ thông tin theo từng lát cắt giúp cho người dùng có thể nhận diện chính xác từ đó lấy dữ liệu và huấn luyện cho mô hình. Tuy nhiên, dữ liệu ghi theo từng lát cắt sẽ rất dài, không có tính tổng quát, vì vậy, nghiên cứu sinh tiến hành ghi thêm một tệp thông tin nữa. Tệp thông tin này có tác dụng tóm tắt thông tin có trong tệp thông tin theo lát cắt để người đọc có thể quan sát một cách tổng quát về thông tin khối u của bệnh nhân, từ đó dễ dàng kiểm tra lại chất lượng phân đoạn thủ công của bộ dữ liệu. Lưu đồ chương trình ghi thông tin kết quả phân đoạn thủ công được thể hiện như trong [Hình 2.8](#).



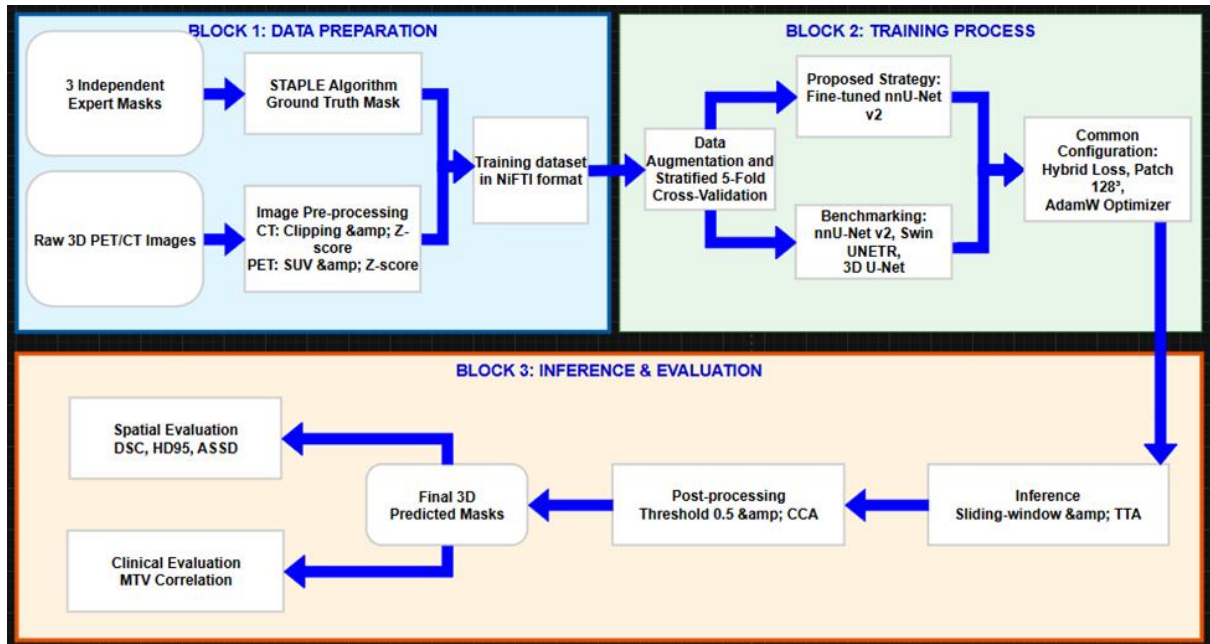
Hình 2.8. Lưu đồ chương trình ghi thông tin kết quả phân đoạn thủ công

2.3. Phát triển nền tảng tự động phân đoạn ảnh PET/CT ứng dụng mô hình học sâu

2.3.1. Cấu trúc chung

Quy trình phân đoạn ung thư phổi không tế bào nhỏ (NSCLC) tự động được mô tả trong Hình 2.9. Quy trình đề xuất vận hành theo ba bước liên tiếp. Đầu tiên, ở giai đoạn chuẩn bị dữ liệu, các ảnh 3D PET/CT thô được tiền xử lý, đồng thời thuật toán STAPLE được áp dụng để tổng hợp kết quả gán nhãn từ ba chuyên gia thành một ảnh chuẩn đồng thuận (consensus ground truth), sau đó dữ liệu được chuyển sang định dạng NiFTI. Tiếp theo, trong giai đoạn huấn luyện, tập dữ liệu được tăng cường (augmented) và đánh giá bằng phương pháp kiểm định chéo K-fold phân tầng (stratified 5-fold cross-validation). Mô hình nnU-Net v2 cải tiến được huấn luyện và so sánh với các mô hình nền tảng (nnU-Net v2, Swin UNETR và 3D U-Net) trong cùng một cấu hình đồng nhất (hàm mất mát hỗn hợp, kích thước vùng cắt patch size $128 \times 128 \times 128$, bộ tối ưu hóa AdamW). Cuối cùng, ở giai đoạn suy luận và đánh giá, các kết quả dự đoán được tạo ra thông qua chiến lược cửa sổ trượt (sliding-window) kết hợp tăng cường dữ liệu tại thời điểm kiểm thử (test-time augmentation), kể đến là bước hậu xử lý để thu được mặt nạ dự

đoán 3D (3D mask) cuối cùng. Hiệu năng của mô hình sau đó được đánh giá dựa trên các chỉ số không gian (DSC, HD95, ASSD) và độ tương quan lâm sàng thông qua thể tích khối u chuyển hóa (MTV).



Hình 2.9. Nền tảng phân đoạn tự động

2.3.2. Chuẩn bị dữ liệu cho mạng học sâu

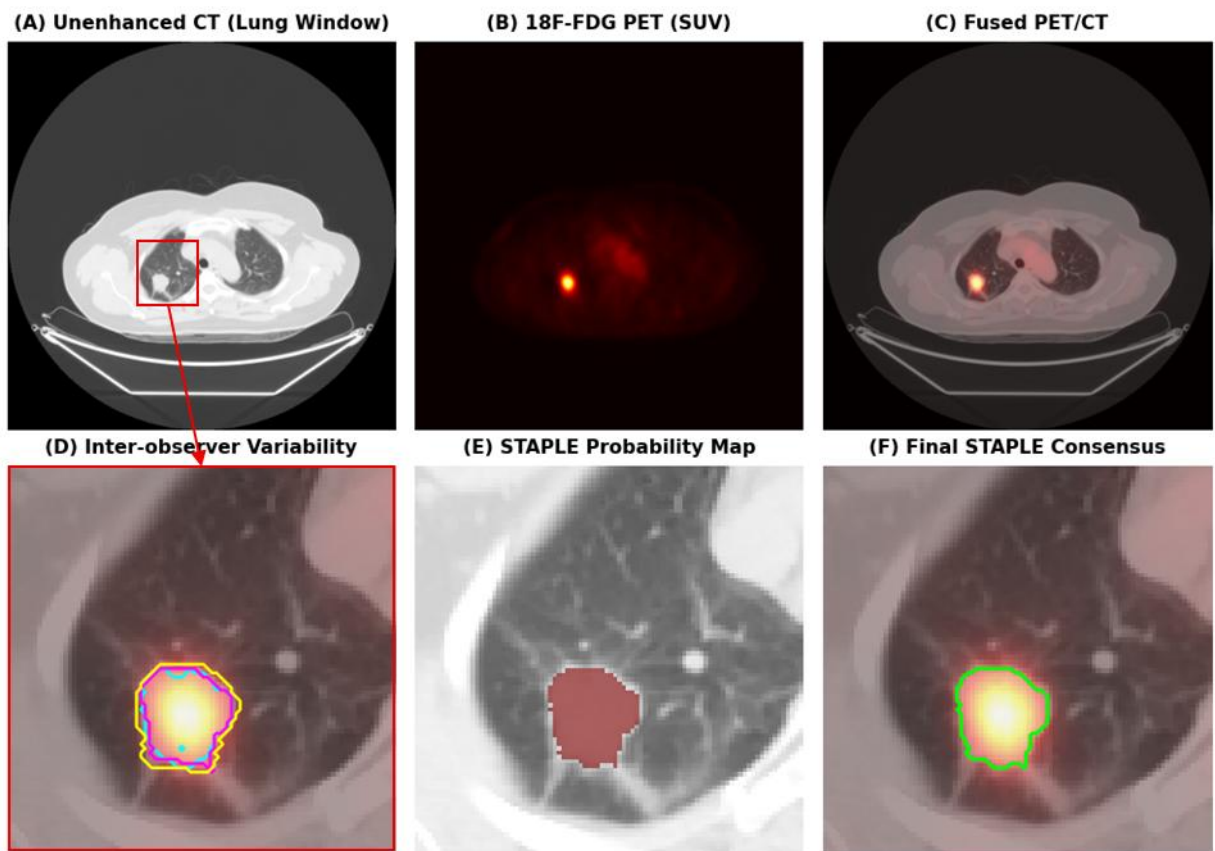
Hình ảnh PET và CT được tái lấy mẫu (resampling) về kích thước voxel đẳng hướng $2.0 \times 2.0 \times 2.0 \text{ mm}^3$; trong đó, phép nội suy spline bậc ba (cubic spline interpolation) được áp dụng cho dữ liệu cường độ pixel, còn phép nội suy láng giềng gần nhất (nearest-neighbor interpolation) được dùng cho các nhãn mặt nạ (mask labels) nhằm bảo toàn tính nguyên vẹn của ranh giới khối u.

- **Kênh CT:** Kỹ thuật cắt ngưỡng cửa sổ (window clipping) được áp dụng trong khoảng $[-1000, 400]$ HU để tập trung vào vùng nhu mô phổi, đồng thời loại bỏ nhiễu từ xương hoặc các vật thể bên ngoài. Sau đó, dữ liệu được chuẩn hóa Z-score dựa trên giá trị trung bình và độ lệch chuẩn của toàn bộ tập dữ liệu.
- **Kênh PET:** Cường độ phóng xạ được chuyển đổi sang giá trị hấp thụ chuẩn hóa (SUV) dựa trên trọng lượng cơ thể và được cắt ngưỡng trong khoảng $[0, 20]$ nhằm giảm thiểu tác động từ các giá trị ngoại lai cực đoan (outliers). Tiếp theo, dữ liệu được chuẩn hóa Z-score để mạng thần kinh nhân tạo có thể tập trung vào độ tương phản chuyển hóa, thay vì các biến thiên về cường độ tuyệt đối.

Các kỹ thuật tăng cường dữ liệu (data augmentation) được áp dụng bao gồm xoay ngẫu nhiên, thay đổi tỷ lệ (scaling), và đặc biệt là bổ sung nhiễu Gaussian

vào dữ liệu PET để mô phỏng đặc tính nhiễu Poisson vốn có trong ảnh y học hạt nhân.

Hình 2.10 minh họa dữ liệu hình ảnh đa phương thức đầu vào cùng quy trình khởi tạo nhãn chuẩn (ground-truth) nghiêm ngặt được áp dụng trong nghiên cứu này. Bằng cách dung hợp các chi tiết giải phẫu từ ảnh chụp CT không cản quang với hoạt tính chuyển hóa từ ảnh PET ^{18}F -FDG, các tổn thương NSCLC đã được định vị một cách chính xác, sau đó được các chuyên gia y học hạt nhân khoanh vùng thủ công bằng phần mềm 3D Slicer. Để giải quyết sự khác biệt vốn có giữa các người đánh giá từ những phần chú giải độc lập, thuật toán STAPLE [66] đã được áp dụng để tổng hợp chúng thành một bản đồ xác suất thống nhất. Quy trình này tạo ra một kết quả đồng thuận cuối cùng mang tính xác suất vững chắc, giúp giảm thiểu một cách hiệu quả các định kiến chủ quan của con người [67], từ đó cung cấp một tiêu chuẩn tham chiếu khách quan và có độ tin cậy cao để huấn luyện các kiến trúc học sâu của nhóm tác giả. Cuối cùng, tập dữ liệu được phân chia theo chiến lược kiểm định chéo 5 phần (5-fold cross-validation), phân tầng theo thể tích khối u cùng chỉ số SUV_{max} , rồi chuẩn hóa thành các ma trận nhị phân 3D, trong đó giá trị 1 đại diện cho vùng khối u và 0 biểu thị cho các mô lành.



Hình 2.10: *Trực quan hóa dữ liệu PET/CT đa phương thức và quy trình khởi tạo nhãn chuẩn. (A) Ảnh CT không cản quang; (B) Ảnh PET ^{18}F -FDG (bản đồ*

SUV); (C) Ảnh dung hợp PET/CT; (D) Sự khác biệt giữa ba chuyên gia khoan vùng; (E) Bản đồ xác suất từ thuật toán STAPLE; và (F) Kết quả đồng thuận xác suất cuối cùng được sử dụng làm tiêu chuẩn tham chiếu.

2.3.3. Lựa chọn và huấn luyện mô hình học sâu

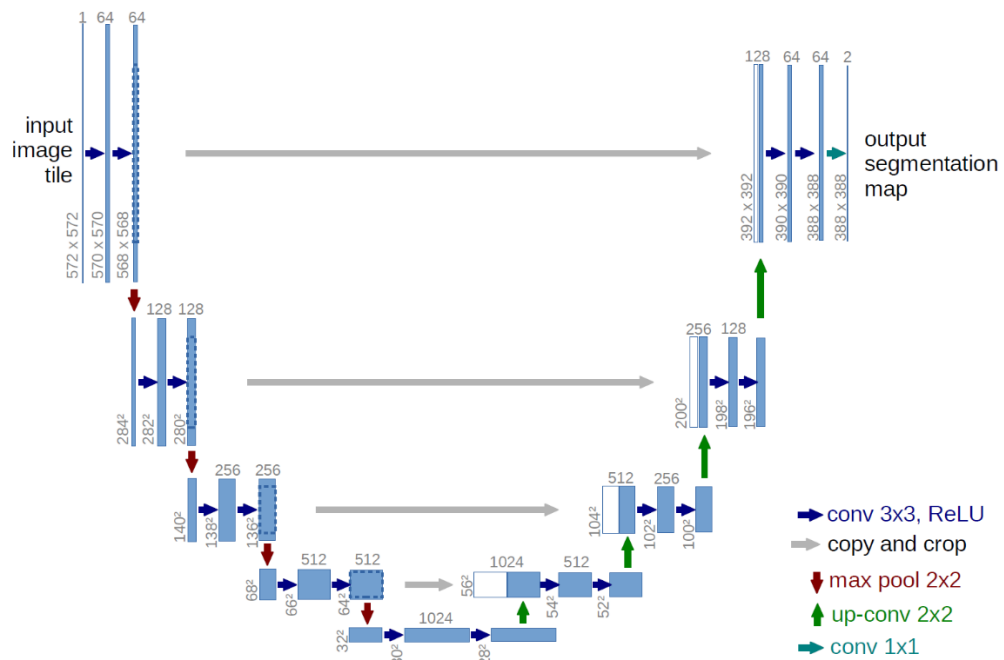
Lựa chọn mô hình học sâu

Nghiên cứu sinh đã thử nghiệm thành công mô hình phân tách u não từ ảnh cộng hưởng từ T1 MRI với cấu trúc mạng U-net có khả năng phân tách u não một cách chính xác [89]. Nghiên cứu tiếp theo này không chỉ hướng tới tối ưu hóa hiệu năng phân đoạn mà còn xây dựng một hệ thống đánh giá chuẩn đối sánh (benchmarking) toàn diện, nhằm kiểm thử khả năng thích ứng của các thuật toán tiên tiến trước đặc điểm nhân trắc học và phân phối hình ảnh đặc thù của bệnh nhân Việt Nam. Ba kiến trúc học sâu chiến lược đã được lựa chọn để hiện thực hóa mục tiêu này:

- **nnU-Net v2 (Mô hình cơ sở):** Nhờ thế mạnh tự cấu hình thông minh theo đặc trưng dữ liệu mà không cần can thiệp thủ công, mô hình tiếp nhận đồng thời dữ liệu đa kênh PET và CT (kích thước khối đầu vào $128 \times 128 \times 128$ voxel), mang lại độ ổn định vượt trội trong việc xác định ranh giới giải phẫu.
- **Swin UNETR (Kiến trúc đột phá ngữ cảnh):** Khắc phục triệt để giới hạn về trường thụ cảm cục bộ của CNN truyền thống bằng cách tích hợp cơ chế tự chú ý (self-attention) của Transformer. Qua việc phân rã dữ liệu 3D thành các token khối nhỏ và xử lý bằng các cửa sổ $7 \times 7 \times 7$, mô hình thiết lập thành công các mối phụ thuộc tầm xa, đặc biệt hiệu quả trong việc nhận diện các khối u NSCLC kích thước lớn hoặc có hình thái xâm lấn phức tạp, từ đó tối ưu hóa độ nhạy tổng thể.
- **3D U-Net với ResNet backbone (Mô hình kiểm chứng):** Được triển khai trên nền tảng MONAI nhằm đánh giá hiệu quả dung hợp đặc trưng ở cấp độ bộ mã hóa. Sự chênh lệch hiệu năng giữa kiến trúc tiêu chuẩn này và phiên bản nnU-Net tinh chỉnh (đạt chỉ số Dice bứt phá 83,44%) là minh chứng rõ nét cho thấy: nếu không qua học chuyển giao và tinh chỉnh sâu trên miền dữ liệu mục tiêu, các mô hình phổ thông rất khó vượt qua được rào cản dịch chuyển miền dữ liệu (domain shift).

a) nnU-Net

Mạng U-Net là một kiến trúc mạng nơ-ron tích chập (CNN) đặc biệt, được phát triển bởi Olaf Ronneberger và các cộng sự vào năm 2015 [90], ban đầu phục vụ cho mục đích phân đoạn hình ảnh y tế (Medical Image Segmentation). Cái tên "U-Net" xuất phát từ hình dáng đặc trưng của cấu trúc mạng, trông giống như chữ U (Hình 2.11).



Hình 2.11: Cấu trúc mạng U-net

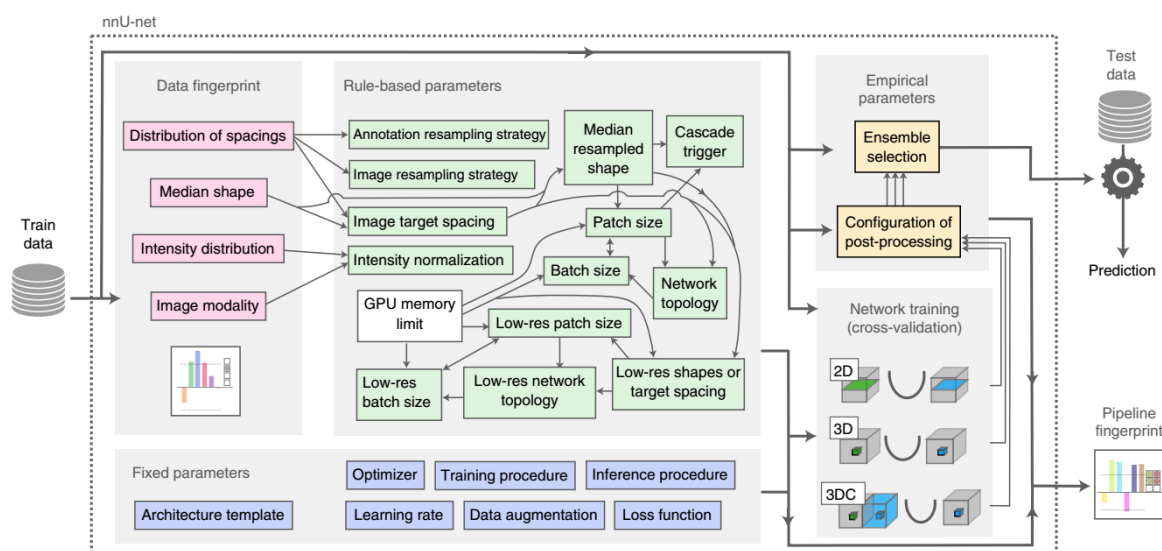
Kiến trúc này được chia làm hai phần chính đối xứng nhau qua một "nút thắt" (bottleneck):

- Nhánh co bóp (Encoding path): Nằm ở bên trái, thực hiện nhiệm vụ trích xuất các đặc trưng ngữ cảnh từ hình ảnh thông qua các lớp tích chập và giảm mẫu (downsampling). Điều này giúp mạng hiểu được "cái gì" đang có trong ảnh. Cấu trúc chi tiết như sau: Lặp lại các khối gồm hai lớp tích chập (Convolution) 3×3, mỗi lớp đi kèm một đơn vị chỉnh lưu tuyến tính (ReLU). Sau đó là lớp Max Pooling 2×2 với bước nhảy (stride) bằng 2 để giảm độ phân giải. Mục tiêu của nhánh này là tại mỗi bước giảm mẫu, số lượng kênh đặc trưng (feature channels) tăng gấp đôi để học các đặc trưng phức tạp hơn, trong khi kích thước không gian giảm xuống.
- Nhánh mở rộng (Decoding path): Nằm ở bên phải, thực hiện nhiệm vụ khôi phục kích thước ảnh ban đầu thông qua các lớp chuyển vị tích chập (up-convolution). Điều này giúp mạng xác định chính xác "vị trí" của đối tượng. Cấu trúc chi tiết như sau: Lặp lại các bước tăng mẫu (upsampling) bằng cách sử dụng tích chập chuyển vị (up-convolution/transposed convolution) 2×2. Mỗi bước này làm giảm một nửa số kênh đặc trưng nhưng tăng gấp đôi kích thước ảnh. Lớp cuối cùng: Sử dụng tích chập 1×1 để ánh xạ các vector đặc trưng vào số lượng lớp phân loại mong muốn.

Điểm đột phá nhất của U-Net chính là các kết nối tắt (skip connections). Các kết nối này truyền thông tin chi tiết từ các tầng ở nhánh co bóp sang thẳng các tầng tương ứng ở nhánh mở rộng. Nhờ đó, mạng có thể kết hợp được cả thông tin ngữ cảnh sâu (tầng thấp) và thông tin không gian chi tiết (tầng cao), giúp việc

phân đoạn ranh giới các đối tượng trở nên cực kỳ chính xác ngay cả khi tập dữ liệu huấn luyện nhỏ. Zhou và cộng sự [91] đã phân tích thực trạng cũng như những thách thức trong việc phân đoạn ung thư phổi, đồng thời nhấn mạnh vai trò ngày càng quan trọng của trí tuệ nhân tạo (AI), đặc biệt là trong việc tích hợp hai phương thức PET và CT để phân đoạn khối u một cách chính xác và toàn diện hơn.

Nhiều biến thể và cải tiến cho cấu trúc U-Net đã được đề xuất trong những năm gần đây. Chẳng hạn, nnU-Net—một phương pháp phân đoạn dựa trên học sâu—có khả năng tự động cấu hình cho bất kỳ nhiệm vụ mới nào, giúp giảm thiểu việc tinh chỉnh các siêu tham số (hyperparameter) thủ công (Isensee và cộng sự, 2021 [92]; Ferrante và cộng sự, 2022 [93]). Hình 2.12 minh họa cách nnU-Net giải quyết một cách có hệ thống bài toán cấu hình toàn bộ quy trình phân đoạn (pipeline), đồng thời trực quan hóa và mô tả các lựa chọn thiết kế cốt lõi nhất. Từ một nhiệm vụ phân đoạn mới, các đặc tính của tập dữ liệu sẽ được trích xuất dưới dạng "dấu bản sắc dữ liệu" (màu hồng). Một tập hợp các quy tắc trực nghiệm sẽ mô hình hóa các mối ràng buộc lẫn nhau giữa các tham số (biểu diễn bằng các mũi tên mảnh) và tác động lên dấu bản sắc này nhằm suy luận ra các "tham số theo quy tắc" phụ thuộc vào dữ liệu (màu xanh lá) của quy trình. Các tham số này được hỗ trợ bởi nhóm "tham số cố định" (màu xanh dương) — vốn đã được định nghĩa trước và không cần điều chỉnh. Có tối đa ba cấu hình mạng được huấn luyện thông qua quá trình kiểm chéo 5 phần (five-fold cross-validation). Cuối cùng, nnU-Net sẽ tự động thực hiện lựa chọn thực nghiệm để tìm ra tổ hợp mô hình (ensemble) tối ưu nhất, đồng thời xác định xem có cần thực hiện hậu xử lý hay không ("tham số thực nghiệm", màu vàng).



Hình 2.12: Quy trình tự động cấu hình phương pháp phân đoạn ảnh sinh y học dựa trên nnU-Net.

Ứng dụng của nnU-Net. Khi áp dụng nnU-Net vào một tập dữ liệu mới, quy trình tự động cấu hình sẽ tự khởi chạy mà không cần đến bất kỳ sự can thiệp thủ công nào. Do đó, ngoại trừ một vài lựa chọn thực nghiệm còn lại cần quyết định, quy trình này không làm phát sinh thêm bất kỳ chi phí tính toán nào ngoài quy trình huấn luyện mạng tiêu chuẩn. Cơ chế tự động cấu hình phương pháp của nnU-Net bắt đầu bằng việc trích xuất "dấu bản sắc dữ liệu", sau đó thực thi các quy tắc thực nghiệm. Theo mặc định, nnU-Net sẽ khởi tạo ba cấu hình U-Net khác nhau: một mạng U-Net hai chiều (2D), một mạng U-Net 3D hoạt động ở độ phân giải ảnh gốc (full resolution), và một chuỗi U-Net 3D nối tầng (cascade) — trong đó mạng U-Net đầu tiên sẽ xử lý các hình ảnh đã được hạ lấy mẫu (downsampled), còn mạng thứ hai được huấn luyện để tinh chỉnh các bản đồ phân đoạn (segmentation maps) do mạng trước đó tạo ra ở độ phân giải gốc.

Sau quá trình kiểm chéo (cross-validation), nnU-Net sẽ dựa trên kết quả thực nghiệm để chọn ra cấu hình đơn lẻ hoặc tổ hợp mô hình (ensemble) cho hiệu năng tốt nhất. Cuối cùng, nnU-Net cũng dùng thực nghiệm để quyết định xem có áp dụng bước hậu xử lý "loại bỏ các thành phần không phải lớn nhất" (non-largest component suppression) hay không, nếu bước này giúp cải thiện hiệu năng đo lường được. Kết quả đầu ra của quá trình cấu hình và huấn luyện tự động này là các mô hình đã được tối ưu hóa hoàn toàn, sẵn sàng để triển khai nhằm dự đoán trên những hình ảnh mới chưa từng thấy. [Isensee và cộng sự \[93\]](#) đã chứng minh khả năng tổng quát hóa của các lựa chọn thiết kế được mã hóa trong các tham số cố định, tham số theo quy tắc và tham số thực nghiệm của nnU-Net bằng cách áp dụng nó vào 13 tập dữ liệu bổ sung.

nnU-Net có khả năng xử lý đa dạng các cấu trúc mục tiêu và đặc tính hình ảnh. [Isensee và cộng sự \[92\]](#) chứng minh giá trị của nnU-Net như một công cụ phân đoạn "ăn liền" (out-of-the-box) hiệu quả bằng cách áp dụng hệ thống này vào 11 cuộc thi phân đoạn ảnh sinh y học quốc tế, bao gồm 23 tập dữ liệu khác nhau và 53 nhiệm vụ phân đoạn {6,14,16–24} (<https://cremi.org/>). Tập hợp thử nghiệm này bao quát rất nhiều đối tượng từ các cơ quan, cấu trúc vi mô của cơ quan, cho đến các khối u, tổn thương và cấu trúc tế bào trên cả ảnh 2D lẫn 3D; được thu nhận từ các kỹ thuật chụp cộng hưởng từ (MRI), cắt lớp vi tính (CT), kính hiển vi điện tử (EM) và kính hiển vi huỳnh quang (FM). Các "Cuộc thi" (Challenges) ở đây là những đấu trường quốc tế nhằm đánh giá hiệu năng của nhiều thuật toán khác nhau trong một môi trường được chuẩn hóa. Trong tất cả các nhiệm vụ phân đoạn, nnU-Net đều được huấn luyện từ đầu (from scratch) và chỉ sử dụng duy nhất dữ liệu do ban tổ chức cuộc thi cung cấp. Về mặt định tính, nghiên cứu sinh nhận thấy nnU-Net có thể thích ứng tốt với sự phân hóa lớn về đặc tính dữ liệu cũng như sự đa dạng của các cấu trúc mục tiêu; nói cách khác, các cấu hình quy trình (pipeline) được hệ thống tự động khởi tạo hoàn toàn tương

đồng với những thiết lập mà các chuyên gia con người đánh giá là hợp lý và đúng đắn.

nnU-Net vượt trội hơn các quy trình chuyên biệt trên hàng loạt nhiệm vụ đa dạng. Mặc dù mang tính tổng quát (generic), nnU-Net vẫn vượt mặt hầu hết các giải pháp phân đoạn hiện có, ngay cả khi các giải pháp này đã được tối ưu hóa một cách chuyên biệt cho từng nhiệm vụ tương ứng. Nhìn chung, nnU-Net đã thiết lập một cấu trúc mạng tốt nhất (state of the art) trên 33/53 cấu trúc mục tiêu, và ở các cấu trúc còn lại, hệ thống cũng thể hiện hiệu năng ngang ngửa hoặc tiệm cận với những vị trí dẫn đầu trên bảng xếp hạng (leaderboard).

Tóm lại, nnU-Net đã thiết lập một cấu trúc mạng tốt nhất (state of the art) trên nhiều cuộc thi phân đoạn ngữ nghĩa khác nhau, đồng thời thể hiện khả năng tổng quát hóa mạnh mẽ mà không đòi hỏi tri thức chuyên gia hay tài nguyên tính toán vượt quá một quy trình huấn luyện mạng tiêu chuẩn. Như nhận định của Litjens và các cộng sự, và nay đã được chứng thực bằng định lượng trong nghiên cứu này, việc cấu hình phương pháp trong xử lý ảnh sinh y học từng được coi là một "bài thực hành mang đậm tính thực nghiệm" mà "không thể có một công thức chung nào". Dựa trên công thức được đề xuất trong nghiên cứu này, nnU-Net có khả năng tự động hóa quy trình vốn thường thiếu tính hệ thống và rườm rà này, từ đó giúp giảm bớt gánh nặng cho các nhà nghiên cứu. Nghiên cứu sinh đề xuất khai thác nnU-Net như một công cụ "ăn liền" để đạt được hiệu năng phân đoạn tiên tiến nhất, như một mô hình cơ sở (baseline) chuẩn hóa và độc lập với tập dữ liệu (dataset-agnostic) để làm mốc so sánh, và như một khung làm việc (framework) nhằm đánh giá trên quy mô lớn các ý tưởng mới mà không tốn công sức thiết kế thủ công.

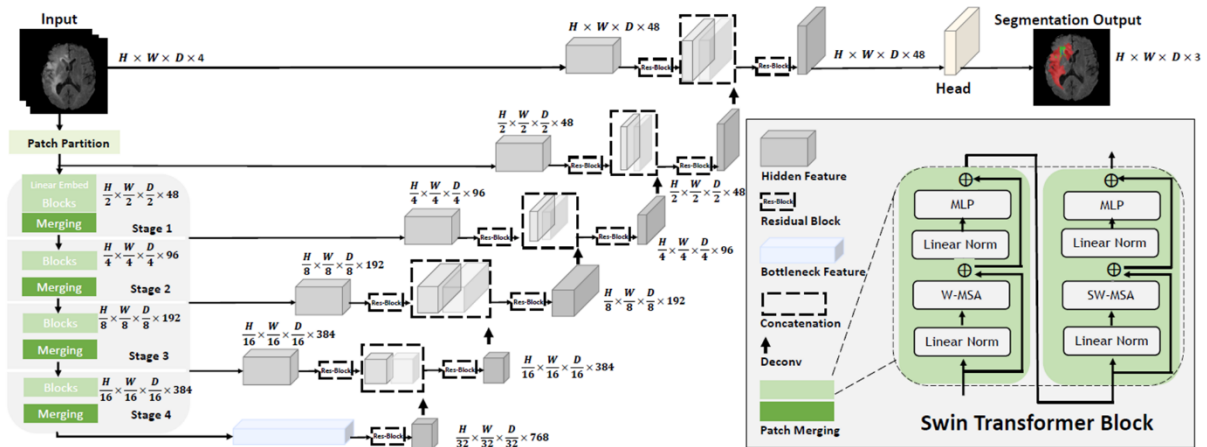
Phiên bản tiền nhiệm (nnU-Net v1) vốn được phát triển thần tốc trong một khoảng thời gian ngắn khi tác giả [Isensee và cộng sự \(2020\) \[92\]](#) tham gia cuộc thi *Medical Segmentation Decathlon* vào năm 2018. Chính vì vậy, cấu trúc và chất lượng mã nguồn lúc đó chưa phải là tối ưu nhất. Nhiều tính năng được bổ sung sau này cũng không còn hoàn toàn tương thích với các nguyên lý thiết kế ban đầu của nnU-Net. nnU-Net V2 là phiên bản được đại tu toàn diện — theo đúng nghĩa "xóa sạch đi và làm lại từ đầu". Nhờ vậy, mọi thứ đã trở nên tốt hơn rất nhiều (ít nhất là theo quan điểm của tác giả). Dù hiệu năng phân đoạn vẫn được giữ vững, phiên bản này đã được tích hợp thêm hàng loạt cải tiến cực kỳ thú vị. Giờ đây, việc sử dụng nnU-Net như một khung phát triển (development framework) hay việc tinh chỉnh thủ công (fine-tune) cấu hình của nó cho các tập dữ liệu mới đã trở nên dễ dàng hơn đáng kể. Một động lực lớn thúc đẩy việc tái cấu trúc này chính là sự ra đời của dự án *Helmholtz Imaging*, thôi thúc các nhà phát triển mở rộng khả năng của nnU-Net sang nhiều định dạng hình ảnh và lĩnh vực mới hơn.

b) Swin UNETR

Trong các cuộc thi BraTS trước đây, các tổ hợp (ensembles) kiến trúc có dạng U-Net đã đạt được nhiều kết quả đầy hứa hẹn trong bài toán phân đoạn khối u não đa phương thức. Kamnitsas và các cộng sự [94] đã đề xuất một mô hình phân đoạn bền vững bằng cách tổng hợp kết quả đầu ra của nhiều mô hình dựa trên mạng CNN khác nhau như 3D U-Net [72], 3D FCN [95] và DeepMedic [94]. Tiếp sau đó, Myronenko và các cộng sự [96] đã giới thiệu SegResNet, một kiến trúc sử dụng bộ mã hóa - giải mã phần dư (residual encoder-decoder), trong đó một nhánh phụ trợ được sử dụng để tái cấu trúc dữ liệu đầu vào thông qua một bộ tự mã hóa biến phân (variational auto-encoder) đóng vai trò như một tác vụ thay thế (surrogate task). Zhou và các cộng sự [97] đã đề xuất sử dụng một tổ hợp gồm các mạng dựa trên CNN khác nhau bằng cách tích hợp thông tin bối cảnh đa quy mô thông qua một khối chú ý (attention block). Zhou và các cộng sự [97] lại sử dụng phương pháp xếp chồng hai giai đoạn (two-stage cascaded approach) cấu thành từ các mô hình U-Net; trong đó, giai đoạn đầu tiên tính toán một dự báo phân đoạn thô, sau đó kết quả này sẽ được tinh chỉnh lại ở giai đoạn thứ hai. Hơn thế nữa, Isensee và các cộng sự [92] đã đề xuất mô hình nnU-Net và chứng minh rằng một kiến trúc U-Net phổ quát với những sửa đổi nhỏ là đủ để đạt được hiệu năng đầy cạnh tranh trong nhiều cuộc thi BraTS liên tiếp.

Gần đây, các mô hình dựa trên nền tảng Transformer đã thu hút được rất nhiều sự chú ý trong cả lĩnh vực thị giác máy tính [98] [99] [100] lẫn phân tích ảnh y tế [101] [102]. Chen và các cộng sự [101] đã giới thiệu một kiến trúc 2D U-Net được hưởng lợi từ một mạng ViT (Vision Transformer) đặt tại vùng thắt cổ chai (bottleneck) của mạng thần kinh. Wang và các cộng sự [103] đã mở rộng phương pháp này cho bài toán phân đoạn khối u não 3D. Ngoài ra, Xie và các cộng sự [104] đã đề xuất sử dụng một mô hình dựa trên ViT với các lớp Transformer có thể biến dạng (deformable transformer layers) nằm giữa bộ mã hóa và giải mã dựa trên CNN của nó, bằng cách xử lý các đặc trưng được trích xuất ở các độ phân giải khác nhau. Khác với các phương pháp này, Hatamizadeh và các cộng sự [102] đã đề xuất kiến trúc UNETR; trong đó, một bộ mã hóa dựa trên ViT sử dụng trực tiếp các vùng ảnh đầu vào 3D (3D input patches) được kết nối với một bộ giải mã dựa trên CNN. UNETR đã cho thấy những kết quả đầy hứa hẹn đối với bài toán phân đoạn khối u não sử dụng bộ dữ liệu MSD [105]. Khác với mô hình UNETR, kiến trúc Swin UNETR do nhóm tác giả đề xuất sử dụng một bộ mã hóa Swin Transformer để trích xuất các biểu diễn đặc trưng ở nhiều độ phân giải khác nhau, kết hợp với cơ chế cửa sổ dịch chuyển (shifted windowing mechanism) để tính toán tính năng tự chú ý (self-attention) (Hình 2.13). Kết quả chứng minh rằng các mạng Swin Transformer [100] sở hữu năng lực tuyệt vời trong việc học các biểu

diễn bối cảnh đa quy mô và mô hình hóa các mối phụ thuộc tầm xa (long-range dependencies) khi so sánh với các phương pháp dựa trên ViT có độ phân giải cố định.



Hình 2.13: Tổng quan về kiến trúc Swin UNETR. Đầu vào của mô hình là khối ảnh MRI đa phương thức 3D gồm 4 kênh. Swin UNETR tạo ra các vùng ảnh (patches) không chồng chập từ dữ liệu đầu vào, rồi sử dụng một lớp phân chia vùng ảnh (patch partition layer) để thiết lập các cửa sổ có kích thước mong muốn phục vụ cho việc tính toán tính năng tự chú ý (self-attention). Các biểu diễn đặc trưng đã được mã hóa trong mạng Swin Transformer sẽ được chuyển đến một bộ giải mã CNN (CNN-decoder) thông qua các kết nối nhảy bậc (skip connections) ở nhiều độ phân giải khác nhau. Kết quả đầu ra phân đoạn cuối cùng.

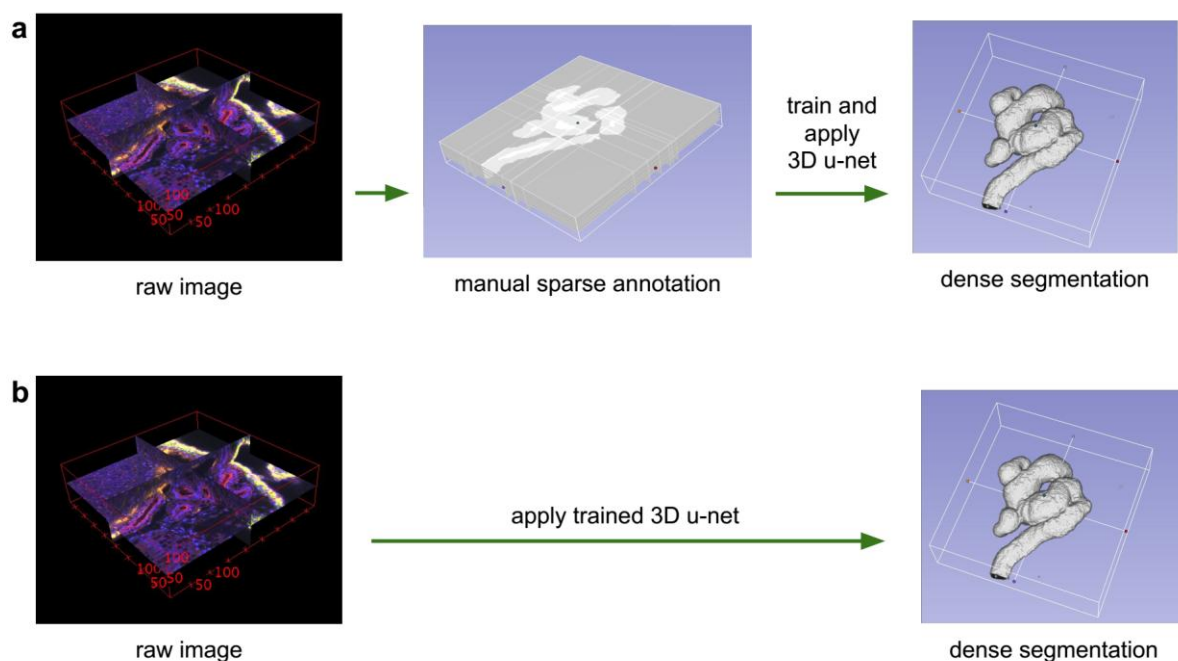
i) 3D U-Net

Dữ liệu khối (volumetric data) vốn rất phổ biến trong lĩnh vực phân tích dữ liệu y sinh. Tuy nhiên, việc gán nhãn phân đoạn cho loại dữ liệu này lại gặp nhiều khó khăn, do màn hình máy tính chỉ có thể hiển thị các lát cắt 2D. Chính vì vậy, việc gán nhãn từng lát cắt cho các khối dữ liệu lớn trở nên vô cùng tẻ nhạt và tốn công sức. Cách làm này cũng rất kém hiệu quả, bởi các lát cắt liền kề nhau hầu như đều chứa thông tin tương tự. Đặc biệt là đối với các phương pháp tiếp cận dựa trên học máy — vốn đòi hỏi một lượng lớn dữ liệu gán nhãn — thì việc dán nhãn toàn bộ các khối 3D không phải là giải pháp hữu hiệu để tạo ra các tập dữ liệu huấn luyện lớn và phong phú nhằm đạt khả năng tổng quát hóa tốt.

Trong nghiên cứu này, mạng học sâu được đề xuất có khả năng học cách tạo ra các phân đoạn khối dày đặc (dense volumetric segmentations), nhưng chỉ yêu cầu một vài lát cắt 2D được gán nhãn để huấn luyện. Mạng này có thể được ứng dụng theo hai cách khác nhau như mô tả trong Hình 1: hướng ứng dụng thứ nhất tập trung vào việc làm dày đặc một tập dữ liệu vốn được gán nhãn thưa thớt (sparsely annotated); hướng thứ hai học từ nhiều tập dữ liệu gán nhãn thưa thớt

để tổng quát hóa (generalize) cho dữ liệu mới. Cả hai trường hợp này đều có tính ứng dụng thực tiễn rất cao. Kiến trúc của mạng được phát triển dựa trên mô hình U-Net trước đây, bao gồm một nhánh mã hóa (encoder) co cụm để phân tích toàn bộ hình ảnh và một nhánh giải mã (decoder) mở rộng kế tiếp nhằm tạo ra kết quả phân đoạn ở độ phân giải đầy đủ [90]. Trong khi U-Net thuần túy là một kiến trúc 2D, mạng được đề xuất trong bài báo này tiếp nhận đầu vào là các khối 3D và xử lý chúng bằng các phép toán 3D tương ứng, cụ thể là tích chập 3D (3D convolutions), max pooling 3D và các tầng tích chập chuyển vị 3D (3D up-convolutional layers). Bên cạnh đó, nhóm nghiên cứu cũng tránh các điểm nghẽn (bottlenecks) trong kiến trúc mạng [106] và áp dụng chuẩn hóa theo lô (batch normalization) [107] để mạng hội tụ nhanh hơn.

Trong nhiều ứng dụng y sinh, chỉ cần một số lượng rất ít hình ảnh là đủ để huấn luyện một mạng có khả năng tổng quát hóa tương đối tốt. Điều này là do bản thân mỗi hình ảnh vốn đã chứa đựng các cấu trúc lặp đi lặp lại với những biến thể tương ứng. Trong ảnh khối (volumetric images), hiệu ứng này thậm chí còn rõ rệt hơn, đến mức có thể huấn luyện mạng chỉ với hai ảnh khối để tổng quát hóa cho ảnh khối thứ ba. Một hàm mất mát có trọng số (weighted loss function) cùng kỹ thuật tăng cường dữ liệu đặc biệt cho phép huấn luyện mạng chỉ với một vài lát cắt được gán nhãn thủ công, tức là từ dữ liệu huấn luyện được gán nhãn thưa thớt.



Hình 2.14: Các kịch bản ứng dụng phân đoạn khối với mạng 3D U-Net. (a) Phân đoạn bán tự động: Người dùng gán nhãn một vài lát cắt trên mỗi khối dữ liệu cần phân đoạn, sau đó mạng sẽ dự đoán để đưa ra kết quả phân đoạn dày

đặc (dense segmentation). **(b) Phân đoạn hoàn toàn tự động:** Mạng được huấn luyện dựa trên các lát cắt đã gán nhãn từ một tập huấn luyện mang tính đại diện, sau đó có thể vận hành trực tiếp trên các khối dữ liệu chưa được gán nhãn.

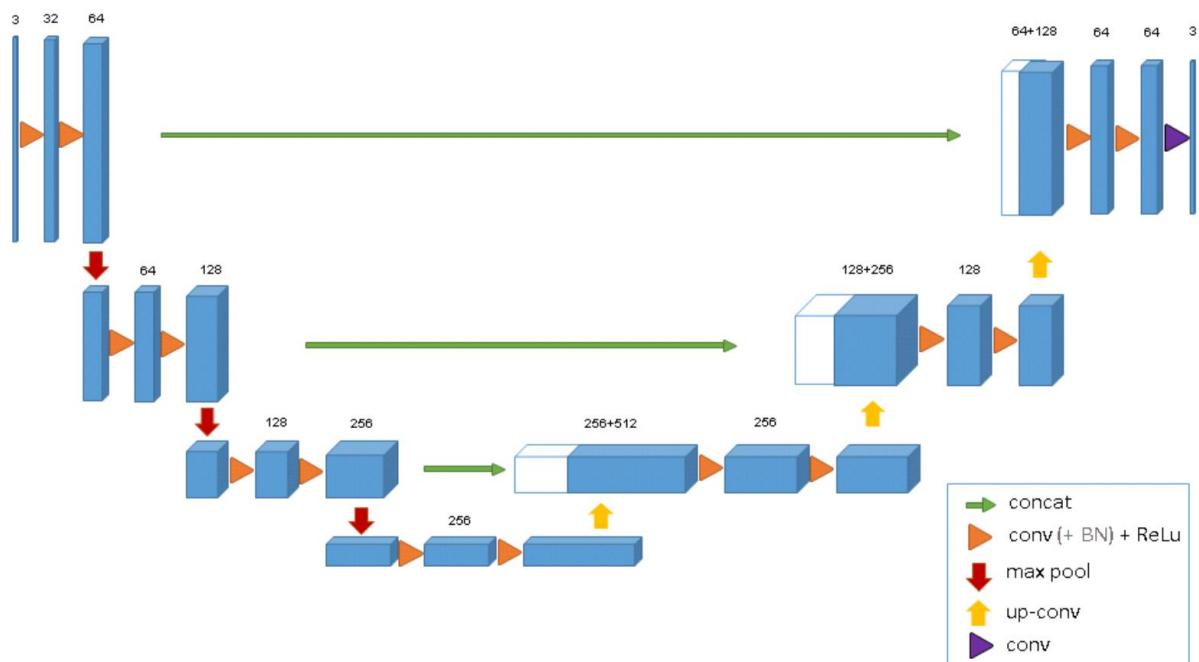
Hiện nay, các mạng tích chập (CNN) đã có thể phân đoạn các ảnh y sinh 2D phức tạp với độ chính xác tiệm cận hiệu năng của con người [90] [108] [109]. Nhờ thành công này, nhiều nỗ lực đã được thực hiện nhằm áp dụng CNN 3D vào dữ liệu khối y sinh. Milletari và các cộng sự [110] đã giới thiệu một mạng CNN kết hợp với phương pháp bỏ phiếu Hough (Hough voting) để phân đoạn 3D. Tuy nhiên, phương pháp của họ không phải là một quy trình liền mạch từ đầu đến cuối (end-to-end) và chỉ hoạt động hiệu quả đối với các cấu trúc dạng khối (blob-like) đặc khít. Tiếp cận của Kleesiek và các cộng sự [111] là một trong số ít các phương pháp CNN 3D end-to-end dành cho phân đoạn 3D. Mặc dù vậy, mạng của họ khá nông và chỉ có duy nhất một tầng max-pooling sau các lớp tích chập đầu tiên; do đó, mô hình không có khả năng phân tích các cấu trúc ở nhiều thang đo tỷ lệ khác nhau (multiple scales).

Nghiên cứu được phát triển dựa trên mô hình U-Net 2D [90] — kiến trúc từng giành chiến thắng tại nhiều cuộc thi phân đoạn và theo vết quốc tế vào năm 2015. Kiến trúc cùng kỹ thuật tăng cường dữ liệu của U-Net cho phép huấn luyện các mô hình có khả năng tổng quát hóa rất tốt chỉ từ một vài mẫu gán nhãn. Mô hình đã khai thác một thực tế rằng: các phép biến đổi cứng (rigid transformations) và các biến dạng đàn hồi nhẹ (elastic deformations) nếu được áp dụng một cách hợp lý vẫn tạo ra các hình ảnh có tính hợp lý về mặt sinh học.

Các kiến trúc tích chập chuyển vị (up-convolutional) như mạng tích chập hoàn toàn (FCN) dùng cho phân đoạn ngữ nghĩa [95] hay U-Net cho đến nay vẫn chưa được phổ biến rộng rãi, và vì thế cần mở rộng kiến trúc này lên không gian 3D [112]. Trong nghiên cứu đó của Tran và các cộng sự, kiến trúc này được áp dụng cho dữ liệu video và có sẵn tập nhãn đầy đủ cho quá trình huấn luyện. Điểm nhấn nổi bật trong bài báo là mô hình có thể được huấn luyện từ con số không (from scratch) trên các khối dữ liệu gán nhãn thưa thớt, đồng thời có thể xử lý các khối dữ liệu có kích thước lớn tùy ý nhờ vào chiến lược ghép mảnh liền mạch (seamless tiling strategy).

Hình 2.15 minh họa kiến trúc của mạng. Tương tự như mô hình U-Net tiêu chuẩn, mạng bao gồm một nhánh phân tích (analysis path) và một nhánh tổng hợp (synthesis path), mỗi nhánh có bốn bước phân giải (resolution steps). Trong nhánh phân tích, mỗi tầng chứa hai lớp tích chập $3 \times 3 \times 3$, mỗi lớp được theo sau bởi một đơn vị tuyến tính chỉnh lưu (ReLU), và kế đến là một tầng max pooling $2 \times 2 \times 2$ với bước nhảy (stride) bằng 2 trên mỗi chiều không gian. Trong nhánh tổng hợp, mỗi tầng gồm một phép tích chập chuyển vị (up-convolution) $2 \times 2 \times 2$ với

bước nhảy bằng 2 trên mỗi chiều, tiếp theo là hai lớp tích chập $3 \times 3 \times 3$, mỗi lớp cũng được theo sau bởi một ReLU. Các kết nối tắt (shortcut connections) từ các tầng có cùng độ phân giải ở nhánh phân tích sẽ cung cấp các đặc trưng độ phân giải cao cốt lõi cho nhánh tổng hợp. Tại tầng cuối cùng, một phép tích chập $1 \times 1 \times 1$ sẽ giảm số lượng kênh đầu ra về bằng số lượng nhân phân đoạn (trong bài toán này là 3 nhân). Tổng số tham số của kiến trúc này là 19.069.955. Giống như đề xuất trong [106], tác giả tránh các điểm nghẽn (bottlenecks) bằng cách nhân đôi số lượng kênh ngay trước khi thực hiện max pooling. Chiến lược này cũng được tác giả áp dụng tương tự cho nhánh tổng hợp.



Hình 2.15: Kiến trúc mạng 3D U-Net. Các khối màu xanh biểu diễn các bản đồ đặc trưng (feature maps). Số lượng kênh (channels) được ký hiệu ngay phía trên mỗi bản đồ đặc trưng.

Đầu vào của mạng là một mảnh (tile) ảnh kích thước $132 \times 132 \times 116$ voxel với 3 kênh. Kết quả đầu ra tại tầng cuối cùng lần lượt là $44 \times 44 \times 28$ voxel theo các hướng x, y và z. Với kích thước mỗi voxel là $1.76 \times 1.76 \times 2.04, \mu\text{m}^3$, vùng cảm nhận (receptive field) xấp xỉ của mỗi voxel trong kết quả phân đoạn dự đoán sẽ là $155 \times 155 \times 180, \mu\text{m}^3$. Nhờ vậy, mỗi voxel đầu ra đều có thể tiếp cận đầy đủ thông tin ngữ cảnh để quá trình học đạt hiệu quả cao.

Nhóm tác giả cũng áp dụng kỹ thuật chuẩn hóa theo lô (Batch Normalization - "BN") trước mỗi tầng ReLU. Theo nghiên cứu [107], mỗi lô dữ liệu trong quá trình huấn luyện sẽ được chuẩn hóa dựa trên giá trị trung bình (mean) và độ lệch chuẩn (standard deviation) của chính nó, đồng thời các chỉ số thống kê toàn cục (global statistics) cũng được cập nhật liên tục từ các giá trị này. Tiếp theo đó là một tầng chức năng để mô hình học các tham số tỷ lệ (scale) và độ chệch (bias) một cách tường minh. Khi kiểm thử (test time), việc chuẩn hóa sẽ được thực hiện thông qua các chỉ số thống kê toàn cục đã tính toán cùng các tham số tỷ lệ và độ chệch đã học được. Tuy nhiên, do kích thước lô (batch size) của tác giả bằng 1 và số lượng mẫu rất ít, nên đối với các ứng dụng như thế này, việc tiếp tục sử dụng các chỉ số thống kê hiện tại (current statistics) ngay cả trong giai đoạn kiểm thử sẽ mang lại hiệu quả tốt nhất.

Thành phần quan trọng trong kiến trúc này — cho phép huấn luyện mô hình trên các dữ liệu gán nhãn thưa thớt — chính là hàm mất mát softmax có trọng số (weighted softmax loss function). Bằng cách đặt trọng số của các điểm ảnh (pixels) chưa được gán nhãn bằng 0, mô hình có thể chỉ tập trung học từ các điểm ảnh đã có nhãn, và từ đó có khả năng tổng quát hóa cho toàn bộ khối dữ liệu.

Chi tiết huấn luyện và triển khai mô hình

Nhằm xử lý khối lượng tính toán khổng lồ từ dữ liệu 3D đa phương thức và các mô hình Transformer phức tạp, toàn bộ quy trình huấn luyện và đánh giá được vận hành trên một hệ thống máy chủ trang bị card đồ họa NVIDIA RTX Pro 6000 phiên bản Blackwell Max-Q (96GB GDDR7). Những lợi thế vượt trội về băng thông và dung lượng VRAM cực đại đã cho phép duy trì kích thước vùng ảnh đầu vào (input patch size) ở mức $128 \times 128 \times 192$ voxel mà không cần phải hạ thấp độ phân giải không gian gốc (native spatial resolution).

Phác đồ huấn luyện

Đối với kiến trúc nnU-Net, quy trình huấn luyện tuân thủ theo lịch trình mặc định (tối đa 1000 kỷ nguyên). Đáng chú ý, mô hình nnU-Net (kết hợp tiền huấn luyện + tinh chỉnh) được khởi tạo bằng các trọng số tối ưu từ giải pháp đoạt giải của cuộc thi AutoPET II 2023 (được chia sẻ công khai trên nền tảng Zenodo bởi Isensee và các cộng sự [113]). Đối với các kiến trúc được triển khai qua khung làm việc MONAI (3D U-Net và Swin UNETR), các mô hình được huấn luyện trong 300 kỷ nguyên bằng thuật toán tối ưu AdamW. Tốc độ học ban đầu được thiết lập ở mức 1×10^{-4} kết hợp với lịch trình suy giảm hàm cos (cosine annealing decay schedule) và hệ số suy giảm trọng số là 1×10^{-5} . Kích thước lô tổng thể (overall batch size) được ấn định bằng 8 thông qua cơ chế tích lũy gradient (gradient accumulation).

Hàm mất mát

Hàm mất mát hỗn hợp (hybrid loss function) đã được sử dụng để tối ưu hóa mạng thần kinh, kết hợp tính ổn định của hàm mất mát phân phối entropy chéo (cross-entropy loss) với khả năng xử lý hiện tượng mất cân bằng lớp (class imbalance) vốn là ưu thế của hàm mất mát Soft-Dice. Công thức được định nghĩa cụ thể như sau:

$$L = L_{CE} + L_{Dice} \quad (2.12)$$

Trong đó, các thành phần được tính toán theo công thức sau:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i) \quad (2.13)$$

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i p_i + \varepsilon}{\sum_{i=1}^N y_i + \sum_{i=1}^N p_i + \varepsilon} \quad (2.14)$$

Trong đó, y_i là nhãn thực tế (ground truth) tại voxel thứ i ; p_i là xác suất dự báo của mô hình tại voxel đó; N là tổng số lượng voxel trong không gian ảnh 3D; và ε là một hằng số làm mịn (smoothing constant) cực nhỏ nhằm ngăn chặn lỗi chia cho số không (0).

Suy luận và Hậu xử lý

Quy trình suy luận (inference) trên tập thử nghiệm đã áp dụng phương pháp cửa sổ trượt (sliding-window) với độ chồng chập biên là 50%, kết hợp cùng các kỹ thuật tăng cường dữ liệu tại thời điểm suy luận (TTA - test-time augmentation) bằng cách lấy giá trị trung bình kết quả dự báo từ các thao tác lật ảnh đối xứng (image mirroring). Bước hậu xử lý sử dụng thuật toán phân tích thành phần kết nối (CCA - connected component analysis) để loại bỏ các vùng nhiễu nhỏ và giữ lại các cụm tổn thương có ý nghĩa. Việc thiết lập một ngưỡng xác suất bằng 0,5 kết hợp cùng TTA và CCA đã chứng minh được hiệu quả rõ rệt thông qua một nghiên cứu thử nghiệm loại trừ thành phần (ablation study) nội bộ; kết quả cho thấy quy trình hậu xử lý này giúp giảm 8% diện tích dương tính giả tại các vùng hấp thu FDG sinh lý (ví dụ: mỡ nâu, cơ tim) so với các mặt nạ dự báo thô ban đầu.

2.3.4. Phương pháp đánh giá

Quy trình đánh giá độ chính xác không gian và hiệu năng phát hiện tổn thương tuân thủ nghiêm ngặt theo các hướng dẫn nghiêm ngặt từ bộ khung tiêu chuẩn Metrics Reloaded [75], với các phép toán được thực hiện thông qua thư viện monai.metrics:

- **DSC (Hệ số tương đồng Dice):** Đo lường mức độ chồng chập không gian giữa nhãn dự báo (P) và nhãn thực tế chuẩn (G).

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \quad (2.15)$$

- **HD95 (Khoảng cách Hausdorff tại phân vị thứ 95):** Tính toán khoảng cách lớn nhất giữa hai tập hợp các điểm trên bề mặt nhưng loại trừ đi 5% số lượng điểm dị biệt (outliers) cao nhất, từ đó giảm thiểu tác động nhiễu gây ra bởi các vùng dự báo cô lập và có độ lệch lớn.

$$HD_{95} = \max \left(\max_{p \in P} \min_{g \in G} d(p, g), \max_{g \in G} \min_{p \in P} d(g, p) \right) \quad (2.16)$$

- **ASSD (Khoảng cách bề mặt đối xứng trung bình):** Đo lường khoảng cách trung bình từ bề mặt của vùng phân đoạn tự động đến bề mặt của vùng dán nhãn thủ công và ngược lại. Chỉ số ASSD càng thấp chứng tỏ đường biên dự báo càng bám sát vào nhãn gốc.

$$ASSD = \frac{\sum_{p \in P} d(p, G) + \sum_{g \in G} d(g, P)}{|P| + |G|} \quad (2.17)$$

Bên cạnh đó, độ nhạy (sensitivity) và độ chính xác (precision) cũng được sử dụng nhằm đánh giá năng lực phát hiện tổn thương bệnh lý của mô hình. Các phân tích thống kê lâm sàng sử dụng phép kiểm định định hạng có dấu Wilcoxon (Wilcoxon signed-rank test, với giá trị $p < 0,05$ được coi là có ý nghĩa thống kê), kết hợp cùng phân tích Bland-Altman [114] và hệ số tương quan Pearson (r), nhằm thiết lập mức độ đồng thuận về thể tích khối u chuyển hóa (MTV) giữa các kết quả dự báo của AI và kết quả đánh giá của bác sĩ.

CHƯƠNG 3. KẾT QUẢ VÀ BÀN LUẬN

3.1. Đánh giá bộ dữ liệu

3.1.1. Thông tin hiển thị và lưu trữ bộ dữ liệu

Bộ dữ liệu của luận án đã được thu thập từ 3 bệnh viện: Bệnh viện Ung Bướu Hà Nội, Bệnh viện K và Bệnh viện Ung Bướu TP. Hồ Chí Minh từ năm 2019 đến năm 2023. Bộ dữ liệu bao gồm dữ liệu ảnh PET/CT dưới dạng tệp DICOM đã được ẩn thông tin cá nhân. Các thông tin liên quan khác như: tuổi, giới tính, cân nặng, liều ^{18}F -FDG được tiêm, giai đoạn bệnh.. của 416 bệnh nhân (có mắc ung thư và không) được cung cấp dưới dạng thông tin văn bản, không phải là dữ liệu ảnh (Bảng 3.1). Trong đó, 138 bệnh nhân điều trị tại Bệnh viện Ung Bướu Hà Nội, 138 bệnh nhân điều trị tại Bệnh viện K và 140 bệnh nhân điều trị tại Bệnh viện Ung Bướu TP. Hồ Chí Minh.

Bảng 3.1. Các đặc điểm về tuổi và giới tính của bộ dữ liệu

Chẩn đoán	Giới tính bệnh nhân	Số ca	Tuổi trung bình
Ung thư phổi không tế bào nhỏ	Nam	189	$64,3 \pm 11,2$
	Nữ	111	$65,0 \pm 12,1$
Không ung thư phổi	Nam	42	$41,1 \pm 18,3$
	Nữ	74	$46,8 \pm 16,8$

Trong bộ dữ liệu này, dữ liệu ảnh PET/CT của một bệnh nhân có khoảng 200-300 lát cắt CT và PET, cả ảnh hồi cứu và ảnh tiến cứu. Tổng số lượng dữ liệu là 249.600 lát cắt cho hai loại ảnh PET và CT. Kết quả cho thấy trong nhóm bệnh nhân nghiên cứu có 231 bệnh nhân nam (55,5%) và 185 bệnh nhân nữ (44,5%). Độ tuổi trung bình của nhóm bệnh nhân nghiên cứu là $61,2 \pm 9,9$ tuổi (xem bảng 3.5). Thấp nhất là 25 tuổi, cao nhất là 80 tuổi. Không có sự khác biệt về tuổi giữa bệnh nhân nam và nữ ($p > 0,05$). Bảng 3.2 cho thấy kích thước khối u dao động từ 1,2 cm đến 7,1 cm, trung bình là $2,8 \pm 1,0$ cm.

Bảng 3.2. Thông tin kích thước khối u

Số lượng bệnh nhân	Tỷ lệ (%)	Kích thước của khối u (cm)
151	50,33	≤ 3
89	29,67	$>3-5$
60	20,00	$>5-7$
Kích thước trung bình của các khối u		$2,8 \pm 1,0$ (1,2 – 7,1) cm

Mức độ hấp thu ^{18}F -FDG của khối u trung bình là $5,8 \pm 3,5$. Khối u phổi trái có nồng độ ^{18}F -FDG $\text{SUV}_{\max}$ cao hơn khối u phổi phải. Tuy nhiên, sự khác biệt về mức độ hấp thu ^{18}F -FDG giữa u phổi phải và u phổi trái là không đáng kể ($p=0,21$). $\text{SUV}_{\max}$ của ung thư phổi không phải tế bào nhỏ tăng theo kích thước khối u, rõ nhất khi khối u $>3\text{cm}$. Có mối tương quan thuận giữa $\text{SUV}_{\max}$ và kích thước khối u ($r=0,58$) ([Bảng 3.3](#))

[Bảng 3.3](#) từ nghiên cứu cho thấy các khối u có mức hấp thu ^{18}F -FDG trung bình là $5,8 \pm 3,5$. Giá trị $\text{SUV}_{\max}$ của ung thư phổi tăng theo kích thước khối u. Khi kích thước khối u nhỏ hơn 2cm , khả năng hấp thu $\text{SUV}_{\max}$ trung bình của ^{18}F -FDG là từ $4,2$ đến $2,4$. Kích thước khối u $>2-3\text{ cm}$ làm tăng khả năng hấp thu ^{18}F -FDG ; $\text{SUV}_{\max} = 5,6 \pm 2,7$. Các khối u $>3-5\text{ cm}$ tiếp tục tăng hấp thu ^{18}F -FDG ; $\text{SUV}_{\max} = 8,0 \pm 3,5$; cao hơn đáng kể so với khối u $< 3\text{ cm}$. Có mối tương quan dương (mức độ vừa phải, hệ số tương quan $r = 0,58$). $\text{SUV}_{\max}$ tăng tỷ lệ thuận với kích thước khối u, tăng theo giai đoạn T. Đối với tổn thương $<1\text{ cm}$, mức độ hấp thu ^{18}F -FDG thường thấp nên ảnh CT cần được xem xét kỹ lưỡng để không bỏ sót tổn thương.

Bảng 3.3 Mức độ hấp thu ^{18}F -FDG ($\text{SUV}_{\max}$) theo kích thước khối u

Kích thước khối u (cm)	Số lượng bệnh nhân	$\text{SUV}_{\max}$
≤ 2	53	$4,2 \pm 1,8$
$>2-3$	98	$5,6 \pm 2,7$
$>3-5$	89	$8,0 \pm 3,5$
$>5-7$	60	$17,5 \pm 12,8$
Tổng	300	$5,8 \pm 3,5$

[Bảng 3.4](#) đưa ra một số thông tin của dữ liệu bệnh nhân trong bộ dữ liệu, dữ liệu của từng bệnh nhân được mã hóa và bảo mật các thông tin cá nhân. Mỗi dữ liệu bệnh nhân sẽ có 01 file chi tiết hơn về kết quả chẩn đoán như [Bảng 3.5](#).

Bảng 3.4. Bảng kết quả chụp PET/CT của bệnh nhân

KẾT QUẢ CHỤP PET/CT TOÀN THÂN

Mã dữ liệu: BVK 34

Năm sinh: 1953 **Giới:** Nam

Chẩn đoán: K phổi trái

Phương pháp tiến hành: bệnh nhân nhịn ăn sáng, nồng độ glucose máu 7,3 mmol/l. Sau khi tiêm 7,5 mCi ¹⁸F-FDG 1 giờ vào vị trí tĩnh mạch vùng mu tay trái, bệnh nhân được chụp CT liều thấp, chụp PET từ đỉnh sọ đến giữa đùi. Hình ảnh PET được hiệu chỉnh sự suy giảm (attenuation corrected) bằng CT. Các giá trị hấp thu chuẩn (standardized uptake values – SUV) là các giá trị tối đa tại vùng quan tâm (region of interest) và tính bằng đơn vị g/ml.

Mô tả hình ảnh chụp cắt lớp toàn thân PET/CT:

Đầu - cổ:

- Hình ảnh bắt xạ theo đặc điểm sinh lý ở não. Tăng nhẹ chuyển hóa ¹⁸F-FDG ở amygdal, rìa lưỡi hai bên theo đặc điểm sinh lý.
- Hình ảnh vài hạch ở phía sau cơ ức đòn chũm trái (nhóm V), không tăng chuyển hóa ¹⁸F-FDG.

Lồng ngực:

- Hình ảnh bắt xạ theo đặc điểm sinh lý ở tim. Không phát hiện hình ảnh tràn dịch màng tim và màng phổi 2 bên.
- Hình ảnh nốt mờ ở ngoại vi thùy trên phổi trái có kích thước 21 x 28 mm, bờ tua gai, tăng chuyển hóa ¹⁸F-FDG (SUVmax: 3,6).
- Hình ảnh vài hạch ở trước khí quản, dưới carina, dưới quai động mạch chủ, dưới phế quản gốc trái, tăng nhẹ chuyển hóa ¹⁸F-FDG. Hạch lớn nhất có kích thước 15 x 11 mm (SUVmax: 3,2).
- Hình ảnh hạch rốn phổi trái có đường kính 13 mm, tăng chuyển hóa ¹⁸F-FDG (SUVmax: 3,2).
- Hình ảnh hạch rốn phổi phải có đường kính 10 mm, tăng nhẹ chuyển hóa ¹⁸F-FDG (SUVmax: 3,2).
- Hình ảnh vài hạch nách nhỏ (vài mm) ở hố nách hai bên, không tăng chuyển hóa ¹⁸F-FDG.

Ổ bụng - khung chậu:

- Hình ảnh bắt xạ theo đặc điểm sinh lý ở gan, lách, thượng thận và đài – bể thận 2 bên.
- Hình ảnh dày thành, tăng chuyển hóa ¹⁸F-FDG ở đại tràng trái và đại tràng sigma (SUVmax: 8,1) có thể do viêm. Đối chiếu với nội soi.
- Hình ảnh vài hạch bẹn, kích thước < 10 mm, không tăng chuyển hóa ¹⁸F-FDG ở hố bẹn hai bên.

Bảng 3.4. Bảng kết quả chụp PET/CT của bệnh nhân (tiếp)

Hệ cơ - xương:

Chưa phát hiện hình ảnh tăng chuyển hóa ^{18}F -FDG bất thường ở hệ thống cơ – xương.

Nhận định kết quả:

1. Hình ảnh nốt mờ ở ngoại vi thùy trên phổi trái có kích thước 21 x 28 mm, bờ tua gai, tăng chuyển hóa ^{18}F -FDG tương ứng với tổn thương ác tính nguyên phát.
2. Hình ảnh hạch trung thất (ở trước khí quản, dưới carina, dưới quai động mạch chủ, dưới phế quản gốc trái), hạch rốn phổi trái, tăng nhẹ chuyển hóa ^{18}F -FDG (như mô tả chi tiết ở trên) khả năng do metastases.
3. Hình ảnh hạch rốn phổi phải có đường kính 10 mm, tăng nhẹ chuyển hóa ^{18}F -FDG. Đối chiếu với CT cản quang và sinh thiết (nếu cần).
4. Chưa phát hiện hình ảnh tổn thương tăng chuyển hóa ^{18}F -FDG khu trú, bất thường ở các vị trí khác trên xạ hình PET/CT toàn thân.

Bên cạnh các thông tin về bệnh nhân, chẩn đoán... bộ dữ liệu còn tập hợp dữ liệu ảnh PET/CT của từng bệnh nhân đã phân loại, khoanh vùng khối u ung thư phổi không tế bào nhỏ: bao gồm 01 bộ ảnh gốc đã mã hóa, bảo mật thông tin của bệnh nhân, 03 bộ ảnh do 03 bác sĩ chẩn đoán hình ảnh có kinh nghiệm 5 năm trở lên. Đây sẽ là dữ liệu để nghiên cứu sinh đánh giá, lựa chọn để dùng làm bộ dữ liệu chuẩn để sử dụng trong các nghiên cứu sau về trí tuệ nhân tạo trong lĩnh vực này.

Sau quá trình thu thập, nghiên cứu sinh đã chọn lọc ra được 416 bộ dữ liệu ảnh PET/CT của 416 cá nhân đã được chụp PET/CT tại các bệnh viện: Bệnh viện Ung Bướu Hà Nội, Bệnh viện K và Bệnh viện Ung Bướu TP. Hồ Chí Minh. Các ca bệnh được lựa chọn trong bộ dữ liệu này là các bệnh nhân ung thư phổi tới thăm khám tại bệnh viện trong giai đoạn từ năm 2019 đến năm 2023. Thông qua việc hồi cứu lại kết quả của các bệnh nhân đã được điều trị, nghiên cứu sinh có thể kiểm chứng độ chính xác của dữ liệu, loại bỏ bớt yếu tố con người có thể tác động lên quá trình phân đoạn hình ảnh.

Kết quả được thu thập từ các bệnh nhân có độ tuổi từ khoảng 25 đến 80 tuổi, được lấy ngẫu nhiên không phân biệt giới tính. Trong đó, nữ giới chiếm 44,5% trong tổng số bệnh nhân có trong bộ dữ liệu. Liều bức xạ tiêm cho bệnh nhân dao động từ 6 đến 10 mCi tùy theo cân nặng của bệnh nhân. Ngoài ra trong quá trình thu thập dữ liệu, nghiên cứu sinh cũng thu thập cả các thông tin về sức khỏe của bệnh nhân như cân nặng, chiều cao, tiền sử hút thuốc, uống rượu.

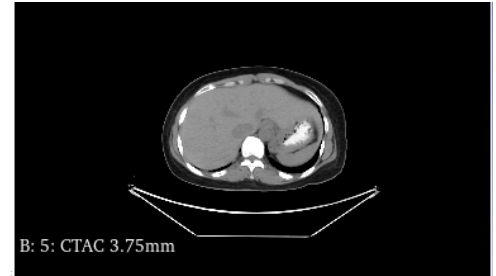
Bảng 3.5: Kết quả đánh giá một số tiêu chí định lượng cho bộ dữ liệu

STT	Tiêu chí	Mức yêu cầu định lượng	Bộ dữ liệu sản phẩm
1.	Quy mô dữ liệu	100 – 200 Bệnh nhân cho nghiên cứu ban đầu	416 bệnh nhân
2.	Chất lượng ảnh	Độ phân giải ảnh PET không thấp hơn 128x128, ảnh CT từ 512x512 (bề dày lớp cắt $\leq 5\text{mm}$)	Độ phân giải ảnh PET 192x192, CT 512x512, chiều dày lớp cắt 3,75mm
3.	Đa dạng thiết bị	Tối thiểu từ 2 hệ thiết bị chụp khác nhau	3 hệ thống chụp ở 3 bệnh viện
4.	Chất lượng gắn nhãn	Có xác nhận từ bác sĩ và thông tin giải phẫu bệnh	Bác sĩ tham gia thực hiện là chuyên gia và có thông tin giải phẫu bệnh đi kèm
5.	Tính đầy đủ của metadata	Đảm bảo $\geq 90\%$ dữ liệu có metadata đầy đủ	95% dữ liệu có thông tin
6.	Định dạng ảnh	DICOM là định dạng chính, có thể xuất thêm NIFTI/PNG nếu cần.	DICOM
7.	Đa dạng dân số học	Có ≥ 3 nhóm tuổi, cả nam và nữ.	Số ca nữ chiếm 40%, độ tuổi từ 25-80
8.	Tính cập nhật	Dữ liệu trong vòng 3-5 năm tại thời điểm nghiên cứu	Dữ liệu 2019-2022
9.	Phiếu đồng thuận bệnh nhân	Có sự thông qua của hội đồng y đức hoặc xác nhận đồng thuận bệnh nhân	Do dữ liệu sử dụng là dữ liệu hồi cứu, việc xin xác nhận tham gia của các bệnh nhân không khả thi, nhóm đã đề xuất và được các cơ sở y tế tham gia đồng ý phối hợp nghiên cứu. Riêng Bệnh viện Ung bướu Tp Hồ Chí Minh đã lập thêm hội đồng đạo đức chấp thuận.

Dưới đây là các thông số đặc trưng hình ảnh của bộ dữ liệu

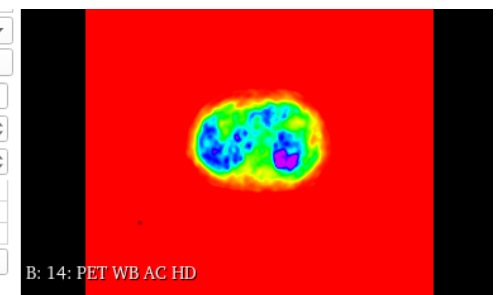
- **Độ phân giải:** ảnh hợp nhất PET/CT thu được trong luận án có kích thước 512 x 512 pixel (điểm ảnh), trong đó ảnh PET có kích thước 192x192 pixel, ảnh CT có kích thước 512 x 512 pixel. Độ phân giải này tương đương với các bộ dữ liệu PET/CT hiện nay sử dụng trong phân đoạn hình ảnh (Hình 3.1 và 3.2).

Active Volume:	5: CTAC 3.75mm		
Volume Information			
Image Dimensions:	512	512	299
Image Spacing:	1.037109mm	1.037109mm	3.260000mm
Image Origin:	265.50mm	265.50mm	-938.73mm
IJK to RAS Direction Matrix:	-1.0000 0.0000 0.0000	0.0000 -1.0000 0.0000	0.0000 0.0000 1.0000
Center Volume			



Hình 3.1. Hình ảnh hiển thị độ phân giải của ảnh CT thu thập trong luận án

Active Volume:	14: PET WB AC HD		
Volume Information			
Image Dimensions:	192	192	299
Image Spacing:	3.6458mm	3.6458mm	3.2600mm
Image Origin:	348.1771mm	348.1771mm	-938.7300mm
IJK to RAS Direction Matrix:	-1.0000	0.0000	0.0000
	0.0000	-1.0000	0.0000
	0.0000	0.0000	1.0000
Center Volume			
Scan Order:	Axial IS		



Hình 3.2. Hình ảnh hiển thị độ phân giải của ảnh PET thu thập trong luận án

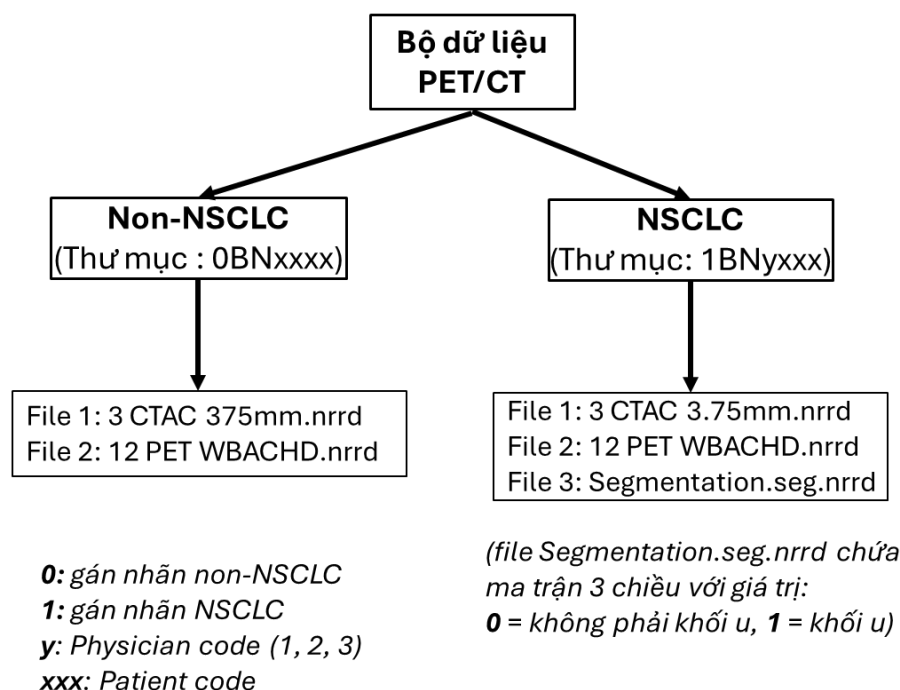
So với nghiên cứu [81], độ phân giải của ảnh CT (512x512 pixel) trong bộ dữ liệu này tương đồng, đảm bảo khả năng phát hiện chi tiết giải phẫu, trong khi độ phân giải của ảnh PET (192x192 pixel) thấp hơn một chút so với các bộ dữ liệu hiện đại (thường là 256x256 pixel) được sử dụng trong nghiên cứu của Lambin [87]. Tuy nhiên, điều này không ảnh hưởng lớn đến phân đoạn do kích thước voxel được chuẩn hóa.

Kích thước của một điểm ảnh trên ảnh CT là $1,037109 \times 1,037109 \text{ mm}^2$ với bề dày của một lát cắt là 3,26 mm. Kích thước của một điểm ảnh trên ảnh PET là $3,6458 \times 3,6458 \text{ mm}^2$ với bề dày của một lát cắt là 3,26 mm. Kích thước voxel trên CT (1,037 mm) nhỏ hơn so với một số nghiên cứu trước đây, thường là 1-2 mm [74], cho thấy độ chi tiết cao hơn, trong khi kích thước voxel PET (3,6458 mm) lớn hơn một chút so với chuẩn EANM (4-6 mm FWHM). Sự khác biệt này có thể yêu cầu hiệu chỉnh thêm trong phân đoạn, nhưng độ dày lát cắt đồng nhất (3,26 mm) giúp tăng tính liên tục, tương tự như dữ liệu trong [77].

- **Độ tương phản chuẩn:** Giá trị HU (ảnh CT) và SUV (ảnh PET) thay đổi giá trị từ thấp đến cao tương ứng với cấu trúc sinh học của từng điểm ảnh thể tích (voxel). Đối với các khối u dương tính với ung thư phổi, giá trị SUV tăng khi kích thước khối u tăng. Các giá trị này đủ để bác sĩ có thể phân biệt được các vùng có khối u phổi và vùng bình thường khác. Xu hướng giá trị SUV tăng theo kích thước khối u phù hợp

với quan sát trong nghiên cứu [79], với $SUV_{max} > 2,5$ được coi là ngưỡng phân biệt khối u ác tính. So với các nghiên cứu trước, bộ dữ liệu này cung cấp dải giá trị HU và SUV đủ rộng, tương tự như dữ liệu trong tài liệu của TCIA [80], hỗ trợ phân đoạn chính xác.

- **Cấu trúc dữ liệu và lưu trữ:**



Hình 3.3. Sơ đồ phân loại và lưu trữ của bộ dữ liệu PET/CT của luận án

Bộ dữ liệu ảnh PET/CT của luận án được phân chia làm 2 nhóm (xem hình 3.3):

- **Nhóm người không mắc bệnh ung thư phổi không tế bào nhỏ:** bao gồm dữ liệu ảnh của những người không bị bệnh và của những bệnh nhân bị ung thư dạng khác. Trong mỗi một thư mục sẽ chứa dữ liệu ảnh PET/CT dưới dạng DICOM của một người không mắc bệnh ung thư phổi không tế bào nhỏ, gồm 02 file có tên và phần mở rộng như hình 3.4a. Người sử dụng có thể đọc 02 file này để xem thông tin về ảnh CT (file 3 CTAC 3.75mm.nrrd) hoặc ảnh PET (file 12 PET WB AC HD.nrrd) của người được chụp. Thư mục chứa dữ liệu ảnh của nhóm này như hình 3.4b. Trong đó, tên của một thư mục được ký hiệu 0BNxxxx, với “0” là nhãn gán cho người không mắc bệnh ung thư phổi không tế bào nhỏ, “xxxx” là mã của cá nhân được chụp ảnh PET/CT. Cấu trúc dữ liệu này tương tự như cách tổ chức trong TCIA [80], nhưng khác với các nghiên cứu trước đây [85] dữ liệu thường không được phân chia rõ ràng giữa nhóm lành tính và ung thư dạng khác, làm tăng tính đại diện của nhóm đối chứng.

Name	Date modified	Type	Size
 3 CTAC 3.75mm	11/3/2023 6:21 PM	NRRD File	70,790 KB
 12 PET WB AC HD	11/3/2023 6:21 PM	NRRD File	15,527 KB

 OBN0001	 OBN0002	 OBN0003
 OBN0004	 OBN0005	 OBN0006
 OBN0007	 OBN0008	 OBN0009
 OBN00010	 OBN00011	 OBN00012
 OBN00013	 OBN00014	 OBN00015
 OBN00016	 OBN00017	 OBN00018
 OBN00019	 OBN00020	 OBN00021
 OBN00022	 OBN00023	 OBN00024
 OBN00025	 OBN00026	 OBN00027

Hình 3.4. Cấu trúc lưu dữ liệu ảnh PET/CT của nhóm người được gán nhãn là không mắc bệnh ung thư phổi không tế bào nhỏ

- Nhóm bệnh nhân mắc bệnh ung thư phổi không tế bào nhỏ:** định dạng ảnh DICOM, chứa thông tin liên quan đến quá trình chụp. Ảnh PET và CT đã được đồng bộ hóa chính xác và được hợp nhất chính xác bằng phần mềm của hãng, điều này rất quan trọng cho việc phân đoạn. Nhóm bệnh nhân này sẽ được các bác sĩ khoanh vùng (phân đoạn khối u). Trong mỗi một thư mục sẽ chứa dữ liệu ảnh PET/CT dưới dạng DICOM của một bệnh nhân mắc bệnh ung thư phổi không tế bào nhỏ, được phân đoạn bởi một trong ba bác sĩ. Thư mục này gồm 03 file có tên và phần mở rộng như hình 3.5. Người sử dụng có thể đọc 03 file này để xem thông tin về ảnh CT (file 3 CTAC 3.75mm.nrrd) hoặc ảnh PET (file 12 PET WB AC HD.nrrd) của người bệnh, file dữ liệu ảnh phân đoạn dưới dạng ma trận ba chiều nhị phân (Segmentation.seg.nrrd). Các điểm ảnh trong file Segmentation.seg.nrrd có giá trị “0” là vị trí đó không phải là khối u, có giá trị “1” thì vị trí đó nằm trong khối u ác tính. Thư mục chứa dữ liệu ảnh của nhóm này như hình 3.5b. Trong đó, tên của một thư mục được ký hiệu 1BNyxxx, với “1” là nhãn gán cho người mắc bệnh ung thư phổi không tế bào nhỏ; “y” là mã của bác sĩ, với $y = 1, 2, 3$ đã thực hiện phân đoạn, “xxx” là mã số bệnh nhân. Việc sử dụng file Segmentation.seg.nrrd và phân đoạn bởi ba bác sĩ là một cải tiến so với các nghiên cứu trước [81] chỉ có một hoặc hai chuyên gia thực hiện, giúp tăng độ tin cậy ($DSC \geq 0.7$). Mức độ đồng bộ hóa PET/CT cũng vượt trội hơn so với các bộ dữ liệu cũ hơn, có độ lệch căn chỉnh thường > 2 mm.

Name	Date modified	Type	Size
 3 CTAC 3.75mm	11/3/2023 6:26 PM	NRRD File	80,134 KB
 12 PET WB AC HD	11/3/2023 6:26 PM	NRRD File	17,265 KB
 Segmentation.seg	11/3/2023 6:26 PM	NRRD File	15 KB

1BN1010	1BN1011	1BN1012
1BN1013	1BN1014	1BN1015
1BN1016	1BN1017	1BN1018
1BN1019	1BN1020	1BN2010
1BN2011	1BN2012	1BN2013
1BN2014	1BN2015	1BN2016
1BN2017	1BN2018	1BN2019
1BN2020	1BN3010	1BN3011
1BN3012	1BN3013	1BN3014
1BN3015	1BN3016	1BN3017
1BN3018	1BN3019	1BN3020

Hình 3.5. Cấu trúc lưu dữ liệu ảnh PET/CT của nhóm bệnh nhân được gán nhãn là mắc bệnh ung thư phổi không tế bào nhỏ

• Dán nhãn và chú thích

- ✓ Chú thích đa chuyên khoa: Các vùng cần phân đoạn đều được dán nhãn bởi 03 chuyên gia có kiến thức về cả hình ảnh giải phẫu và hình ảnh chức năng có kinh nghiệm 5 năm trở lên đã được đề cập trong mục Tiêu chí lựa chọn bác sĩ.
- ✓ Vùng phân đoạn (Region of Interest): Các vùng phân đoạn đã được các bác sĩ khoan, bao gồm khối u, vùng tăng chuyển hóa chức năng từ PET. Các thể tích khối u đều được định dạng dưới dạng ma trận dữ liệu riêng để phù hợp với quá trình học máy tiếp theo. Định dạng ma trận nhị phân (0 và 1) phù hợp với các mô hình học máy hiện đại, vượt trội hơn so với các nghiên cứu cũ chỉ lưu trữ dữ liệu thô [77]. Nghiên cứu của Zhang [88] xác nhận rằng định dạng này tối ưu hóa hiệu suất huấn luyện mô hình deep learning.

• Tính đồng nhất của hình ảnh

- ✓ Độ dày lát cắt: Hình ảnh CT và PET của bộ dữ liệu có độ dày lát cắt đồng nhất là 3,26 mm (xem hình 3.1 và 3.2), đảm bảo tính liên tục và chính xác trong việc phân đoạn.
- ✓ Chất lượng hình ảnh PET: quy trình chụp được tuân thủ theo tiêu chuẩn của bệnh viện nên chất lượng ảnh PET đảm bảo cho việc phân đoạn.

Kích thước bộ dữ liệu

- Số lượng bệnh nhân và hình ảnh: Bộ dữ liệu gồm 416 dữ liệu ảnh, trong đó có 300 dữ liệu ảnh PET/CT của bệnh nhân mắc ung thư phổi không tế bào nhỏ, phần còn lại là dữ liệu ảnh PET/CT của người không mắc ung thư phổi không tế bào nhỏ. So sánh với các nghiên cứu đã phân tích, số lượng của bộ dữ liệu trong nghiên cứu này đủ thống kê cho việc huấn luyện, kiểm tra, lớn hơn ngưỡng tối thiểu 200-300 ảnh được đề xuất trong nghiên cứu [85].
- Tính đa dạng: Bộ dữ liệu khá đa dạng với nhiều lứa tuổi ứng với 2 giới tính nam và nữ, được thu thập tại 03 bệnh viện lớn của Việt Nam.

Bảo mật và quyền riêng tư

- Ẩn danh hóa thông tin: dữ liệu PET/CT đã được ẩn danh trong quá trình tiền xử lý để bảo vệ quyền riêng tư của bệnh nhân, tuân thủ các quy định về bảo mật thông tin y tế.
- Cấp phép sử dụng: Bộ dữ liệu này đã có sự phê duyệt và cấp phép hợp pháp của 03 bệnh viện để sử dụng cho nghiên cứu và học máy.

Tiền xử lý và chuẩn hóa dữ liệu

- Tiền xử lý hình ảnh PET: quá trình tiền xử lý đã sử dụng lọc Gaussian để giảm nhiễu, hiệu chỉnh suy giảm (attenuation correction) và đồng bộ hóa với hình ảnh CT.
- Chuẩn hóa dữ liệu: dữ liệu đã được chuẩn hóa trong quá trình tiền xử lý, phù hợp kích thước và định dạng theo tiêu chuẩn của các bộ dữ liệu dùng cho phân đoạn hình ảnh.

Như vậy, chúng ta có thể thấy bộ dữ liệu đã tuân thủ các yêu cầu cơ bản của bộ dữ liệu sử dụng trong phân đoạn hình ảnh. Bộ dữ liệu PET/CT có chất lượng tốt, phù hợp cho việc phát triển các mô hình học máy nhằm phân đoạn chính xác và hiệu quả các khối u phổi. [Bảng 3.6](#) dưới đây tổng hợp một số thông số chi tiết về bộ dữ liệu:

Bảng 3.6. Một số thông số chi tiết của bộ dữ liệu ảnh PET/CT

Đặc điểm của ảnh			
Độ phân giải (pixel)	Kích thước của một điểm ảnh (mm ²)	Bề dày của một lát cắt (mm)	Ngưỡng chẩn đoán khối u
Ảnh PET			
192x192	3,6458 x 3,6458	3,26	SUV > 2,5
Ảnh CT			
512x512	1,037 x 1,037	3,26	-
File dữ liệu nhóm mắc NSCLC	3 CTAC 3.75mm.nrrd	12 PET WB AC HD.nrrd	Segmentation.seg.nrrd
File dữ liệu nhóm non- NSCLC	3 CTAC 3.75mm.nrrd	12 PET WB AC HD.nrrd	-

3.1.2. Chất lượng bộ dữ liệu PET/CT qua các chỉ số

Tính nhất quán và độ tin cậy của việc phân đoạn rất quan trọng đối với bộ dữ liệu hình ảnh PET/CT, đặc biệt là trong việc xác định thể tích khối u và đòi hỏi mức độ đồng thuận cao giữa các bác sĩ trong các ứng dụng lâm sàng [3, 8]. Trong nghiên cứu này, dữ liệu hình ảnh PET/CT của mỗi bệnh nhân mắc NSCLC được phân đoạn độc lập bởi ba bác sĩ có chuyên môn về hình ảnh ung thư. Để đánh giá mức độ tương

đồng giữa các tập dữ liệu phân đoạn và đảm bảo chất lượng của nhãn phân đoạn, độ sai lệch về thể tích khối u (Volume error-VE) và chỉ số tương đồng Dice (DSC) đã được tính toán, theo các số liệu đã thiết lập để đánh giá phân đoạn hình ảnh y tế [115] [116].

Độ sai lệch tương đối về thể tích khối u (VE) và Chỉ số tương đồng (DSC)

Độ sai lệch tương đối về thể tích khối u chính là sự khác nhau giữa thể tích khối u giữa hai dữ liệu phân đoạn của hai bác sĩ khác nhau. Thông qua việc tính thể tích của khối u từ tệp segmentation, nghiên cứu sinh đã tính toán giá trị độ lệch về thể tích như trong [bảng 3.7](#). với kết quả độ lệch trung bình khoảng 0.2. Mức độ sai lệch thấp này — đặc biệt là giữa bác sĩ thứ nhất và bác sĩ thứ hai — cho thấy sự đồng thuận cao giữa các người quan sát, tương đương hoặc thậm chí vượt trội so với kết quả từ các nghiên cứu PET/CT tương tự [65].

Bảng 3.7. So sánh giá trị độ lệch về thể tích của tập hợp phân đoạn giữa các bác sĩ

So sánh giữa bác sĩ	Khoảng giá trị VE	Số lượng (bệnh nhân)	Tỷ lệ (%)
Bác sĩ 1 và Bác sĩ 2	0,0-0,1	30	10,0
	0,1-0,2	179	59,6
	0,2-0,3	91	30,4
Bác sĩ 2 và Bác sĩ 3	0,0-0,1	21	7,00
	0,1-0,2	148	49,3
	0,2-0,3	131	43,7
Bác sĩ 1 và Bác sĩ 3	0,0-0,1	29	9,7
	0,1-0,2	157	52,3
	0,2-0,3	114	38,0

Chỉ số tương đồng (Dice Similarity Coefficient-DSC) là chỉ số phổ biến nhất để đánh giá mức độ trùng khớp giữa hai tập hợp phân đoạn, chẳng hạn như giữa phân đoạn của hai chuyên gia hoặc giữa chuyên gia và mô hình. Chỉ số này được biểu diễn theo công thức dưới đây:

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (3.1)$$

Trong đó A và B là hai tập hợp thể tích được phân đoạn, $\cap$ là phần giao nhau giữa chúng, và |A|, |B| là kích thước của từng tập hợp.

Để đánh giá chất lượng của bộ dữ liệu, giá trị của DSC dao động từ 0 (không trùng khớp) đến 1 (trùng khớp hoàn toàn). Đối với phân đoạn khối u, các chuyên gia thường yêu cầu DSC đạt ít nhất **0,7 đến 0,8** để coi là đủ trùng khớp.

Bằng cách so sánh giữa các dữ liệu phân đoạn của 03 bác sĩ ứng với từng bệnh nhân, các chỉ số đã được tính toán và có kết quả như trong [bảng 3.8. \[117\]](#)

Bảng 3.8. So sánh giá trị chỉ số tương đồng của tập hợp phân đoạn giữa các bác sĩ

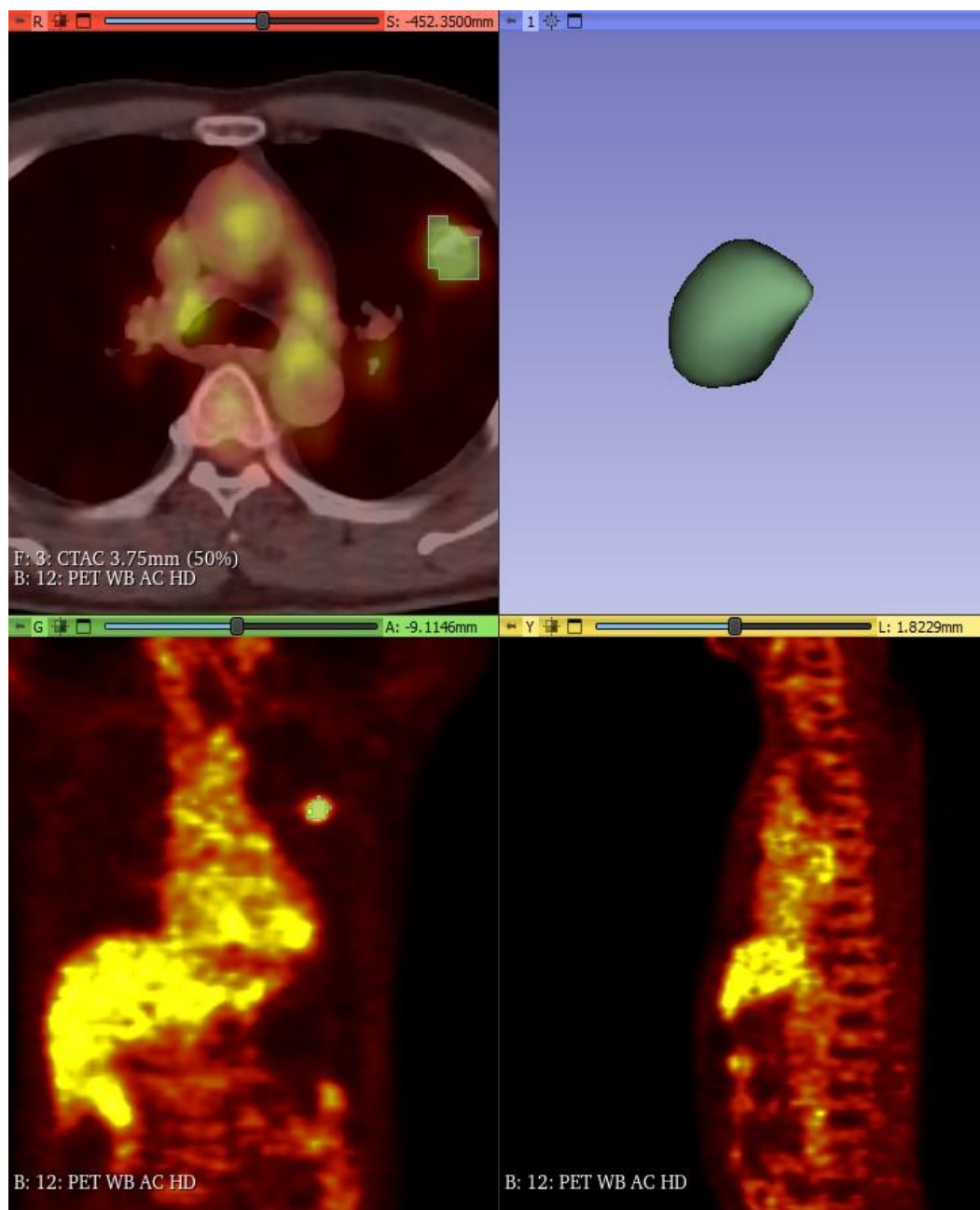
So sánh giữa bác sĩ	Khoảng giá trị DSC	Số lượng (bệnh nhân)	Tỷ lệ (%)
Bác sĩ 1 và Bác sĩ 2	0,9-1,0	32	10,67
	0,8-0,9	183	61,00
	0,7-0,8	85	28,33
Trung bình	0,82±0,08		
Bác sĩ 2 và Bác sĩ 3	0,9-1,0	25	8,33
	0,8-0,9	154	51,33
	0,7-0,8	121	40,33
Trung bình	0,78±0,06		
Bác sĩ 1 và Bác sĩ 3	0,9-1,0	34	11,33
	0,8-0,9	169	56,33
	0,7-0,8	97	32,33
Trung bình	0,80±0,07		
Trung bình cả bộ dữ liệu	0,80±0,07		

Thông qua tính toán, nghiên cứu sinh đã xác định 100% dữ liệu phân đoạn của các bệnh nhân đều đạt độ trùng khớp từ 70% trở lên, vượt qua ngưỡng chấp nhận theo tiêu chuẩn của Image Biomarker Standardization Initiative [\[86\]](#) với $DSC \geq 0,7$ được coi là dấu hiệu của sự đồng thuận tốt giữa các chuyên gia trong phân đoạn y học, đặc biệt với các khối u phức tạp như ung thư phổi không tế bào nhỏ (NSCLC). Về tương quan giữa các bác sĩ, cặp bác sĩ 1 và bác sĩ 2 đạt độ trùng khớp tốt nhất là 82,3%, cao hơn so với mức trung bình 0,75-0,80 trong nghiên cứu của Lambin [\[87\]](#) cho thấy sự phối hợp hiệu quả, có thể do kinh nghiệm tương đồng hoặc sử dụng công cụ như 3D Slicer [\[77\]](#).

Chỉ số tiên lượng dương tính (Positive Predictive Value - PPV) và Độ nhạy (Sensitivity - SE)

PPV và SE là hai chỉ số quan trọng trong đánh giá hiệu quả của các mô hình chẩn đoán hoặc phân loại, bao gồm việc đánh giá bộ dữ liệu y học như ảnh PET/CT trong nghiên cứu ung thư phổi không tế bào nhỏ (NSCLC). Tuy nhiên, việc sử dụng hai chỉ số này để đánh giá trực tiếp chất lượng của bộ dữ liệu (chẳng hạn như tính đại

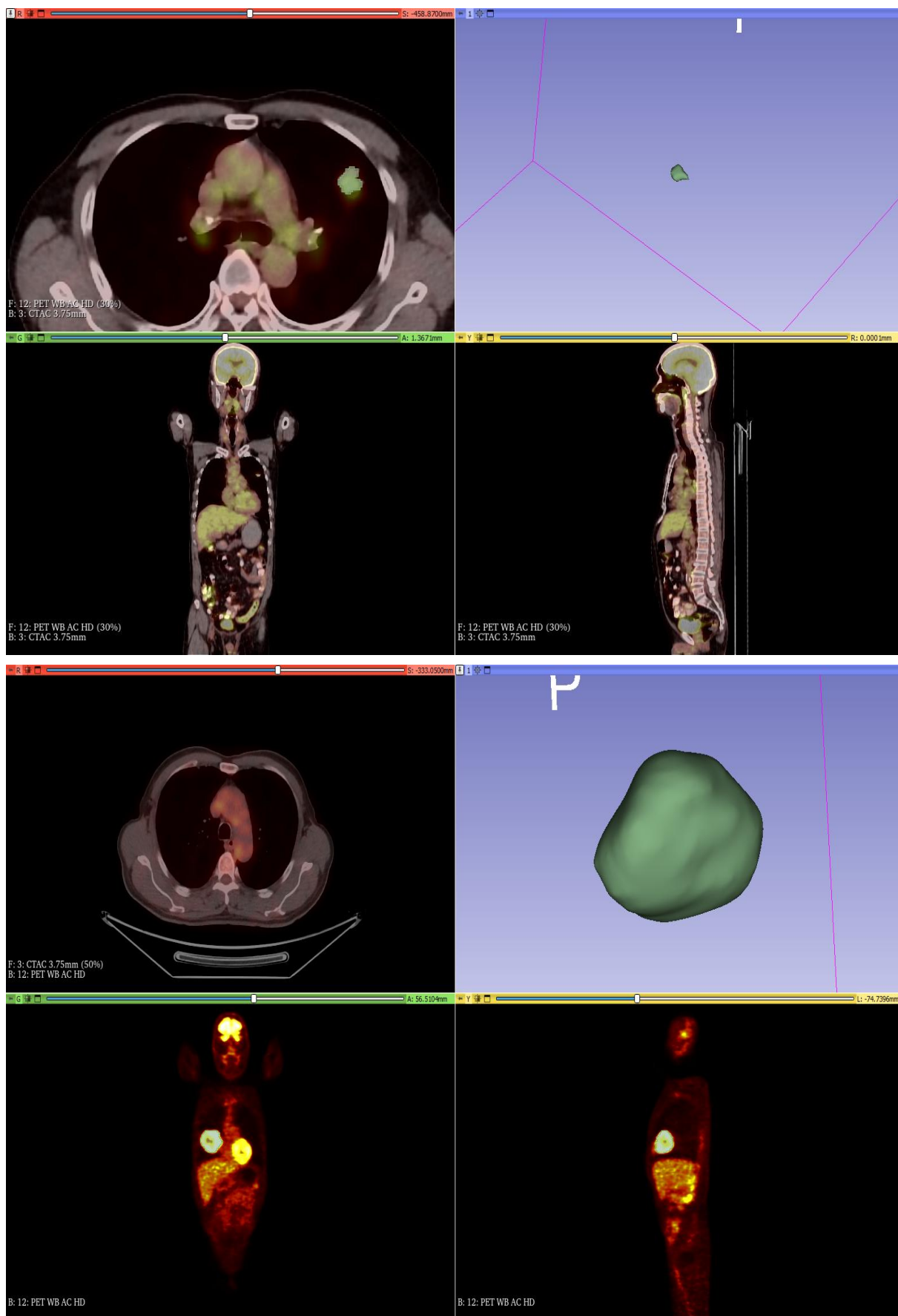
diện, đồng nhất, hoặc chất lượng hình ảnh) không phải là tiêu chuẩn chính. Thay vào đó, PPV và SE thường được sử dụng để đánh giá hiệu suất của các mô hình học máy hoặc thuật toán phân đoạn được huấn luyện trên bộ dữ liệu đó, sau khi dữ liệu đã được xử lý và phân loại [87]. Nghiên cứu [81] nhấn mạnh rằng PPV và SE là các chỉ số hiệu suất (performance metrics) được áp dụng để đánh giá độ chính xác của các mô hình radiomics trong phân loại NSCLC, chứ không phải để đo lường chất lượng dữ liệu thô. Tương tự, Parmar [85] đã chỉ ra rằng các chỉ số này được tính toán dựa trên kết quả dự đoán của mô hình so với tiêu chuẩn vàng (như sinh thiết), thay vì đánh giá các đặc điểm nội tại của bộ dữ liệu như độ phân giải hoặc độ đồng nhất.



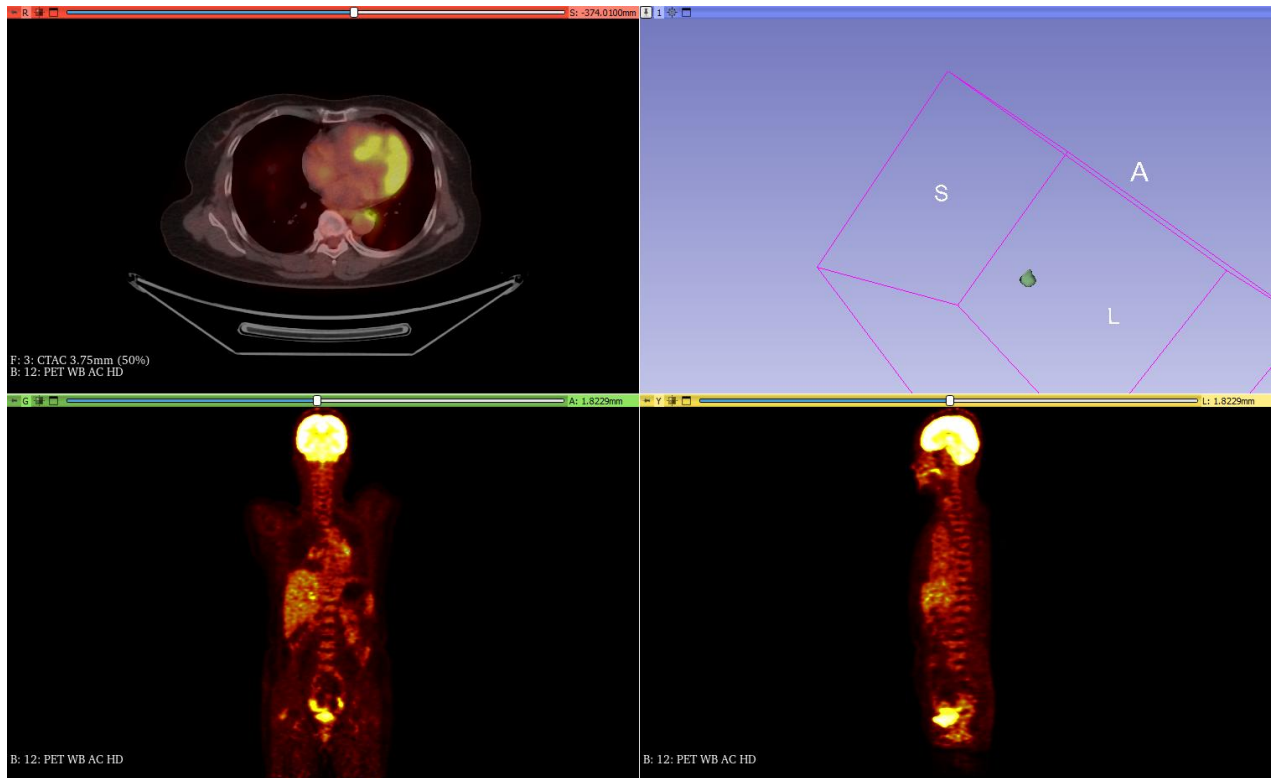
Hình 3.6. Ảnh PET/CT và một ảnh lát cắt của bệnh nhân UBHN27, đã được bác sĩ 01 khoanh vùng và trích trọt thể tích u, xóa các bộ phận khác xung quanh giúp chúng ta có thể quan sát rõ hơn về khối u.



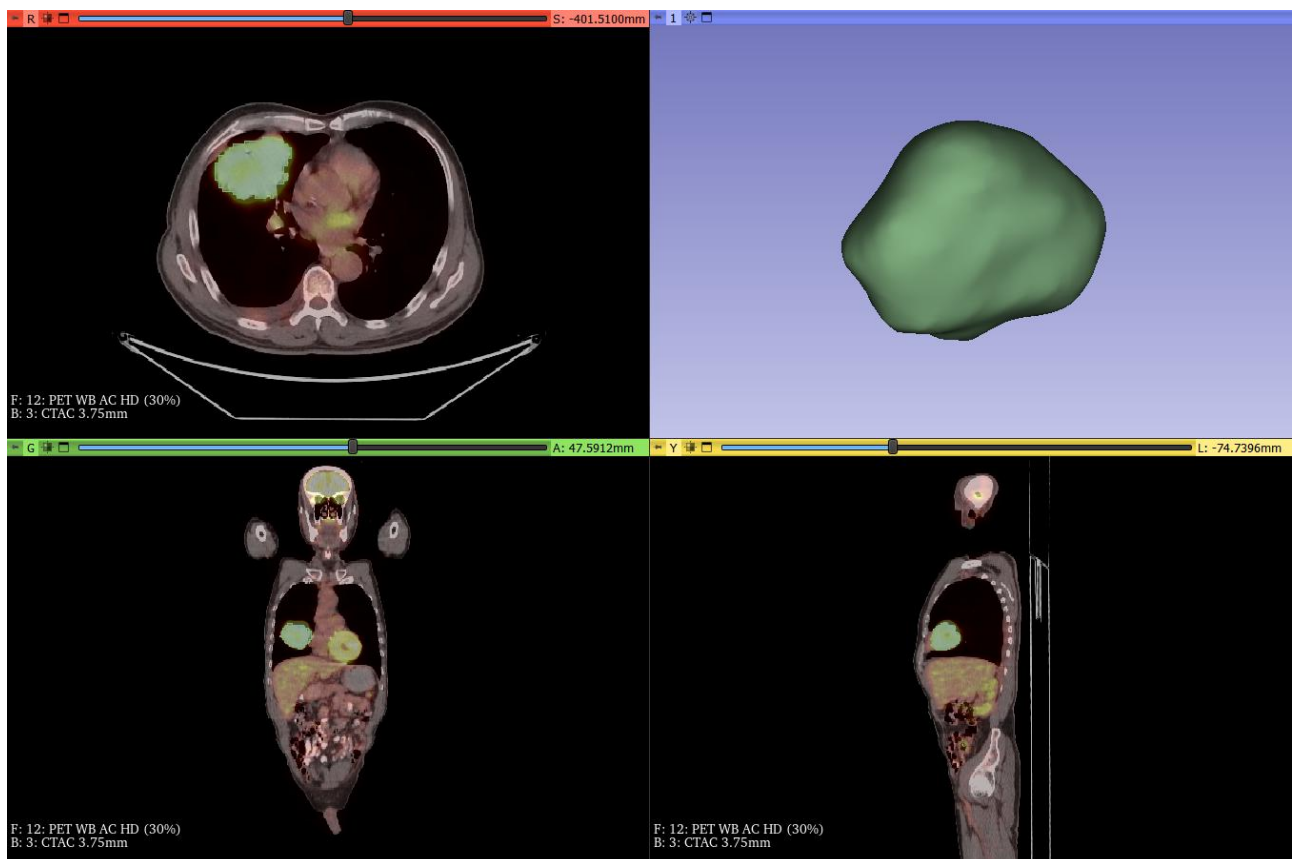
Hình 3.7. Ảnh PET/CT và một ảnh lát cắt của bệnh nhân UBHN27, đã được bác sĩ 02 khoanh vùng và trích trọn thể tích u, xóa các bộ phận khác xung quanh giúp chúng ta có thể quan sát rõ hơn về khối u.



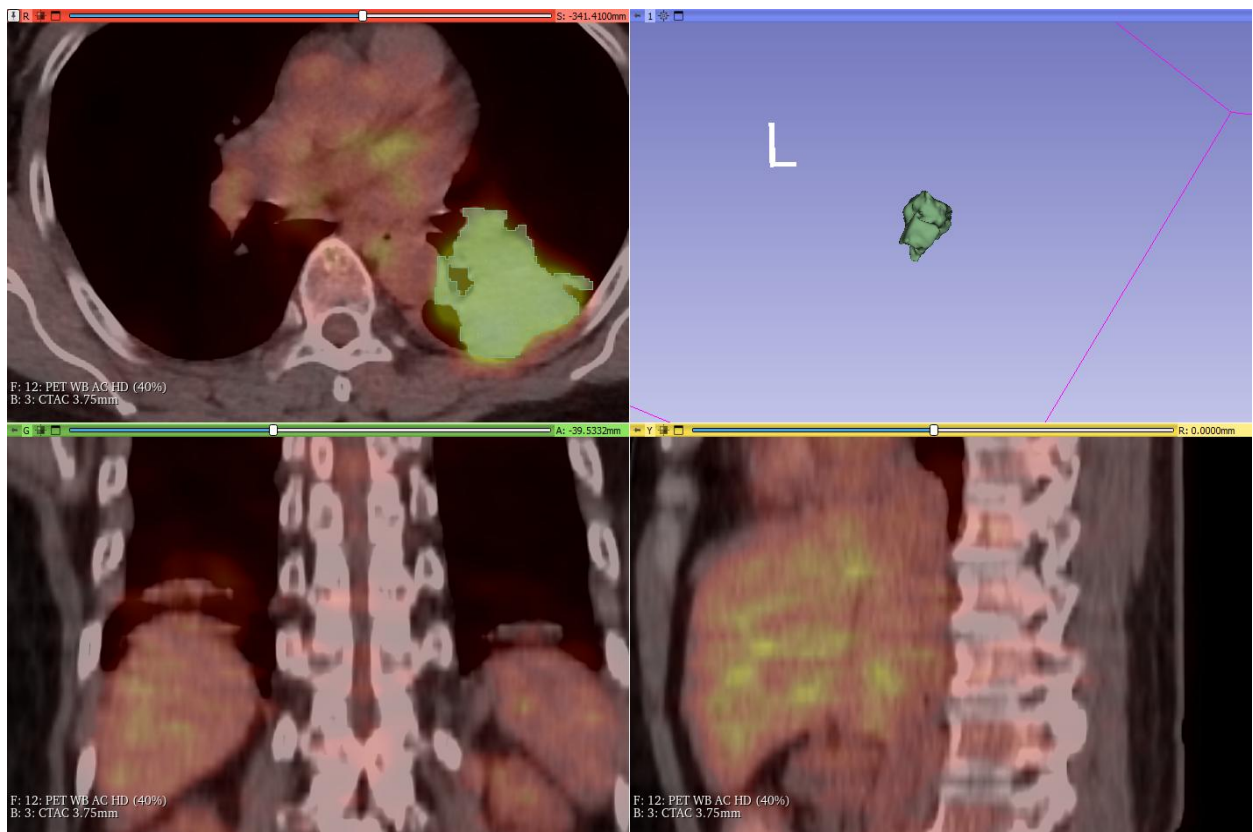
Hình 3.8. Ảnh PET/CT và một ảnh lát cắt của bệnh nhân UBHN27 ở các góc nhìn khác nhau.



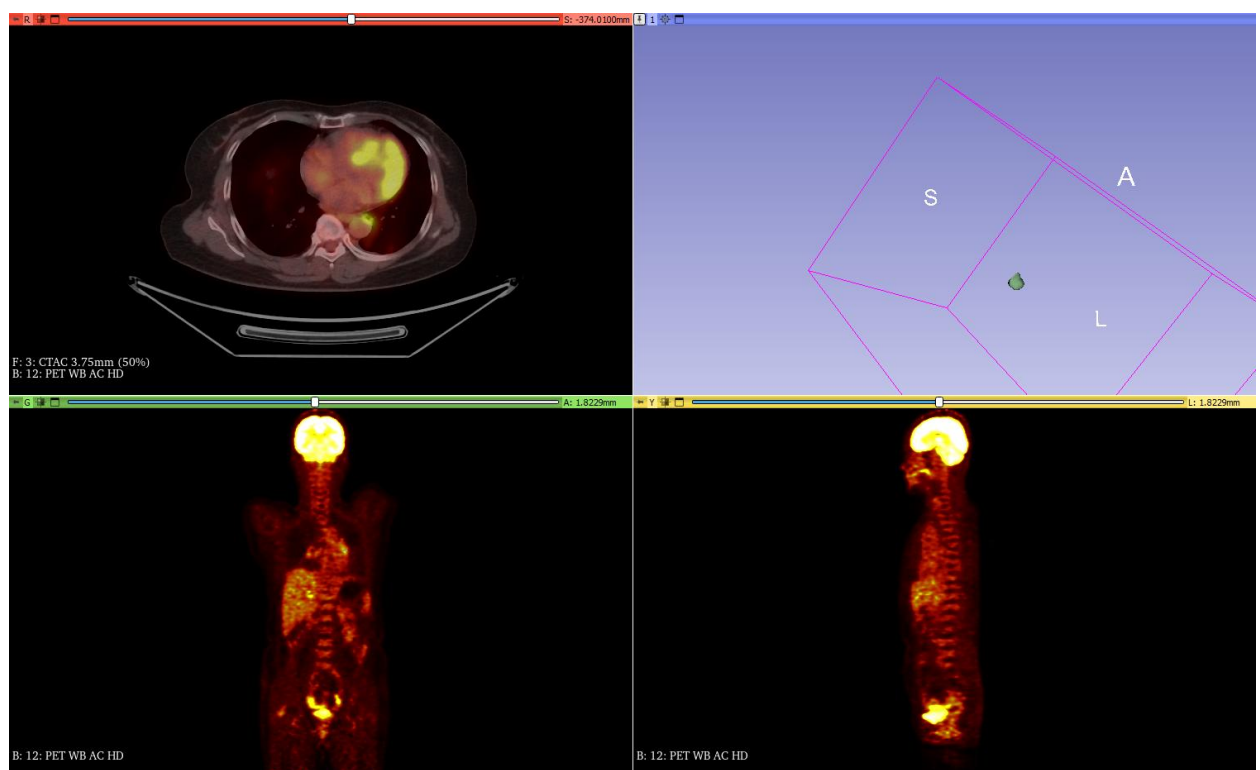
Hình 3.9: Ảnh PET/CT của bệnh nhân BVK – 21 do bác sĩ 01 khoanh vùng khối u



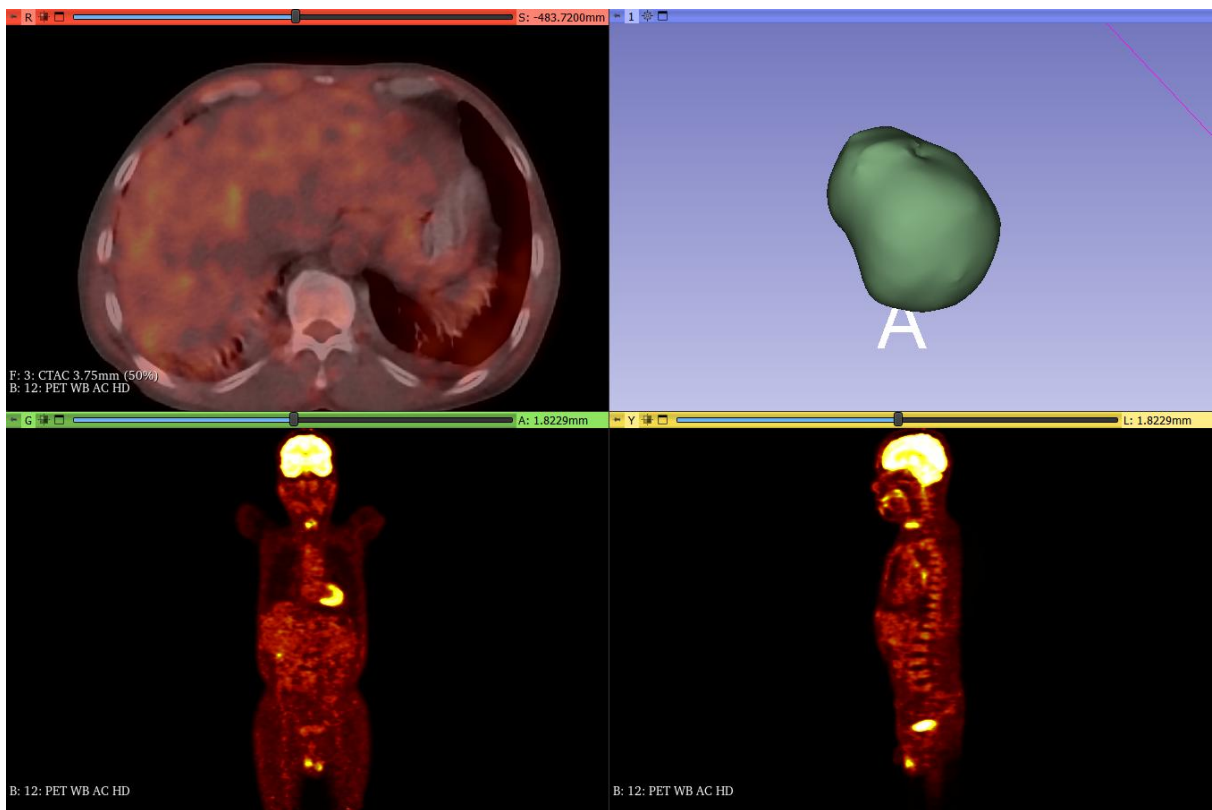
Hình 3.10: Ảnh PET/CT của bệnh nhân BVK – 21 do bác sĩ 02 khoanh vùng khối u



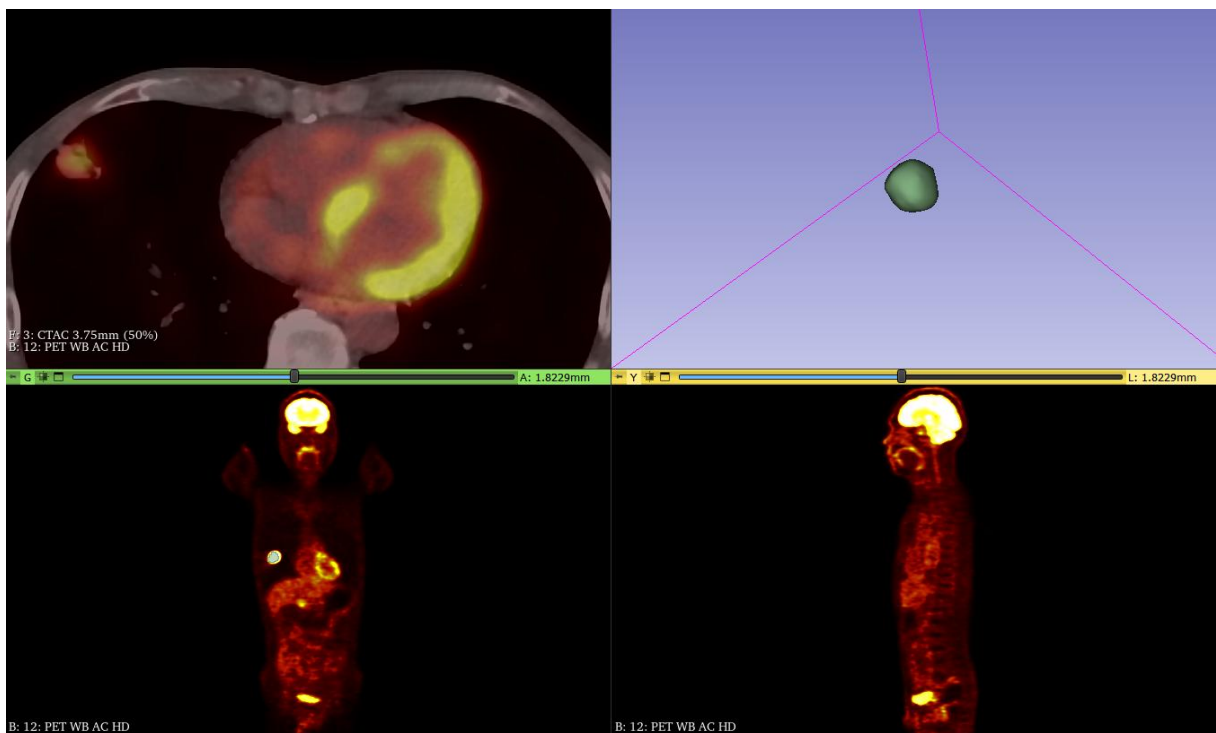
Hình 3.11: Ảnh PET/CT của bệnh nhân UHCM-05 do bác sĩ 01 khoanh vùng khối u



Hình 3.12. Ảnh PET/CT của bệnh nhân UHCM-05 do bác sĩ 02 khoanh vùng khối u



Hình 3.13. Ảnh PET/CT của bệnh nhân UHCM-94 do bác sĩ 01 khoanh vùng khối u



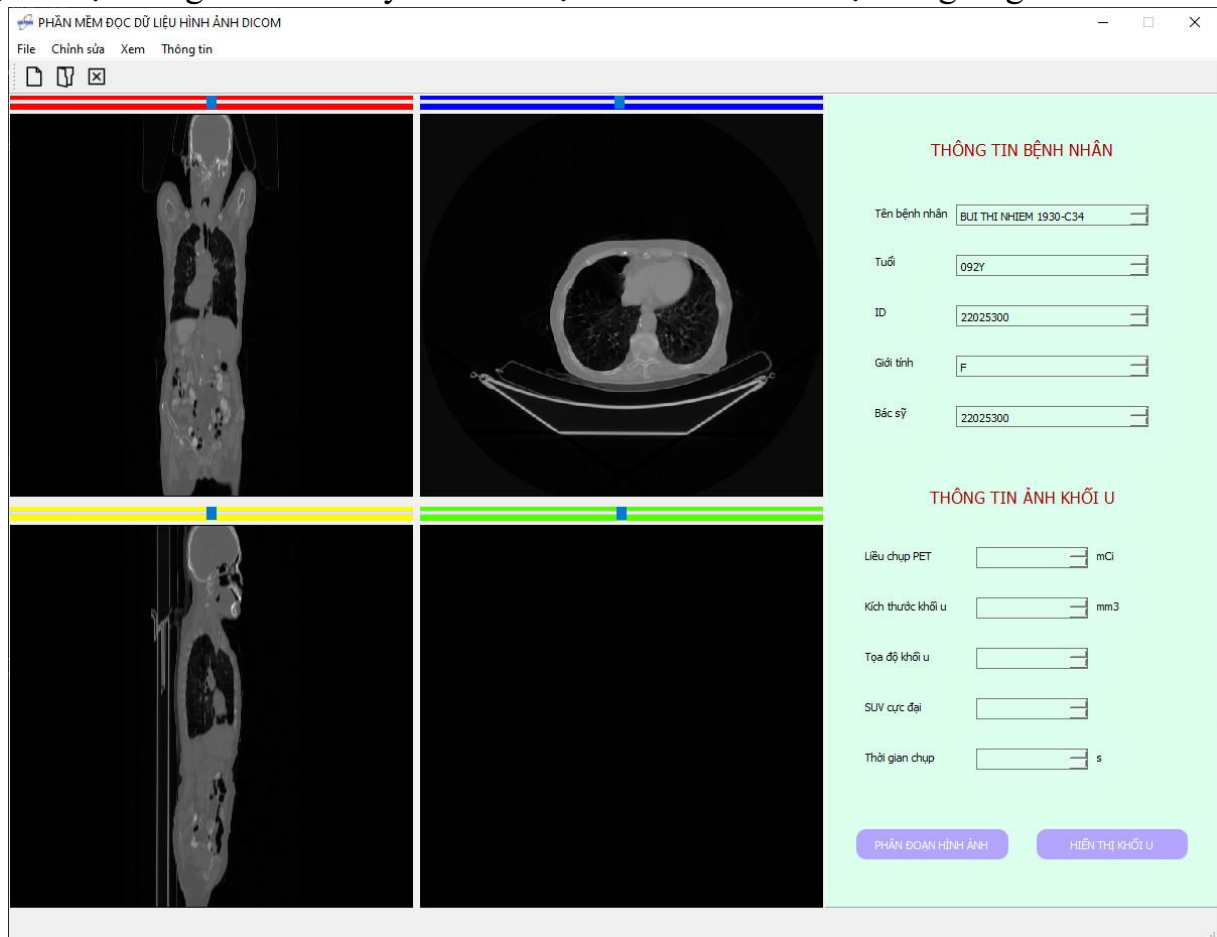
Hình 3.14: Ảnh PET/CT của bệnh nhân UHN-105 do bác sĩ 02 khoanh vùng khối u

3.2. Đánh giá công cụ đọc và tiền xử lý ảnh PET/CT

3.2.1. Giao diện đồ họa của phần mềm hiển thị

Sau khi kết thúc luận án ngoài việc thu thập và phân đoạn thủ công được bộ dữ liệu PET/CT có thể được sử dụng để huấn luyện cho các mô hình học máy giúp hỗ trợ chẩn đoán ung thư phổi. Nghiên cứu sinh cũng xây dựng thành công phần mềm đọc và tiền xử lý hình ảnh PET/CT. Phần mềm có khả năng đọc hình ảnh PET/CT với định dạng ảnh DICOM, tiền xử lý và hiển thị hình ảnh. Giao diện đồ họa của phần mềm được thể hiện như trong [Hình 3.15](#).

Giao diện hiển thị chính của chương trình được chia thành bốn khu vực ([Hình 3.15](#)) để lần lượt hiển thị lát cắt ngang, lát cắt đứng và lát cắt bên của dữ liệu. Một phần hiển thị còn lại của phần mềm sẽ được dùng để hiển thị hình ảnh khối u đã được bác sĩ phân đoạn thủ công. Phần mềm cho phép có thể hiển thị nhiều lát cắt khác nhau trong bộ dữ liệu bằng cách di chuyển thanh dịch ở cửa sổ hiển thị tương ứng.



Hình 3.15. Giao diện chương trình và hiển thị ảnh DICOM

Ngoài phần hiển thị hình ảnh, phần mềm còn có mục bên phía tay phải cửa sổ giao diện sử dụng để hiển thị thông tin của bệnh nhân đọc được từ bộ dữ liệu. Chương trình sẽ truy xuất dữ liệu từ tệp tiêu đề của hình ảnh DICOM sau đó hiển thị. Tương tự

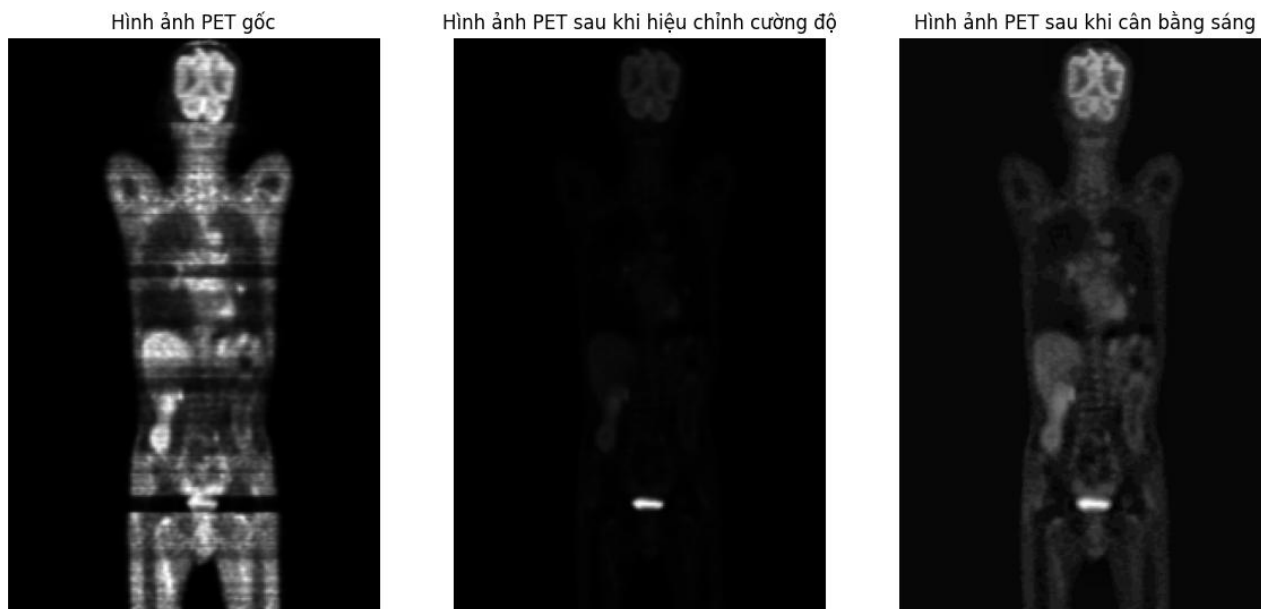
như vậy, thông tin của khối u sau khi đã được bác sĩ phân đoạn sẽ được phần mềm tích toán và hiển thị để người dùng có thể kiểm tra và xác nhận lại độ chính xác của bộ dữ liệu thu được.

Phần mềm có thể cài đặt được trên các máy tính sử dụng hệ điều hành Windows hoặc Linux, nhìn chung, phần mềm hiển thị đã đáp ứng được đầy đủ các yêu cầu đề ra của luận án. Tốc độ xử lý của phần mềm còn chưa được tốt vì được xây dựng trên nền tảng đơn nhân, thời gian xử lý các bộ dữ liệu lớn sẽ rất lâu. Phần mềm còn cần được tiếp tục phát triển khả năng xử lý đa luồng, xử lý dữ liệu kết hợp với card đồ họa để cho thời gian xử lý thực tiễn hơn. Trên cơ sở phần mềm đọc dữ liệu, các mô đun học máy có thể được tích hợp thêm vào để tạo nên phần mềm hỗ trợ chẩn đoán hoàn chỉnh.

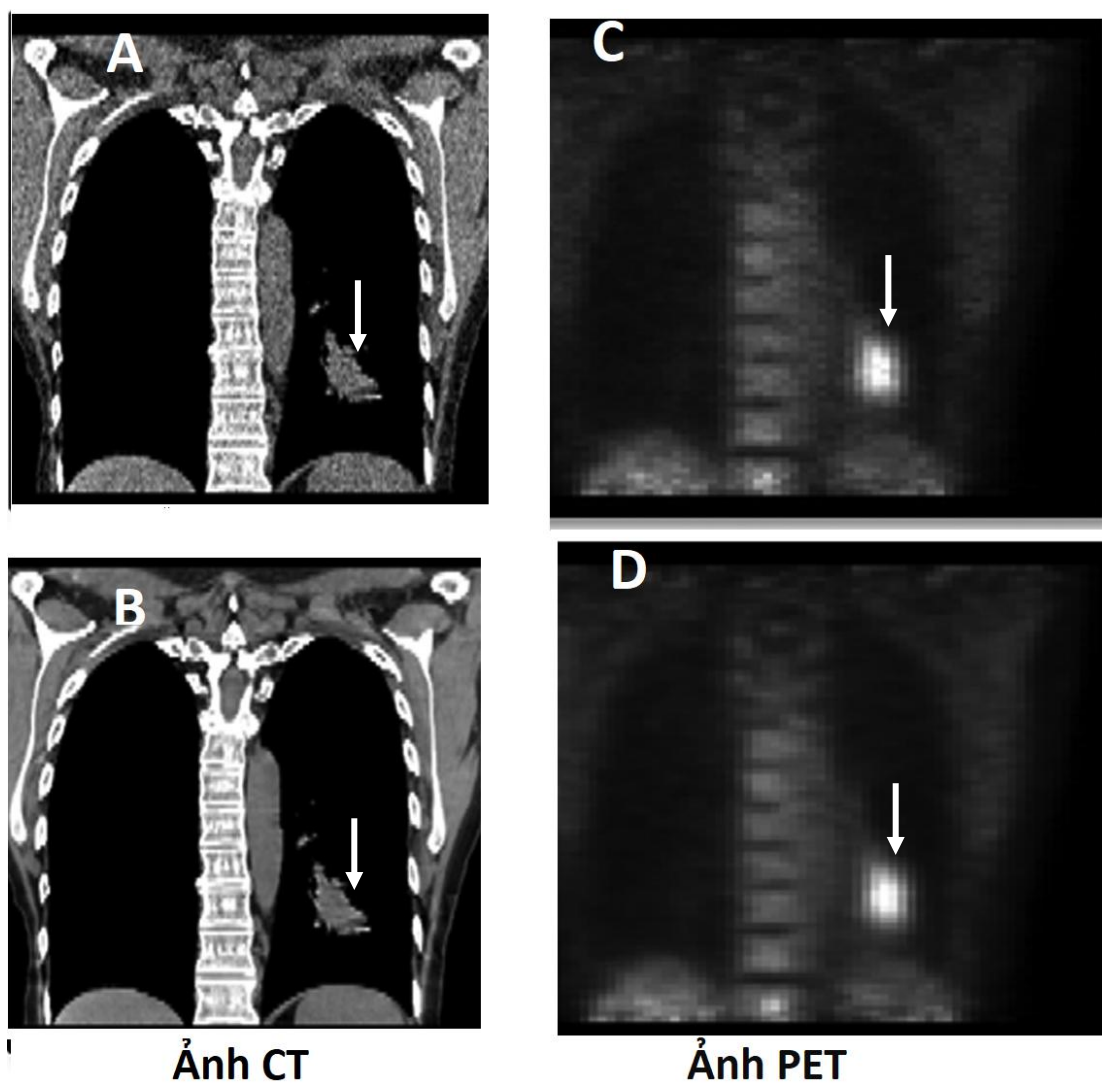
3.2.2. Chức năng tiền xử lý của phần mềm

Sau khi xây dựng hoàn thiện phần mềm đọc và xử lý hình ảnh PET/CT, nghiên cứu sinh đã tiến hành đọc hình ảnh từ bộ dữ liệu mà nhóm đã thu thập được tại các bệnh viện trên cả nước. Hình ảnh của quá trình trước và sau khi hiệu chuẩn cường độ điểm ảnh của các lát cắt hình ảnh PET được thể hiện trên [Hình 3.16](#). Kết quả chỉ ra rằng, trước khi thực hiện hiệu chuẩn cường độ, hình ảnh PET nhìn rất mờ, với nhiều sọc ngang thể hiện sự sai lệch cường độ hiển thị của các lát cắt. Sau khi hiệu chỉnh sự đồng nhất của cường độ, cường độ của mỗi lát cắt sẽ được lấy tỉ lệ theo cường độ của vùng có chỉ số SUV lớn nhất. Thông thường thận là vùng bài tiết phóng xạ nên sẽ có sự tập trung hoạt độ lớn nhất, lớn hơn nhiều so với cả khối u hấp thu. Do đó, để quan sát hình ảnh rõ hơn ta cần phải thực hiện cân bằng sáng của hình ảnh. Hình ảnh sau khi cân bằng sáng đã trở nên dễ dàng quan sát hơn, với vùng cường độ phân bố đều theo toàn thang hiển thị của hình ảnh.

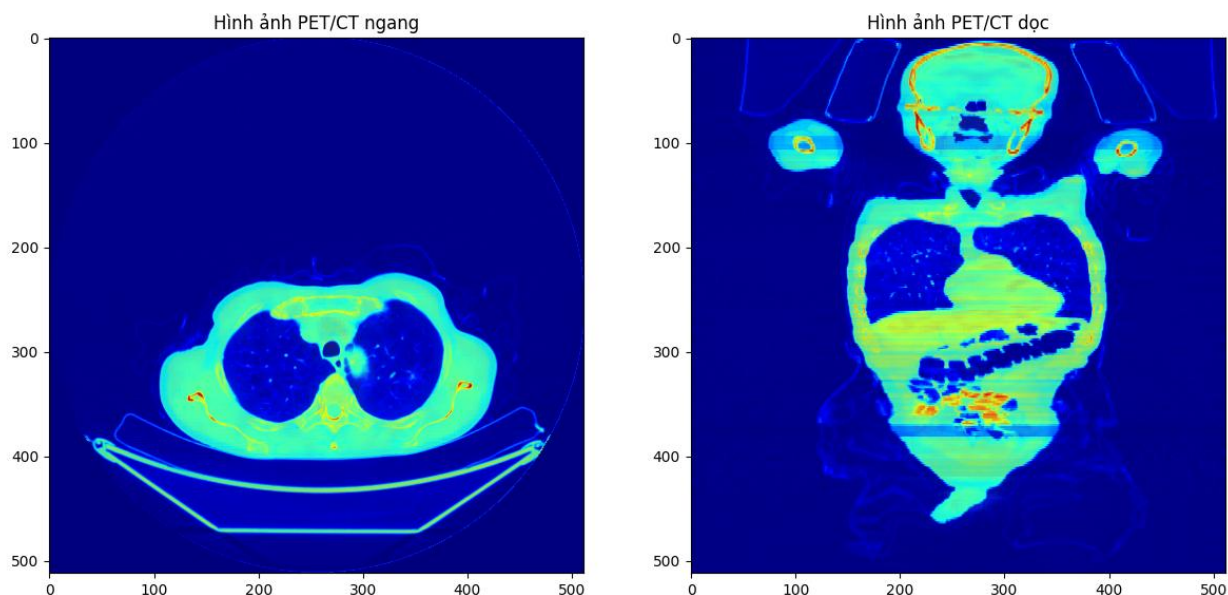
[Hình 3.17](#) chỉ ra ảnh CT và ảnh PET trước và sau khi loại bỏ nhiễu bằng cách sử dụng hàm Gauss. [Hình 3.17A](#), [3.17C](#) lần lượt là hình ảnh CT và hình ảnh PET khi chưa sử dụng hàm lọc. [Hình 3.17B](#), [3.17D](#) là hình ảnh CT và PET đã được áp dụng hàm lọc Gauss với cửa sổ hàm lọc chỉ bằng 3. Dựa trên hình ảnh ta có thể thấy rằng, hình ảnh sau khi lọc trở nên mịn hơn, ồn nhiễu đã giảm đi đáng kể. Tính toán tỷ số tín hiệu trên tạp âm đỉnh của hình ảnh (Peak Signal-to-Noise Ratio: PSNR), ta có thể thấy rằng, giá trị PSNR của ảnh CT tăng lên từ 19 thành 28, trong khi giá trị PSNR của ảnh PET tăng lên từ 14 thành 24. Ảnh PET phản ánh hoạt độ phóng xạ thực tế. Do đó, các kỹ thuật tiền xử lý (đặc biệt là lọc nhiễu hay chuẩn hóa) không được làm mất đi tổng lượng tín hiệu (hoạt độ) trong toàn bộ thể tích quét hoặc trong một cơ quan đích. Đối với các kỹ thuật lọc giữ nguyên cân bằng phóng xạ, tổng giá trị cường độ pixel/voxel trước và sau khi tiền xử lý này phải được giữ nguyên hoặc hiệu chỉnh đúng hệ số tỷ lệ. Hơn nữa, mối quan hệ giữa cường độ tín hiệu pixel sau tiền xử lý và nồng độ phóng xạ thực tế vẫn giữ nguyên tính tuyến tính.



Hình 3.16. Hình ảnh PET trước và sau khi thực hiện hiệu chỉnh cường độ

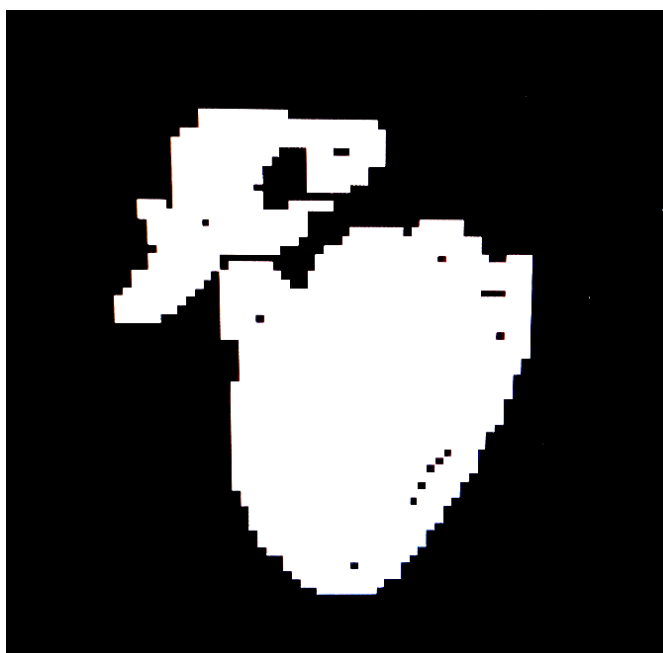


Hình 3.17. Ảnh PET/CT trước (A và C) và sau khi sử dụng hàm lọc Gauss (B và D)



Hình 3.18. Hình ảnh hợp nhất PET và CT theo trục ngang và theo trục dọc

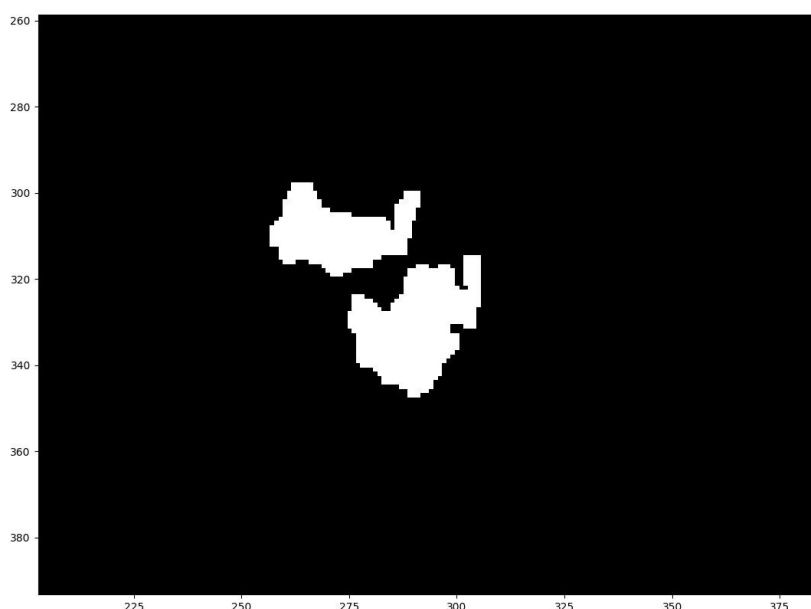
Hình 3.18 mô tả kết quả hợp nhất hình ảnh PET/CT sau khi đã thực hiện các bước hiệu chỉnh, lọc nhiễu. Hình ảnh PET và hình ảnh CT được hợp nhất với nhau dựa trên góc tọa độ là tâm của hình ảnh (tương ứng với trục Z của hệ PET/CT). Do kích thước trường nhìn của hình ảnh PET và hình ảnh CT có khác nhau, nên ta cần phải tiến hành cắt ảnh (tính từ tâm ra) để hai hình ảnh có cùng kích thước. Trong chương trình này, hình ảnh PET sẽ được lấy tỉ lệ để có cùng kích thước với hình ảnh CT, như vậy sẽ dễ dàng hơn cho việc khoanh vùng, tính kích thước khối sau này.



Hình 3.19. Thẻ tích khối u tại một lát cắt xuất hiện vài điểm không được đánh dấu là khối u phổi

Ngoài ra, chương trình có thể xoay ảnh, tăng độ tương phản, độ sáng của ảnh, chuẩn hóa về kích thước giữa ảnh của các bệnh nhân khác nhau... để quá trình huấn luyện máy học tăng độ chính xác.

Hình 3.19 và 3.20 thể hiện kết quả đọc thông tin của khối u từ bộ dữ liệu đã được phân đoạn thủ công bởi bác sĩ. Khối u từ hình ảnh đã được bác sĩ đánh dấu, chương trình sẽ đọc giá trị và tọa độ của các điểm ảnh đã được đánh dấu và phân loại vào hình ảnh khối u. Các điểm đã phân loại này sau đó sẽ được lưu lại để huấn luyện học máy. Tuy nhiên, có một vấn đề, trong quá trình bác sĩ phân đoạn thủ công có thể không đánh dấu kín được khối u (Hình 3.19). Việc này làm dữ liệu phân đoạn trở nên không chính xác, khó xác định vị trí và thể tích khối u. Nghiên cứu sinh đã viết chương trình xử lý và chuẩn hóa lại hình ảnh phân đoạn từ dữ liệu mà bác sĩ đã cung cấp. Kết quả Hình 3.20 thể hiện khối u sau khi được đọc thông tin đã được xử lý để trở nên liền mạch hơn mà đảm bảo không làm mất đi hay thêm các thông tin thừa vào khối u. Khối u sau khi được xử lý vẫn có biên dạng đồng nhất với khối u được đánh dấu bởi bác sĩ.



Hình 3.20. Hình ảnh của một lát cắt dữ liệu sau khi đã hiệu chỉnh lại

3.2.3. Kết quả trích xuất thông tin khối u

Việc tính toán một số thông số của bộ dữ liệu PET/CT sau khi đã qua tiền xử lý là việc rất quan trọng để cho người dùng nắm được các thông tin cụ thể hơn về bộ dữ liệu. Các thông số này sẽ được lưu dưới dạng file excel, với các thông số cụ thể như: số lượng khối u, vị trí tâm khối u, kích thước khối u theo ba chiều khác nhau, thể tích, chỉ số HU và SUV trong khu vực mà bác sĩ đang nghi ngờ là khối u (mô tả trên hình 3.21 và hình 3.22).

PatientID					PatientID												
A	B	C	D	E	A	B	C	D	E	F	G	H	I				
Node1		Node1 Vol			PatientName	PatientID	No of Nod	CoordX1	CoordY1	CoordZ1	W1	H1					
HOANG QUANG HANH 1965 - C34	(75, 99, 96, 120, 186, 201)	109327.25454693074			1	HOANG QUANG HANH 1965 - C34	19048659	1	96	78	186	13	12				
NGUYEN DINH BINH 1954 - C34	(78, 90, 77, 91, 154, 160)	12306.35762636874			2	HOANG QUANG HANH 1965 - C34	19048659	1	96	76	187	14	15				
NGUYEN QUOC CHI 1950 - C34	(109, 118, 105, 115, 191, 197)	7279.817187429395			3	HOANG QUANG HANH 1965 - C34	19048659	1	97	76	188	14	16				
NGUYEN THI LU 1943 - C34	(70, 76, 106, 112, 156, 163)	2729.931445286023			4	HOANG QUANG HANH 1965 - C34	19048659	1	99	76	189	16	18				
TRIEU NGUYEN TINH 1953 - C34	(107, 120, 92, 107, 191, 202)	25176.034439859992			5	HOANG QUANG HANH 1965 - C34	19048659	1	98	76	190	17	18				
					6	HOANG QUANG HANH 1965 - C34	19048659	1	98	75	191	18	19				
					7	HOANG QUANG HANH 1965 - C34	19048659	1	97	75	192	17	19				
					8	HOANG QUANG HANH 1965 - C34	19048659	1	97	75	193	18	19				
					9	HOANG QUANG HANH 1965 - C34	19048659	1	98	75	194	17	19				
					10	HOANG QUANG HANH 1965 - C34	19048659	1	98	76	195	16	18				
					11	HOANG QUANG HANH 1965 - C34	19048659	1	99	76	196	16	17				
					12	HOANG QUANG HANH 1965 - C34	19048659	1	100	77	197	15	16				
					13	HOANG QUANG HANH 1965 - C34	19048659	1	99	77	198	14	11				
					14	HOANG QUANG HANH 1965 - C34	19048659	1	100	78	199	11	9				
					15	HOANG QUANG HANH 1965 - C34	19048659	1	101	79	200	9	8				
					16	HOANG QUANG HANH 1965 - C34	19048659	1	102	80	201	7	6				
					17	NGUYEN DINH BINH 1954 - C34	19050100	1	79	80	154	5	3				
					18	NGUYEN DINH BINH 1954 - C34	19050100	1	77	79	155	10	6				
					19	NGUYEN DINH BINH 1954 - C34	19050100	1	77	78	156	11	8				
					20	NGUYEN DINH BINH 1954 - C34	19050100	1	77	79	157	10	7				

Hình 3.21. Một số thông số về các node đã phân loại

seriesuid	coordX	coordY	coordZ	class	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-56.08	-67.85	-311.92	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	53.21	-244.41	-245.17	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	103.66	-121.8	-286.62	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-33.66	-72.75	-308.41	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-32.25	-85.36	-362.51	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-26.65	-203.07	-165.07	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-74.99	-114.79	-311.92	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-16.14	-248.61	-239.55	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	135.89	-141.41	-252.2	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	90.10228	-68.4308	-218.243	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-45.8738	-164.127	-252.946	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-49.2502	-126.472	-280.633	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-27.7315	-132.699	-276.784	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	104.1648	-211.686	-227.011	1	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-28.8342	-104.182	-199.992	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	81.16818	-126.902	-275.587	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-54.8174	-163.714	-265.795	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-37.3001	-144.115	-207.072	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-94.6956	-84.5251	-315.802	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-115.03	-96.1194	-347.971	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	54.76565	-108.049	-263.287	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	77.90487	-75.2714	-341.653	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-56.78	-153.32	-249.39	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	54.62	-149.12	-262.74	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-83.4	-98.67	-324.57	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	68.63	-126	-169.29	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	81.24	-187.66	-270.46	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-35.06	-229.69	-279.6	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-84.8	-83.96	-394.83	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-138.05	-150.52	-335.81	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-77.8	-85.36	-405.37	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-77.8	-78.36	-394.83	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-33.9897	-95.7604	-257.353	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	111.8843	-100.634	-372.928	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	-80.5917	-77.265	-286.578	0	
1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222365663678666836860	74.1224	62.0870	206.505	0	

Hình 3.22. Một số thông số của bộ dữ liệu

3.3. **Đánh giá chất lượng nền tảng tự động phân đoạn hình ảnh ứng dụng mô hình học sâu**

3.3.1. **Kiểm định chéo mô hình**

Hiệu năng trên bộ dữ liệu nội bộ

Các kết quả từ quy trình kiểm định chéo 5 tập (5-fold cross-validation) trên bộ dữ liệu PET/CT của bệnh nhân Việt Nam được báo cáo chi tiết nhằm xác định kiến trúc mô hình tối ưu. [Bảng 3.5](#) trình bày hiệu năng phân đoạn định lượng của ba mô hình học sâu tiên tiến (được huấn luyện từ đầu - trained from scratch), đặt trong tương quan so sánh với mô hình áp dụng chiến lược học chuyển giao (tinh chỉnh từ các trọng số tiền huấn luyện - fine-tuned from pretrained weights). Có thể thấy rõ từ [Bảng 3.9](#), mô hình nnU-Net (tiền huấn luyện + tinh chỉnh) đã chứng minh được sự vượt trội rõ rệt. Việc tận dụng các đặc trưng tổng thể kết hợp cùng quy trình tinh chỉnh trên nguồn dữ liệu bản địa đã giúp mô hình này đạt được chỉ số DSC cao nhất, ở mức $83,44 \pm 6,5\%$. Đồng thời, mô hình này cũng ghi nhận các sai số khoảng cách không gian thấp nhất, với chỉ số HD95 chỉ ở mức $5,1 \pm 3,6$ mm và ASSD đạt $2,0 \pm 2,2$ mm, thể hiện khả năng bám sát cực tốt và chính xác vào các đường biên của khối u. Độ chính xác (precision) của kiến trúc tinh chỉnh này cũng đạt hiệu năng cao nhất ($89,60 \pm 8,2\%$), cho thấy khả năng kiểm soát xuất sắc đối với các trường hợp dương tính giả tại các vùng hấp thu FDG sinh lý.

Bảng 3.9: So sánh định lượng hiệu năng phân đoạn giữa các kiến trúc học sâu khác nhau trên bộ dữ liệu nội bộ Việt Nam

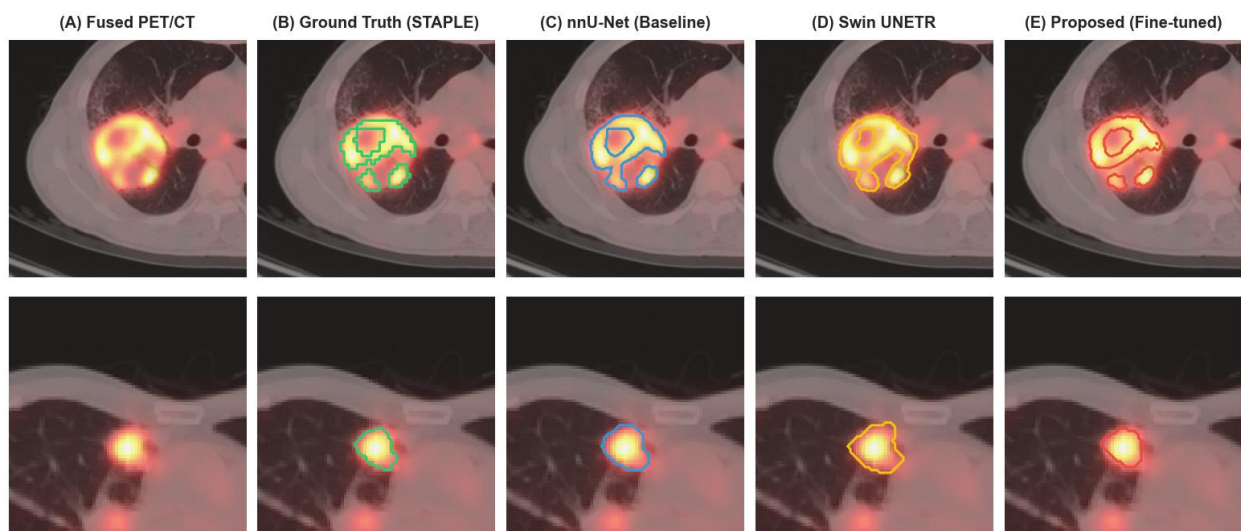
Model	DSC (%) ↑	HD95 (mm) ↓	ASSD (mm) ↓	Sensitivity (%) ↑	Precision (%) ↑	p-value (vs. Fine-tuned)
3D U-Net (MONAI)	65.0 ± 10.2	15.8 ± 8.5	6.2 ± 2.1	66.2 ± 10.5	68.1 ± 12.0	< 0.001
Swin UNETR	71.2 ± 8.6	9.5 ± 6.2	3.1 ± 1.5	80.4 ± 8.1	77.5 ± 10.2	< 0.001
nnU-Net v2	76.6 ± 7.5	7.3 ± 8.6	2.5 ± 2.3	77.2 ± 7.4	81.3 ± 8.5	< 0.01
nnU-NET (pretrained + fine-tuned)	83.44 ± 6.5	5.1 ± 3.6	2.0 ± 2.2	79.19 ± 8.2	89.60 ± 8.2	-

Lưu ý: Ký hiệu ↑ biểu thị giá trị càng cao càng tốt; ↓ biểu thị giá trị càng thấp càng tốt.

Trong số các kiến trúc được huấn luyện từ đầu, mô hình nnU-Net v2 (mô hình cơ sở - baseline) thể hiện độ ổn định cao nhất với chỉ số DSC đạt $76,6 \pm 7,5\%$ và HD95 đạt $7,3 \pm 8,6$ mm. Swin UNETR xếp vị trí tiếp theo khi đạt chỉ số DSC là $71,2 \pm 8,6\%$. Mặc dù chỉ số DSC tổng thể của Swin UNETR thấp hơn so với nnU-Net, nhưng kiến trúc này lại ghi nhận độ nhạy (sensitivity) cao nhất trong tất cả các mô hình, đạt mức $80,4 \pm 8,1\%$. Năng lực giảm thiểu các tổn thương bị bỏ sót (âm tính giả) của Swin UNETR có được là nhờ cơ chế tự chú ý (self-attention mechanism) đặc trưng của các mạng Transformer, giúp mô hình hóa các bối cảnh tổng thể (global contexts) một cách hiệu quả hơn. Ngược lại, mô hình 3D U-Net (MONAI) cho thấy hiệu năng kém nhất

trên bộ dữ liệu này, khi chỉ đạt chỉ số DSC là $65,0 \pm 10,2\%$ và HD95 lên tới $15,8 \pm 8,5$ mm.

Hình 3.23 trình bày các kết quả định tính mang tính đại diện từ hai ca bệnh thực tế, so sánh một cách trực quan kết quả đầu ra phân đoạn của các kiến trúc huấn luyện khác nhau. Qua quan sát trực quan, các mô hình nnU-Net cơ sở và Swin UNETR gặp nhiều khó khăn trong việc phác họa một cách chính xác các đường biên khối u phức tạp. Ngược lại, mô hình nnU-Net tinh chỉnh thể hiện độ chồng chấp tốt nhất đối với kết quả đồng thuận STAPLE (STAPLE consensus), đặc biệt là tại các vùng có mức hấp thu FDG không đồng nhất (heterogeneous). Nhìn chung, với độ chính xác không gian và độ chính xác phân loại (precision) vượt trội, mô hình nnU-Net tinh chỉnh đã được lựa chọn để sử dụng cho toàn bộ các phân tích chuyên sâu tiếp theo (downstream analyses) trong nghiên cứu này.



Hình 3.23. Kết quả định tính tiêu biểu với các kiến trúc huấn luyện khác nhau

Độ đồng thuận giữa các quan sát viên

Việc đánh giá một mô hình AI đặt trong mối tương quan với mức độ biến thiên giữa các quan sát viên người (inter-observer variability) đóng vai trò quyết định trong việc xác định tính khả thi của mô hình khi đưa vào ứng dụng lâm sàng. Trong nghiên cứu này, độ đồng thuận giữa ba bác sĩ chẩn đoán hình ảnh dao động từ mức trung bình đến tốt, với hệ số Dice bắt cặp (pairwise Dice coefficient) đạt $0,78 \pm 0,06$. Khi tiến hành so sánh độc lập với nhãn chú giải của từng bác sĩ (được trình bày ở [Bảng 3.10](#)), mô hình nnU-Net tinh chỉnh đã đạt được chỉ số Dice trung bình vô cùng ấn tượng là 0,818. Cụ thể, mô hình đạt giá trị DSC lần lượt là 0,815, 0,822 và 0,818 khi đánh giá đối chiếu với Người đánh giá 1 (Rater 1), Người đánh giá 2 (Rater 2) và Người đánh giá 3 (Rater 3); đồng thời đạt đỉnh với giá trị DSC là 0,834 khi so sánh với nhãn đồng thuận STAPLE. Đáng chú ý, hiệu năng phân đoạn của AI đã vượt qua biên độ sai số giữa các đánh giá viên người (human inter-rater error margin) trong 68% số ca được khảo sát. Những phát hiện này cung cấp bằng chứng thuyết phục cho thấy mô hình tinh chỉnh mang lại một mức độ nhất quán ít nhất là tương đương với các chuyên gia con

người, từ đó đáp ứng đầy đủ các yêu cầu về độ tin cậy để triển khai thực tế trên lâm sàng.

Bảng 3.10. Phân tích DSC về sự phù hợp giữa mô hình đề xuất và ý kiến của từng chuyên gia chẩn đoán hình ảnh.

Model	Rater 1	Rater 2	Rater 3	STAPLE Consensus
nnU-Net (pretrained + fine-tuned)	0.815	0.822	0.818	0.834

3.3.2. Kiểm định chéo dữ liệu

Hiệu năng trên các bộ dữ liệu quốc tế

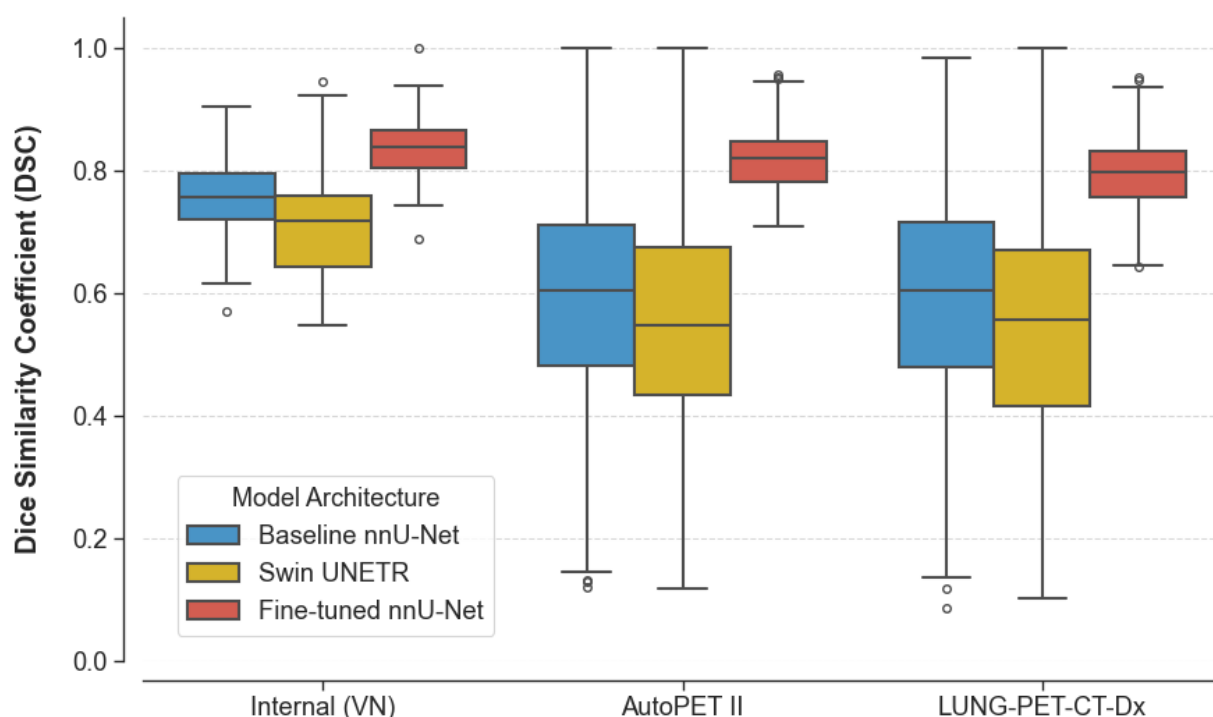
Nhằm đánh giá một cách nghiêm ngặt khả năng tổng quát hóa miền dữ liệu (domain generalization) và độ nhạy của mô hình đối với hiện tượng dịch chuyển miền (domain shift), các kiến trúc phân đoạn đã được áp dụng trực tiếp (dưới dạng zero-shot, không qua huấn luyện lại) trên hai bộ dữ liệu độc lập hoàn toàn từ bên ngoài: bộ dữ liệu AutoPET II (n = 200 ca thử nghiệm) và bộ dữ liệu LUNG-PET-CT-Dx (TCIA) (n = 50 ca thử nghiệm). Như được minh họa ở [Bảng 3.11](#), các mô hình được huấn luyện từ đầu trên bộ dữ liệu nội bộ đã bị suy giảm hiệu năng nghiêm trọng khi đối chiếu với dữ liệu bên ngoài. Các mô hình nnU-Net v2 cơ sở và Swin UNETR ghi nhận các mức sụt giảm chỉ số DSC trung bình khổng lồ, lần lượt lên tới -21,0% và -20,1%. Sự suy giảm hiệu năng rõ rệt này chủ yếu bắt nguồn từ những khác biệt mang tính hệ thống trong giao thức máy quét (scanner protocols), kỹ thuật chuẩn hóa chỉ số SUV, cũng như sự biến thiên về nhân khẩu học giữa hai quần thể bệnh nhân châu Á và phương Tây. Ngược lại, mô hình nnU-Net tinh chỉnh lại cho thấy khả năng chống chịu (resilience) vượt trội khi duy trì chỉ số DSC ở mức cao, đạt 0,811 trên tập AutoPET II và 0,798 trên tập LUNG-PET-CT-Dx. Với mức sụt giảm hiệu năng trung bình cực kỳ thấp — chỉ -3,5% — khung làm việc dựa trên chiến lược tinh chỉnh đã chứng minh được năng lực tổng quát hóa vượt xa các kiến trúc chỉ được huấn luyện cục bộ từ đầu.

Bảng 3.11. Hiệu suất tổng quát hóa không cần huấn luyện của các mô hình phân đoạn trên các bộ dữ liệu quốc tế độc lập (AutoPET II và LUNG-PET-CT-Dx).

Model	Internal (VN)	AutoPET II	LUNG-PET-CT-Dx	Average Drop
nnU-Net v2 (baseline)	0.766	0.612	0.598	-21.0%
Swin UNETR	0.712	0.577	0.560	-20.1%
nnU-Net (pretrained + fine-tuned)	0.834	0.811	0.798	-3.5%

Sự vượt trội về khả năng tổng quát hóa của mô hình tinh chỉnh được minh chứng rõ nét hơn thông qua phân phối thông kê ở [Hình 3.24](#). Biểu đồ chỉ ra rằng khi đối mặt

với hiện tượng dịch chuyển miền trên các bộ dữ liệu độc lập bên ngoài, các kiến trúc cơ sở (nnU-Net baseline và Swin UNETR) không chỉ bị sụt giảm nghiêm trọng về giá trị trung vị (median values) mà còn bộc lộ mức độ phân tán dữ liệu khổng lồ. Khoảng tứ phân vị (IQR) của các mô hình này bị giãn rộng đáng kể, đi kèm với hàng loạt điểm dữ liệu dị biệt (outliers) có hiệu năng lao dốc xuống các ngưỡng cực thấp ($< 0,4$). Ngược lại, mô hình nnU-Net tinh chỉnh vẫn duy trì được độ ổn định đáng kinh ngạc. Biên độ phân tán của mô hình đề xuất trên cả hai bộ dữ liệu AutoPET II và LUNG-PET-CT-Dx đều rất hẹp và tương đương với mức phân phối ghi nhận trên bộ dữ liệu nội bộ. Điều này minh chứng rằng hệ thống AI dựa trên chiến lược tinh chỉnh không chỉ bảo toàn được độ chính xác tổng thể, mà còn giảm thiểu tối đa các lỗi thất bại thảm hại (catastrophic failures) gây ra bởi sự khác biệt trong giao thức hình ảnh và các chủng loại thiết bị khác nhau.



Hình 3.24. Phân bố DSC trên các tập dữ liệu nội bộ và bên ngoài. Biểu đồ hộp minh họa giá trị trung vị, khoảng tứ phân vị (IQR) và các vòng tròn bao cho các mô hình được đánh giá chuẩn, làm nổi bật tính ổn định vượt trội của khung được tinh chỉnh dưới sự thay đổi miền dữ liệu.

Nghiên cứu loại trừ thành phần: hiệu quả của quy trình tinh chỉnh từ mô hình tiền huấn luyện

Nhằm định lượng một cách hệ thống các lợi ích thực nghiệm của việc khởi tạo mạng thần kinh bằng các trọng số tiền huấn luyện (pretrained weights) từ dự án AutoPET II so với việc khởi tạo ngẫu nhiên (random initialization), nghiên cứu sinh đã tiến hành một nghiên cứu loại trừ thành phần (ablation study) có kiểm soát, sử dụng phương pháp kiểm định chéo 5 tập trên bộ dữ liệu nội bộ. [Bảng 3.12](#) cho thấy những lợi ích sâu sắc mà chiến lược học chuyển giao (transfer learning) mang lại trên cả hai

phương diện: hiệu năng mô hình và thời gian tính toán. Việc huấn luyện mô hình nnU-Net cơ sở từ đầu (train from scratch) cho chỉ số DSC đạt 0,766 và khoảng cách đường biên HD95 đạt 7,30 mm, đòi hỏi thời gian xử lý xấp xỉ 16 giờ cho mỗi tập kiểm định. Trong khi đó, quy trình tinh chỉnh (fine-tuning) đã thúc đẩy đáng kể độ chính xác phân đoạn, nâng chỉ số DSC lên mức 0,834, đồng thời rút ngắn rõ rệt chỉ số HD95 xuống còn 5,10 mm.

Hơn thế nữa, quy trình tinh chỉnh này đã giúp cắt giảm tới 37,5% tổng thời gian huấn luyện (rút ngắn xuống còn 10 giờ cho mỗi tập kiểm định) và đạt được sự hội tụ (convergence) chỉ sau 500 kỷ nguyên bổ sung. Các kết quả từ phân tích loại trừ thành phần này một lần nữa khẳng định rằng: việc tận dụng các đặc trưng tổng thể (global features) từ các bộ dữ liệu quy mô lớn là một chiến lược tối ưu, đặc biệt phát huy hiệu quả cao khi được kết hợp với một quy trình tinh chỉnh nghiêm ngặt trên miền dữ liệu đích (target data domain).

Bảng 3.12. Kết quả nghiên cứu loại bỏ đánh giá tác động của học chuyển giao lên độ chính xác phân đoạn và hiệu quả huấn luyện.

Configuration	DSC	HD95 (mm)	Training time (hours/fold)
Baseline nnU-Net	0.766	7.30	16
Fine-tuned nnU-Net	0.834	5.10	10

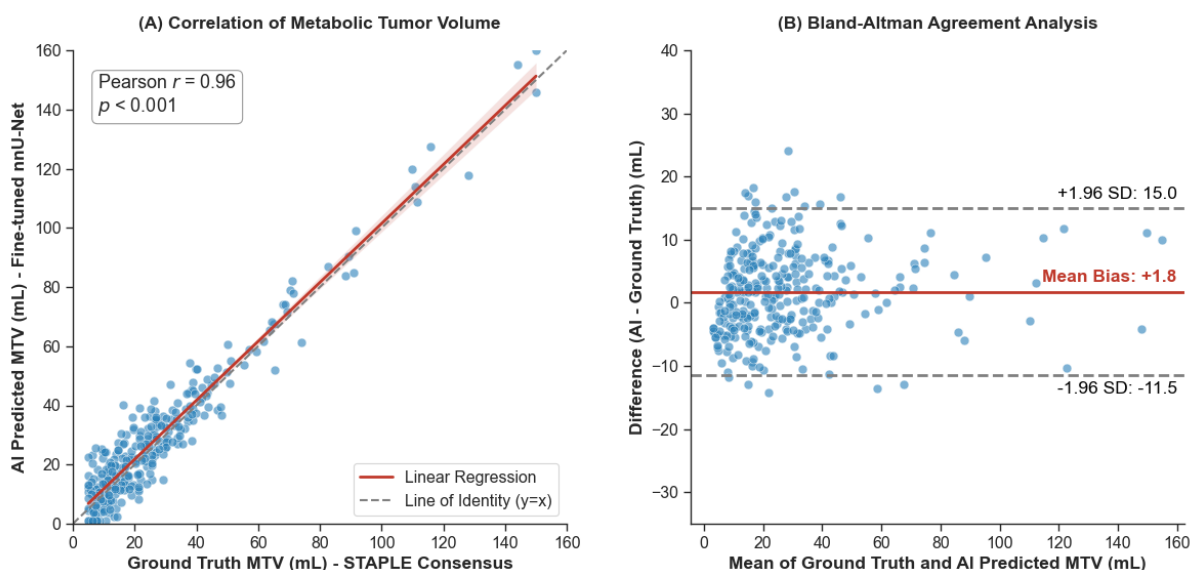
Kiểm định lâm sàng

Vượt ra ngoài các chỉ số đánh giá độ chồng chập không gian (như DSC hay HD95), giá trị cốt lõi của một mô hình AI trong ngành xạ trị ung thư (radiation oncology) nằm ở độ chính xác của nó khi trích xuất các thông số sinh học định lượng, đặc biệt là chỉ số tổng thể tích khối u chuyển hóa (MTV). Để đánh giá năng lực này, nghiên cứu sinh đã tiến hành so sánh đối chiếu thể tích khối u do mô hình dự báo tự động với thể tích thu được từ các nhãn đồng thuận STAPLE của các chuyên gia.

[Hình 3.25A](#) trình bày các kết quả phân tích thống kê cho thấy một mối tương quan tuyến tính cực kỳ mạnh mẽ giữa kết quả dự báo của AI và kết quả đánh giá thủ công của bác sĩ, đạt hệ số tương quan Pearson lên tới $r = 0,96$ ($p < 0,001$). Bên cạnh đó, biểu đồ phân tích Bland-Altman ([Hình 3.25B](#)) minh chứng cho mức độ đồng thuận lâm sàng ở mức cao, ghi nhận độ sai lệch trung bình (mean bias) chỉ ở mức +1,8 mL. Giới hạn đồng thuận 95% (95% limits of agreement) nằm trong một khoảng hẹp từ -11,5 mL đến +15,0 mL, đây là mức sai số hoàn toàn chấp nhận được trong phạm vi dung sai cho phép của quy trình lập kế hoạch xạ trị thực tế.

Đặc biệt, quy trình tinh chỉnh trên nguồn dữ liệu nội bộ đã giải quyết thành công một trong những điểm hạn chế lớn nhất của các mô hình học sâu trên ảnh PET: hiện tượng dương tính giả tại các vùng hấp thu FDG sinh lý. So với mô hình nnU-Net cơ sở, tỷ lệ nhận diện sai tại các vùng cơ quan như mỡ nâu, cơ tim hoặc não đã giảm đáng

kể từ 14% xuống còn 6% sau khi áp dụng chiến lược tinh chỉnh. Sự cải tiến này chứng minh rằng mô hình không chỉ học được đặc điểm hình thái của khối u, mà còn phân biệt hiệu quả giữa tín hiệu bệnh lý và các hoạt động chuyển hóa sinh lý đặc trưng của từng bệnh nhân.

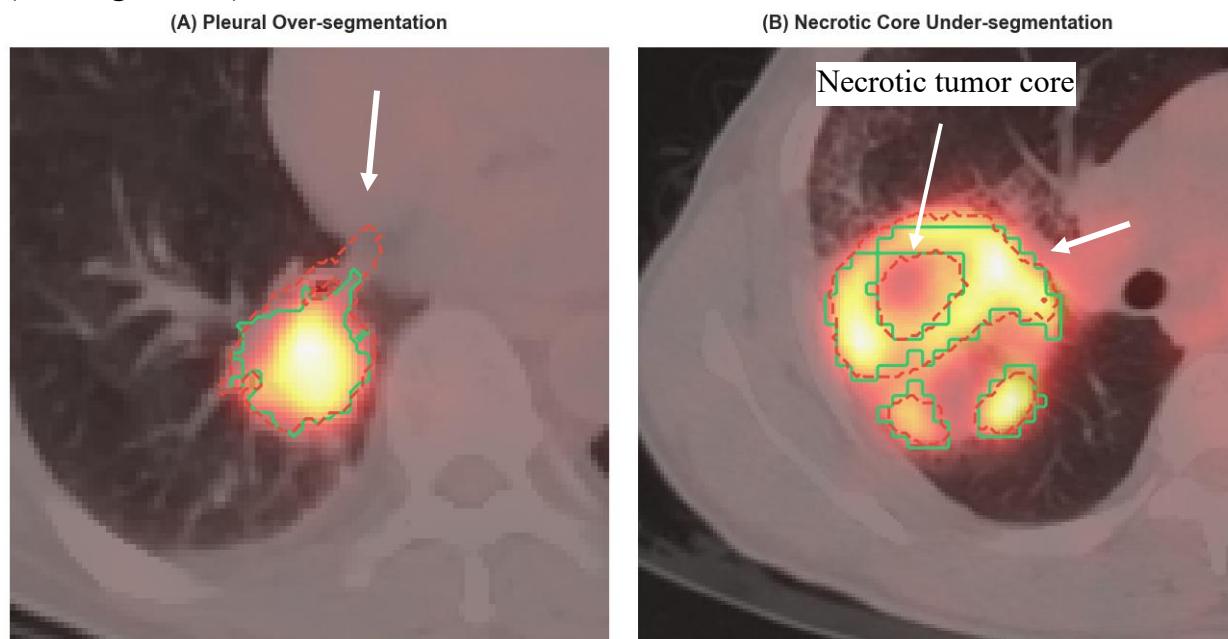


Hình 3.25: Phân tích sự tương đồng lâm sàng của MTV được trích xuất bởi mô hình nnU-Net tinh chỉnh so với sự đồng thuận của chuyên gia STAPLE.

Hình 3.26 minh họa những điểm khác biệt giữa nhãn đồng thuận STAPLE (đường nét liền màu xanh lá) và mô hình nnU-Net tinh chỉnh (đường nét đứt màu đỏ), trong đó các mũi tên màu trắng chỉ ra các vùng bất đồng nhất đáng chú ý. Dù sở hữu độ bền vững (robustness) tổng thể cao, mô hình nnU-Net tinh chỉnh đôi khi vẫn bộc lộ hiện tượng phân đoạn bất đồng nhất tại các vùng ranh giới phức tạp. Trong các trường hợp có vị trí giải phẫu cận kề nhau, chẳng hạn như vùng tiếp giáp giữa khối u và màng phổi, mô hình có xu hướng phân đoạn quá mức (over-segmentation) do ảnh hưởng của hiệu ứng thể tích từng phần (partial volume effects) làm mờ các đường biên (Hình 3.26A). Ngược lại, Hình 3.26B trình bày một trường hợp phân đoạn thiếu hụt (under-segmentation) điển hình đối với khối u có chứa lõi hoại tử không hấp thu FDG (non-FDG-avid necrotic core). Trong những kịch bản như vậy, kết quả dự báo của mô hình chỉ khu trú duy nhất tại vùng viền tăng chuyển hóa (hypermetabolic rim) mà bỏ sót, không bao trọn được toàn bộ thể tích hoại tử bên trong vốn đã được phác họa bởi sự đồng thuận của các chuyên gia. Những ca thất bại cụ thể này đã làm nổi bật các thách thức vẫn đang tồn tại trong việc phân đoạn các khối u có tính chất đa dạng

(heterogeneous)

cao.



Hình 3.26. Phân tích định tính sự không nhất quán phân đoạn ở các vùng ranh giới giải phẫu phức tạp: (A) phân đoạn quá mức màng phổi; (B) phân đoạn thiếu lõi khối u hoại tử không hấp thụ FDG

Bàn luận

Nghiên cứu này đã phát triển và kiểm định một hệ thống phân đoạn tự động khối u ung thư phổi không tế bào nhỏ (NSCLC) trên hình ảnh đa phương thức ^{18}F -FDG PET/CT dựa trên kiến trúc nnU-Net v2. Hệ thống này được tối ưu hóa chuyên biệt cho quần thể bệnh nhân Việt Nam thông qua một bộ dữ liệu gồm 416 ca bệnh (100% đã được xác chuẩn bằng kết quả mô bệnh học, với nhãn đồng thuận STAPLE từ ba chuyên gia y học hạt nhân). Các kết quả định lượng tại Mục 3 đã cung cấp minh chứng rõ ràng về hiệu năng vượt trội của kiến trúc nnU-Net tinh chỉnh so với các kiến trúc được huấn luyện từ đầu, đồng thời khẳng định khả năng tổng quát hóa mạnh mẽ của mô hình trên nguồn dữ liệu độc lập bên ngoài.

Trong quy trình đánh giá nội bộ bằng phương pháp kiểm định chéo 5 tập, mô hình nnU-Net tinh chỉnh đã đạt được chỉ số DSC cao nhất ($83,44 \pm 6,5\%$), chỉ số HD95 thấp nhất ($5,1 \pm 3,6$ mm) và độ chính xác phân loại đứng đầu ($89,60 \pm 8,2\%$). Khi đặt trong tương quan so sánh với mô hình nnU-Net v2 cơ sở (DSC đạt $76,6 \pm 7,5\%$; HD95 đạt $7,3 \pm 8,6$ mm), kiến trúc nnU-Net tinh chỉnh đã cải thiện đáng kể cả về độ chính xác không gian lẫn độ chính xác phân loại, đồng thời làm giảm sai số bề mặt (chỉ số ASSD giảm từ $2,5 \pm 2,3$ mm xuống còn $2,0 \pm 2,2$ mm). Kết quả này hoàn toàn đồng nhất với các khuyến nghị của Isensee và các cộng sự [68] về những lợi ích của việc tinh chỉnh nnU-Net trên miền dữ liệu đích, cũng như phù hợp với các nghiên cứu gần đây tập trung vào kỹ thuật thích ứng miền (domain adaptation) trong bài toán phân đoạn ảnh PET/CT.

Mô hình nnU-Net tinh chỉnh cũng đã chứng minh được tính nhất quán lâm sàng vượt trội so với mức độ biến thiên giữa các quan sát viên người (human inter-observer variability). Trong khi hệ số Dice bắt cặp giữa ba bác sĩ chỉ đạt $0,78 \pm 0,06$, mô hình tinh chỉnh đã đạt được chỉ số DSC trung bình là 0,818 khi đối chiếu độc lập với từng người đánh giá, và đạt đỉnh ở mức 0,834 so với nhãn đồng thuận STAPLE. Hiệu năng của AI đã vượt qua biên độ sai số của các chuyên gia trong 68% số ca bệnh, khẳng định một cách chắc chắn tính khả thi về mặt lâm sàng của mô hình.

Khả năng tổng quát hóa (generalization capability) của mô hình đã được chứng minh thông qua quy trình đánh giá chéo trên hai bộ dữ liệu độc lập bên ngoài (AutoPET II và LUNG-PET-CT-Dx). Các mô hình được huấn luyện từ đầu (nnU-Net v2 cơ sở và Swin UNETR) đều bị sụt giảm nghiêm trọng về chỉ số DSC trung bình, lần lượt là -21,0% và -20,1%; trong khi đó, mô hình nnU-Net tinh chỉnh chỉ giảm -3,5% (duy trì chỉ số DSC ở mức 0,811 trên tập AutoPET II và 0,798 trên tập LUNG-PET-CT-Dx). Biểu đồ hộp (boxplot) đã minh họa rõ nét việc mô hình tinh chỉnh sở hữu biên độ phân tán hẹp và ít điểm dữ liệu dị biệt (outliers) hơn, chứng minh độ bền vững (robustness) vượt trội của nó trước hiện tượng dịch chuyển miền (domain shift) gây ra bởi các khác biệt về nhân khẩu học, phân phối chỉ số SUV và giao thức máy quét.

Phân tích loại trừ thành phần (ablation analysis) đã khẳng định rõ ràng giá trị của chiến lược học chuyển giao (transfer learning): việc khởi tạo mạng bằng các trọng số tiền huấn luyện kết hợp với quy trình tinh chỉnh không chỉ nâng chỉ số DSC từ 0,766 lên 0,834, mà còn rút ngắn khoảng cách đường biên HD95 từ 7,30 mm xuống còn 5,10 mm, đồng thời tiết kiệm được 37,5% thời gian huấn luyện (giảm từ 16 giờ xuống còn 10 giờ cho mỗi tập kiểm định). Kết quả này một lần nữa củng cố vai trò sống còn của học chuyển giao trong việc cắt giảm chi phí tính toán và đẩy nhanh tốc độ hội tụ (convergence rates) mà vẫn bảo toàn được hiệu năng ở mức cao.

Về mặt giá trị lâm sàng, mô hình tinh chỉnh đã đạt được mối tương quan tuyến tính rất tốt với chỉ số thể tích khối u chuyển hóa (MTV) đo đặc thủ công bởi các chuyên gia (hệ số Pearson $r = 0,96$; $p < 0,001$), cùng độ sai lệch trung bình (mean bias) trong phân tích Bland-Altman chỉ ở mức +1,8 mL (giới hạn đồng thuận 95% LoA: từ -11,5 đến 15,0 mL). Đồng thời, tỷ lệ dương tính giả tại các vùng hấp thu sinh lý (ví dụ: mỡ nâu, cơ tim, não) đã giảm từ 14% xuống còn 6%, minh chứng rằng mô hình không chỉ học được các đặc trưng hình thái của khối u, mà còn phân biệt hiệu quả giữa tín hiệu bệnh lý và các hoạt động chuyển hóa sinh lý bình thường.

Mặc dù đạt được độ chính xác tổng thể đầy ấn tượng, kết quả phân tích định tính chuyên sâu (Hình 3.26) đã chỉ ra một số hạn chế nhất định của mô hình tinh chỉnh tại các vùng ranh giới giải phẫu phức tạp. Diễn hình như quan sát ở Hình 3.26A, mô hình đôi khi còn gặp khó khăn trong việc phác họa ranh giới chính xác giữa khối u và các cấu trúc lân cận như màng phổi, dẫn đến hiện tượng phân đoạn quá mức (over-segmentation). Tình trạng này có khả năng trở nên nghiêm trọng hơn do hiệu ứng thể tích từng phần (partial volume effect) vốn có của kỹ thuật ghi hình ảnh PET, làm nhòe đi các tín hiệu ở vùng biên. Ngược lại, đối với các khối u lớn có hiện tượng hoại tử

trung tâm (Hình 3.26B), mạng thần kinh lại bộc lộ xu hướng phân đoạn thiếu hụt (under-segmentation). Do vùng hoại tử không có đặc tính hấp thu FDG (lacks FDG avidity), mô hình chủ yếu dựa vào dải tín hiệu PET cao để phác họa đường biên, dẫn đến việc bỏ sót lõi khối u bất chấp kênh dữ liệu CT đã cung cấp thông tin về mật độ mô. Những điểm bất đồng nhất cục bộ tại vùng biên này chính là nguyên nhân lý giải cho biên độ sai số còn tồn dư trong chỉ số Dice ($\sim 16,5\%$) và khoảng cách HD95 (5,1 mm). Những cải tiến thuật toán trong tương lai có thể tập trung vào việc tích hợp các hàm mất mát nhận biết đường biên (boundary-aware loss functions) hoặc tăng cường các cơ chế chú ý (attention mechanisms) cho luồng trích xuất đặc trưng của ảnh CT, từ đó thúc đẩy AI nhận diện hình thái một cách toàn diện thay vì phụ thuộc quá mức vào mức độ chuyển hóa FDG.

KẾT LUẬN

Luận án đã thực hiện được các mục tiêu và nội dung nghiên cứu đề ra, các kết quả nghiên cứu chính thể hiện đóng góp mới bao gồm:

- 1) Luận án đã tiến hành thu thập được các dữ liệu từ 03 bệnh viện: Bệnh viện Ung Bướu Hà Nội, Bệnh viện K và Bệnh viện Ung Bướu TP. Hồ Chí Minh với 416 tập dữ liệu ảnh PET/CT lồng ngực của 416 cá nhân khác nhau. Bộ dữ liệu bao gồm: Các thông tin liên quan như: tuổi, giới tính, cân nặng, liều ^{18}F -FDG được tiêm, giai đoạn bệnh, có bệnh hay không có bệnh ung thư phổi không tế bào nhỏ.; Thông tin chẩn đoán của từng bệnh nhân từ hình ảnh PET/CT. Bộ dữ liệu bao gồm ảnh PET/CT của 300 bệnh nhân mắc ung thư phổi không tế bào nhỏ và 116 bệnh nhân không mắc ung thư phổi không tế bào nhỏ, được gán nhãn và phân loại thành các thư mục riêng lẻ. Đối với mỗi bệnh nhân được gán nhãn ung thư không tế bào nhỏ sẽ có 04 dữ liệu ảnh PET/CT bao gồm: ảnh gốc, và 03 ảnh do 03 bác sĩ khác nhau khoanh vùng khối u. Trong đó, 138 bệnh nhân điều trị tại Bệnh viện Ung Bướu Hà Nội, 138 bệnh nhân điều trị tại Bệnh viện K và 140 bệnh nhân điều trị tại Bệnh viện Ung Bướu TP. Hồ Chí Minh. Trong bộ dữ liệu này, dữ liệu ảnh PET/CT của một bệnh nhân có khoảng 200-300 lát cắt CT và PET.
- 2) Luận án đã xây dựng thành công chương trình đọc dữ liệu phân đoạn thủ công được thực hiện bởi bác sĩ, tiến hành xử lý dữ liệu và xác định biên dạng, đường bao của các khối u, từ đó trích xuất và lưu lại thông tin của các khối u phục vụ cho việc kiểm tra và để huấn luyện các mô hình học máy. Thông tin trích xuất ra bao gồm: thông tin tổng quát về các khối u có trong cơ sở dữ liệu của các bệnh nhân và thông tin chi tiết của khối u trên từng lát cắt của bộ dữ liệu. Với kết quả này, nghiên cứu sinh sẽ tiếp tục hoàn thiện chương trình, giao diện đồ họa để xây dựng phần mềm đọc dữ liệu hình ảnh chẩn đoán PET/CT, trích xuất thông tin phân đoạn thủ công của bác sĩ, từ đó có thể xây dựng và chuẩn hóa dữ liệu hình ảnh PET/CT phục vụ cho việc huấn luyện cho các mô hình, chương trình chẩn đoán ung thư phổi bằng cách ứng dụng trí tuệ nhân tạo và học máy.
- 3) Luận án đã phát triển và kiểm định thành công hệ thống phân đoạn tự động khối u ung thư phổi không tế bào nhỏ (NSCLC) trên hình ảnh ^{18}F -FDG PET/CT dựa trên khung làm việc nnU-Net v2 tinh chỉnh. Hệ thống này được tối ưu hóa cho bệnh nhân Việt Nam dựa trên một bộ dữ liệu mạnh mẽ gồm 416 ca bệnh. Các kết quả cho thấy mô hình nnU-Net tinh chỉnh đạt hiệu năng vượt trội so với các kiến trúc được huấn luyện từ đầu, bao gồm cả mô hình nnU-Net cơ sở và Swin UNETR. Mô hình nnU-Net tinh chỉnh cũng thể hiện tính nhất quán lâm sàng cao (chỉ số DSC đạt 0,834 so với nhãn đồng thuận STAPLE), mối tương quan cực kỳ mạnh mẽ về chỉ số MTV ($r = 0,96$), và giảm đáng kể tỷ lệ dương tính giả tại các vùng hấp thu sinh lý (từ 14% xuống còn 6%). Đáng chú ý, mô hình vẫn duy trì được hiệu năng ổn định trên các nguồn

dữ liệu độc lập bên ngoài (DSC đạt 0,811 trên tập AutoPET II và 0,798 trên tập LUNG-PET-CT-Dx) với mức sụt giảm trung bình chỉ -3,5%, hoàn toàn trái ngược với mức giảm lớn -21,0% của mô hình cơ sở. Kết quả này khẳng định khả năng tổng quát hóa vượt trội có được nhờ chiến lược học chuyển giao. Quy trình tinh chỉnh cũng giúp tiết kiệm đến 37,5% thời gian huấn luyện, mang lại những lợi ích thực tiễn lớn về mặt tối ưu hóa tài nguyên tính toán. Những phát hiện này minh chứng rằng việc bản địa hóa (localizing) các mô hình học sâu tiên tiến nhất hiện nay (state-of-the-art) trên một bộ dữ liệu bản địa chất lượng cao là giải pháp cực kỳ hiệu quả để vượt qua rào cản dịch chuyển miền và đưa AI chuyển giao vào ứng dụng lâm sàng thực tế. Hệ thống nnU-Net tinh chỉnh không chỉ đạt đến độ chính xác tương đương cấp độ chuyên gia, mà còn cung cấp một công cụ tự động hóa đáng tin cậy, đóng góp vào việc chuẩn hóa quy trình phân chia giai đoạn ung thư phổi và lập kế hoạch xạ trị tại Việt Nam cũng như các quốc gia đang phát triển khác.

Hướng phát triển tiếp theo của đề tài

Kết quả nghiên cứu của luận án bước đầu mở ra hướng đi cho việc tích hợp AI vào thực hành y học chính xác, hướng tới mục tiêu nâng cao tính công bằng và khả năng tiếp cận trong chẩn đoán hình ảnh PET/CT. Để định hướng phát triển tiếp theo cho một ứng dụng học máy tự động phân đoạn ảnh PET/CT, chúng ta cần giải quyết các bài toán cốt lõi về độ chính xác kỹ thuật, khả năng ứng dụng lâm sàng và tính đa nhiệm, cụ thể là:

- Dựa trên các đặc tính vật lý ảnh hưởng trực tiếp đến độ phân giải không gian của PET (như sai số do quãng chạy hiệu dụng của positron hay góc lệch không đồng trục), hướng phát triển tiếp theo có thể đưa AI tham gia sâu hơn vào giai đoạn tiền xử lý như sử dụng mạng nơ-ron để học các mô hình nhiễu vật lý, từ đó "khôi phục" và làm sắc nét các đường biên khối u bị mờ do hiệu ứng thể tích bán phần (partial volume effect) hoặc do chuyển động hô hấp của phổi.
- Phát triển các thuật toán học máy có khả năng tự động phân đoạn và đồng bộ theo chu kỳ thở của bệnh nhân mà không cần đến các thiết bị phần cứng gating phức tạp.
- Áp dụng thuật toán Học tự giám sát (Self-supervised learning) trên các tập dữ liệu lớn chưa được dán nhãn để mô hình tự học các đặc trưng hình học của phổi trước, sau đó chỉ cần một lượng nhỏ dữ liệu "chuẩn" (Few-shot learning) để tinh chỉnh (fine-tune) cho tác vụ phân đoạn tổn thương.

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

Danh sách các bài báo quốc tế

1. **Ho Quang Tuan**, Tran Thuy Duong, Bui Ngoc Ha, Nguyen Huu Quyet, Le Van Tinh, Chu Van Tuynh, Vo Khac Nam, Le Thi Quynh Dao, Chu Van Luong, Dang Thi My Linh, Duong Thi Nhung, Nguyen Dang Nguyen, Vo Quynh Trang; *Development of a Vietnamese PET/CT Dataset for Machine Learning-Based Analysis of Non-Small Cell Lung Cancer Images*; Atom Indonesia, Vol. 51 No.2 (2025), 137-143. (<https://atomindonesia.brin.go.id/index.php/aij/article/view/1645>)
2. Bui Tien Hung, Tran Thuy Duong, Bui Ngoc Ha, **Ho Quang Tuan**; *A Monte Carlo simulation of Gamma ART-6000 devices to investigate beam characteristics in a homogeneous environment using Geant4 simulation toolkits*; Radiation Physics and Chemistry - 2025, Vol. 229 (<https://doi.org/10.1016/j.radphyschem.2025.112542>)
3. Ioannis Delakis, Zoe Brady, Md Nahid Hossain, Rath Beauta, Sann Chanthoubopha, Chhoeurt Koeurn, Yun Seang Yi, Rajesh Kumar, Lukmanda Evan Lubis, Supriyanto Ardjo Pawiro, Indah Lestariningsih, Hideyuki Mizuno, Kosuke Matsubara, Keiichi Akahane, Siti Normasitah Masduki, Siti Norsyafiqah Mohd Mustafa, Kanchan Prasad Adhikari, Darin O’Keeffe, Atif Masood, Luwis Gabriel, Sanja Ipu, Melanie Dela Fuente Marquez, Ramon Carlo L. Cruzpero, Soma Somanesan, Anchali Krisanachinda, **Tuan Quang Ho**, Olivera Ciraj-Bjelac; *Status of the medical physics profession in diagnostic radiology and image-guided interventional procedures in the Asia–Pacific region: initial findings from an IAEA project*; Physical and Engineering Sciences in Medicine Volume 48, pages 867–880, (2025) (<https://doi.org/10.1007/s13246-025-01554-w>)
4. Quang Tuan Ho, Ngoc Ha Bui, Thuy Duong Tran, Quang Huy Khuat, Ngoc Toan Tran, Xuan Chung Le, Huu Quyet Nguyen, Tat Thang Nguyen, Van Thai Nguyen, Dinh Thuy Mai, Quang Duy To, Dinh Chau Nguyen, Nguyen Huong Giang Trinh, Van Chinh Cao, Tien Hung Bui, Thu Trang Vu, Khac Nam Vo and Hai Quan Ho; *Benchmarking Deep Learning for NSCLC PET/CT Segmentation on a Histologically Confirmed Vietnamese Dataset: Validation and Generalization*; Journal of Imaging Volume 12(8), 364 (2026), (<https://doi.org/10.3390/jimaging12080364>)

Danh sách các bài báo trong nước

1. **Hồ Quang Tuấn**, Đặng Thị Mỹ Linh, Dương Thị Nhung, Nguyễn Hữu Quyết, Nguyễn Ngọc Quỳnh, Bùi Tiến Hưng, Lê Xuân Chung, Trần Thuỳ Dương, Bùi Ngọc Hà, Nguyễn Văn Thái, Phạm Lâm Sơn, Lê Văn Tình, Chu Văn Lương, Chu Văn Tuynh, Bùi Xuân Cường, Hoàng Văn Toán, Võ Khắc Nam, Nguyễn Trung Hiếu, Đặng Thị Minh Tâm, Lê Thị Quỳnh Dao, Đặng Minh Phước; *Nghiên cứu thuật toán định vị và phân đoạn khối u phổi từ các thể tích lồng ngực PET/CT dựa trên phân tích đặc điểm hình ảnh*, Tạp chí Y học Việt Nam, 2023, tập 533 số Đặc biệt.
2. Bui Ngoc Ha, Nguyen Trung Kien, Tran Thi Thuy Duong, Tran Ngoc Toan, **Ho Quang Tuan**, Pham Cam Phuong, Pham Van Thai, Vu Thi Trang *Development of Brain MRI Image Segmentation program using UNET++ network for radiosurgery planning*; Journal of Nuclear Science and Technology, 2023, Vol. 13 No. 1.
(<https://doi.org/10.53747/nst.v13i1.398>)

TÀI LIỆU THAM KHẢO

- [1] Ioannis Delakis, Zoe Brady, Ho Quang Tuan, et al., "Status of the medical physics profession in diagnostic radiology and image-guided interventional procedures in the Asia-Pacific," *Physical and Engineering Sciences in Medicine*, vol. 48, pp. 867-880, 2025.
- [2] George Cris, Nastasia Sanda Moldovean-Cioroianu, Diana-Gabriela Timaru, Călin Căinap and Vasile Chis, "Radiopharmaceuticals for PET and SPECT Imaging: A Literature Review over the Last Decade," *International Journal of Molecular Sciences*, vol. 23, no. 5023, 2022.
- [3] JERROLD T. BUSHBERG, J. ANTHONY SEIBERT, EDWIN M. LEIDHOLDT JR, JOHN M. BOONE, The Essential Physics of Medical Imaging, Third Edition, Philadelphia, Mỹ: LIPPINCOTT WILLIAMS & WILKINS, 2012.
- [4] B. v. 103, "DCPX trong ghi hình PX PET," 09 October 2020. [Online]. Available: <https://benhvien103.vn/dcp-x-trong-ghi-hinh-px-pet/>.
- [5] Hossein Jadvar, Abass Alavi, and Sanjiv S. Gambhir, "18F-FDG Uptake in Lung, Breast, and Colon Cancers: Molecular Biology Correlates and Disease Characterization," *J Nucl Med*, vol. 50, no. 11, p. 1820–1827, 2009.
- [6] Khushica Purbhoo, Mboyoy Di-Tamba Vangu, "Normal Variants and Pitfalls of 18F-FDG PET/CT Imaging in Pediatric Oncology," *Front Nucl Med.*, vol. 2, no. 825891, 2022.
- [7] Ha, L.N., Phuong, N.T. & Son, M.H, "The impact of qualitative 18F-FDG PET/CT in predicting clinical outcomes of post-surgical differentiated thyroid cancer patients with elevated thyroglobulin and negative radioiodine whole-body scan," *BMC Surg*, vol. 24, no. 377, 2024.
- [8] Ferdinando Corica, Maria Silvia De Feo, Maria Lina Stazza, Maria Rondini , Andrea Marongiu, Viviana Frantellizzi, , Susanna Nuvoli, Alessio Farcomeni , Giuseppe De Vincentis, and Angela Spanu, "Qualitative and Semiquantitative Parameters of 18F-FDG-PET/CT as Predictors of Malignancy in Patients with Solitary Pulmonary Nodule," *cancers*, vol. 15, no. 1000, 2023.
- [9] M E J Callister, D R Baldwin, A R Akram, S Barnard, P Cane, J Draffan et al, "British Thoracic Society Pulmonary Nodule Guideline Development Group;

British Thoracic Society Standards of Care Committee," *Thorax*, vol. 70, pp. ii1-ii54, 2015.

- [10] Katherine L Ordidge, Nemi Gandy, Mubarik A Arshad, Kathryn Wallitt, Neil Soneji, Sameer Khan, Tara D Barwick, "Interobserver agreement of the visual Herder scale for the assessment of solitary pulmonary nodules on 18F Fluorodeoxyglucose PET/computed tomography," *Nucl Med Commun*, vol. 41, no. 3, pp. 235-240, 2020.
- [11] Stephen J. Swensen, MD; Marc D. Silverstein, MD; Duane M. Ilstrup, MS; Cathy D. Schleck; Eric S. Edell, MD, "The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules," *Arch Intern Med*, vol. 157, pp. 849-855, 1997.
- [12] Gerarda J. Herder, MD; Harm van Tinteren, MSc; Richard P. Golding, MD; Piet J. Kostense, PhD; Emile F. Comans, PhD; Egbert F. Smit, PhD; Otto S. Hoekstra, PhD, "Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography," *Chest*, vol. 128, no. 4, pp. 2490-6, 2005.
- [13] John W. Keyes, Jr., "SUV: Standard Uptake or Silly Useless Value," *J Nucl Med*, vol. 36, pp. 1836-1839, 1995.
- [14] Joseph A Thie, "Understanding the standardized uptake value, its methods, and implications for usage," *J Nucl Med*, vol. 45, no. 9, pp. 1431-4, 2004.
- [15] L. Sokoloff, M. Reivich, C. Kennedy, M. H. Des Rosiers, C. S. Patlak, K. D. Pettigrew, O. Sakurada, M. Shinohara, "The [14C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat," *Journal of Neurochemistry*, vol. 28, pp. 897-916, 1977.
- [16] Steven M. Larson, Yusuf Erdi, Timothy Akhurst, Madhu Mazumdar, Homer A. Macapinlac, Ronald D. Finn, Cecille Casilla, Melissa Fazzari, Neil Srivastava, Henry W.D. Yeung, John L. Humm, Jose Guillem, Robert Downey, Martin Karpeh, Alfred E. Cohen, Robert Gi, "Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis," *Clin Positron Imaging*, vol. 2, no. 3, pp. 159-171, 1999.
- [17] Johannes B. Roedl, Rivka R. Colen, Nagaraj S. Holalkere, Alan J. Fischman, Noah C. Choi, Michael A. Blake, "Adenocarcinomas of the esophagus: response to chemoradiotherapy is associated with decrease of metabolic

tumor volume as measured on PET-CT—comparison to histopathologic and clinical response evaluation," *Radiotherapy and Oncology*, vol. 89, no. 3, pp. 278-286, 2008.

- [18] Y E Erdi, H Macapinlac, K E Rosenzweig, J L Humm, S M Larson, A K Erdi, E D Yorke, "Use of PET to monitor the response of lung cancer to radiation treatment," *Eur J Nucl Med*, vol. 27, no. 7, pp. 861-6, 2000.
- [19] Patrick Veit-Haibach, Niklaus G Schaefer, Hans C Steinert, Jan D Soyka, Burkhardt Seifert, Rolf A Stahel, "Combined FDG-PET/CT in response evaluation of malignant pleural mesothelioma," *Lung Cancer*, vol. 67, no. 3, pp. 311-7, 2010.
- [20] Jose G Guillem, Harvey G Moore, Timothy Akhurst, David S Klimstra, Leyo Ruo, Madhu Mazumdar, Bruce D Minsky, Leonard Saltz, W Douglas Wong, Steven Larson, "Sequential preoperative fluorodeoxyglucose positron emission tomography assessment of response to preoperative chemoradiation: a means for determining longterm outcomes of rectal cancer," *J Am Coll Surg*, vol. 199, no. 1, pp. 1-7, 2004.
- [21] Genevieve B Melton, William C Lavelly, Heather A Jacene, Richard D Schulick, Michael A Choti, Richard L Wahl, Susan L Gearhart, "Efficacy of preoperative combined 18 fluorodeoxyglucose positron emission tomography and computed tomography for assessing primary rectal cancer response to neoadjuvant therapy," *J Gastrointest Surg*, vol. 11, no. 8, pp. 961-9, 2007.
- [22] Matthias R Benz, Martin S Allen-Auerbach, Fritz C Eilber, Hui J J Chen, Sarah Dry, Michael E Phelps, Johannes Czernin, Wolfgang A Weber, "Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas," *J Nucl Med*, vol. 49, no. 10, pp. 1579-84, 2008.
- [23] Patrick Flamen, Bruno Vanderlinden, Philippe Delatte, Ghanem Ghanem, Lieveke Ameye, Marc Van Den Eynde, Alain Hendlisz, "Multimodality imaging can predict the metabolic response of unresectable colorectal liver metastases to radioembolization therapy with Yttrium- 90 labeled resin microspheres," *Phys Med Biol*, vol. 53, no. 22, pp. 6591-603, 2008.
- [24] Fatma Yıldırım, Ahmet Selim Yurdakul, Sevket Özkaya, Ümit Özgür Akdemir, Can Öztürk, "Total lesion glycolysis by 18F-FDG PET/CT is independent prognostic factor in patients with advanced non-small cell lung cancer," *Clin Respir J*, vol. 11, no. 5, pp. 602-611, 2017.

- [25] D Groheux, G Quere, E Blanc, C Lemarignier, L Vercellino, C de Margerie-Mellon, P Merlet, S Querellou, "FDG PET-CT for solitary pulmonary nodule and lung cancer: Literature review," *Diagnostic and Interventional Imaging*, vol. 97, no. 10, pp. 1003-1017, 20216.
- [26] M K Gould, C C Maclean, W G Kuschner, C E Rydzak, D K Owens, "Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis," *JAMA*, vol. 285, no. 7, pp. 914-924, 2001.
- [27] Paul Cronin, Ben A. Dwamena, Aine Marie Kelly, Ruth C. Carlos, "Solitary pulmonary nodules: meta-analytic comparison of cross-sectional imaging modalities for diagnosis of malignancy," *Radiology*, vol. 246, no. 3, 2008.
- [28] Valentina Ambrosini, Silvia Nicolini, Paola Caroli, Cristina Nanni, Arianna Massaro, Maria Cristina Marzola, Domenico Rubello, Stefano Fanti, "PET/CT imaging in different types of lung cancer: an overview," *Eur J Radiol*, vol. 81, no. 5, pp. 988-1001, 2012.
- [29] Richard L Wahl, Heather Jacene, Yvette Kasamon, Martin A Lodge, "From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors," *J Nucl Med*, vol. 1, no. 1, pp. 122S-150S, 2009.
- [30] Norbert Avril, Stefanie Sassen, Barbara Schmalfeldt, Joerg Naehrig, Stephan Rutke, Wolfgang A Weber, Martin Werner, Henner Graeff, Markus Schwaiger, Walther Kuhn, "Prediction of response to neoadjuvant chemotherapy by sequential F-18-fluorodeoxyglucose positron emission tomography in patients with advanced-stage ovarian cancer," *J Clin Oncol*, vol. 23, no. 30, pp. 7445-7453, 2005.
- [31] Cleveland Clinic, Emeritus Staff, Cleveland, OH, USA, "radiologykey," [Online]. Available: <https://radiologykey.com/performance-characteristics-of-pet-scanners/>.
- [32] Hsin-Hon Lin, Keh-Shih Chuang, Cheng-Chang Lu et al, "Use of beam stoppers to correct random and scatter coincidence in PET: A Monte Carlo simulation," *Nuclear Instruments and Methods in Physics Research*, vol. 711, pp. 27-37, 2013.
- [33] Bui Tien Hung, Tran Thuy Duong , , Bui Ngoc Ha, Ho Quang Tuan, "A Monte Carlo simulation of Gamma ART-6000 devices to investigate beam characteristics in a homogeneous environment using Geant4 simulation toolkits," *Radiation Physics and Chemistry*, vol. 229, no. 112542, 2025.

- [34] Jeffrey P Schmall, Joel S Karp, Matt Werner, and Suleman Surti, "Parallax error in long-axial field-of-view PET scanners—a simulation study," *Phys Med Biol*, vol. 61, no. 14, pp. 5443-5455, 2016.
- [35] IAEA Human Health Series No.27. PET/CT Atlas on Quality Control and Image Artefact, Vienna: IAEA, 2014.
- [36] Marine Soret, Stephen L. Bacharach, and Irene Buvat, "Partial-Volume Effect in PET Tumor Imaging," *J Nucl Med*, vol. 48, no. 932-945, 2007.
- [37] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, "Deep learning," *nature*, vol. 521, pp. 436-444, 2015.
- [38] Danilo Bzdok, Naomi Altman & Martin Krzywinski, "Statistics versus machine learning," *Nature methods*, vol. 15, no. 4, pp. 233-234, 2018.
- [39] "Deep learning vs. Machine learning," IBM, 2023. [Online]. Available: <https://www.ibm.com/topics/deep-learning>.
- [40] Lei Cai, Jingyang Gao, Di Zhao, "A review of the application of deep learning in medical image classification and segmentation," *Annals of Translational Medicine*, vol. 8, no. 11, 2020;.
- [41] Md. Eshmam Rayed, S.M. Sajibul Islam, Sadia Islam Niha, Jamin Rahman Jim, Md Mohsin Kabir, M.F. Mridha, "Deep learning for medical image segmentation: State-of-the-art advancements and challenges," *Informatics in Medicine Unlocked*, vol. 47, 2024.
- [42] Maryam Fallahpoor, Subrata Chakraborty, Biswajeet Pradhan, Oliver Faust, Prabal Datta Barua, Hossein Chegeni, Rajendra Acharya, "Deep learning techniques in PET/CT imaging: A comprehensive review from sinogram to image space," *Comput Methods Programs Biomed*, vol. 243, no. 107880, 2024.
- [43] Yan Sun, Xinyu Ge, Rong Niu, Jianxiong Gao, Yunmei Shi, Xiaoliang Shao, Yuetao Wang and Xiaonan Shao, "PET/CT radiomics and deep learning in the diagnosis of benign and malignant pulmonary nodules: progress and challenges," *Front. Oncol*, vol. 14, 2024.
- [44] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud G P M van Stiphout, Patrick Granton, Catharina M L Zegers, Robert Gillies, Ronald Boellard, André Dekker, Hugo J W L Aerts, "Radiomics: extracting more information from medical images using advanced feature analysis," *Eur J Cancer*, vol. 48, no. 4, pp. 441-6, 2012.

- [45] Song Chen, Stephanie Harmon, Timothy Perk, Xuena Li, Meijie Chen, Yaming Li & Robert Jeraj, "Diagnostic classification of solitary pulmonary nodules using dual time (18)F-fdg pet/ct image texture features in granuloma-endemic regions," *Scientific Reports*, vol. 7, no. 9370, 2017.
- [46] Song Chen, Stephanie Harmon, Timothy Perk, Xuena Li, Meijie Chen, Yaming Li, Robert Jeraj, "Using neighborhood gray tone difference matrix texture features on dual time point pet/ct images to differentiate Malignant from benign fdg-avid solitary pulmonary nodules," *Cancer Imaging*, vol. 19, no. 1, 2019.
- [47] Jianping Zhang, Guang Ma, Jingyi Cheng, Shaoli Song, Yingjian Zhang, L Q Shi, "Diagnostic classification of solitary pulmonary nodules using support vector machine model based on 2-[18f]Fluoro-2-deoxy-D-glucose pet/computed tomography texture features," *Nucl Med Commun*, vol. 41, no. 6, pp. 560-6, 2020.
- [48] Barbara Palumbo 1, Francesco Bianconi, Isabella Palumbo, Mario Luca Fravolini, Matteo Minestrini, Susanna Nuvoli, Maria Lina Stazza, Maria Rondini, Angela Spanu, "Value of shape and texture features from (18)F-fdg pet/ct to discriminate between benign and Malignant solitary pulmonary nodules: an experimental evaluation.," *Diagnostics (Basel)*, vol. 10, no. 9, 2020.
- [49] Yujing Hu, Xinming Zhao, Jianyuan Zhang, Jingya Han, Meng Dai, "Value of (18)F-fdg pet/ct radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis," *Eur J Nucl Med Mol Imaging*, vol. 48, no. 1, pp. 231-240, 2021.
- [50] Hyungjin Kim, Dongheon Lee, Woo Sang Cho, Jung Chan Lee, Jin Mo Goo, Hee Chan Kim, Chang Min Park, "CT-based deep learning model to differentiate invasive pulmonary adenocarcinomas appearing as subsolid nodules among surgical candidates: comparison of the diagnostic performance with a size-based logistic model and radiologists," *European Radiology*, vol. 30, no. 6, pp. 3295-3305, 2020.
- [51] Wei Zhao, Jiancheng Yang, Yingli Sun, Cheng Li, Weilan Wu, Liang Jin, Zhiming Yang, Bingbing Ni, Pan Gao, Peijun Wang, Yanqing Hua, Ming Li, "3d deep learning from ct scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas," *Cancer Res*, vol. 78, no. 24, pp. 6881-6889, 2018.

- [52] Victor Manuel Alves, Jaime Dos Santos Cardoso, João GamaVictor Manuel Alves, Jaime Dos Santos Cardoso, João Gama, "Classification of pulmonary nodules in 2-[(18)F]Fdg pet/ct images with a 3d convolutional neural network," *Nucl Med Mol Imaging*, vol. 58, no. 1, pp. 9-24, 2024.
- [53] Yung-Chi Lai, Kuo-Chen Wu, Neng-Chuan Tseng, Yi-Jin Chen, Chao-Jen Chang, Kuo-Yang Yen, Chia-Hung Kao, "Differentiation between Malignant and benign pulmonary nodules by using automated three-dimensional high-resolution representation learning with fluorodeoxyglucose positron emission tomography-com," *Front Med*, no. 18:9:773041, 2022.
- [54] Yong-Jin Park, Dongmin Choi, Joon Young Choi, Seung Hyup Hyun, "Performance Evaluation of a Deep Learning System for Differential Diagnosis of Lung Cancer With Conventional CT and FDG PET/CT Using Transfer Learning and Metadata," *Clin Nucl Med*, vol. 46, no. 8, pp. 635-640, 2021.
- [55] Xiaonan Shao^{1,2†}, Rong Niu^{1,2†}, Xiaoliang Shao, Jianxiong Gao, Yunmei Shi, Zhenxing Jiang and Yuetao Wang, "Application of dual-stream 3d convolutional neural network based on (18)F-fdg pet/ct in distinguishing benign and invasive adenocarcinoma in ground-glass lung nodules," *EJNMMI Phys*, vol. 8, no. 74, 2021.
- [56] Xiaolei Zhang, Xianling Dong, M Iqbal Bin Saripan, Dongyang Du, Yanjun Wu 2, Zhongxiao Wang 2, Zhendong Cao, Dong Wen, Yanli Liu, Mohammad Hamiruce Marhaban, "Deep learning pet/ct-based radiomics integrates clinical data: A feasibility study to distinguish between tuberculosis nodules and lung cancer," *Thorac Cancer*, vol. 14, no. 19, pp. 1802-1811, 2023.
- [57] Jiatian Zhang, Yajing Zhao , Yiping Lu, Peng Li, Shijie Dang, Xuanxuan Li, Bo Yin, Lingxiao Zhao, "Meningioma consistency assessment based on the fusion of deep learning features and radiomics features," *Eur J Radiol*, no. 170:111250, 2024.
- [58] Jingchi Zheng, Yue Hao, Yan Guo, Ming Du, Pengyuan Wang, Jun Xin, "An 18f-fdg-pet/ct-based radiomics signature for estimating Malignance probability of solitary pulmonary nodule," *Clin Respir J*, vol. 18, no. 5, 2024.
- [59] Jing Ning, Can Li, Peng Yu, Jingjing Cui, Xiaodan Xu, Yan Jia, Panli Zuo, Jiahe Tian, Lukas Kenner and Baixuan Xu, "Radiomic analysis will add differential diagnostic value of benign and malignant pulmonary nodules: a hybrid imaging study based on [18F]FDG and [18F]FLT PET/CT," *Insights into Imaging*, vol. 14, no. 197, 2023.

- [60] Atsushi Teramoto, Masakazu Tsujimoto, Takahiro Inoue, Tetsuya Tsukamoto, Kazuyoshi Imaizumi, Hiroshi Toyama, Kuniaki Saito, Hiroshi Fujita, "Automated classification of pulmonary nodules through a retrospective analysis of conventional ct and two-phase pet images in patients undergoing biopsy," *Asia Ocean J Nucl Med Biol.*, vol. 7, no. 1, pp. 29-37, 2019.
- [61] Wei Zhao, Jiancheng Yang, Bingbing Ni, Dexi Bi, Yingli Sun, Mengdi Xu, Xiaoxia Zhu, Cheng Li, Liang Jin, Pan Gao, Peijun Wang, Yanqing Hua, Ming Li, "Toward automatic prediction of EGFR mutation status in pulmonary adenocarcinoma with 3D deep learning," *Cancer Med*, vol. 8, no. 7, pp. 3532-3543, 2019.
- [62] Andrea Ciarmiello, Nikola Yosifov, Donatella Masciale, Ornella Ferrando, Franca Foppiano, Amalia Milano, Massimo Canevari, Luigia Florimonte, Massimo Castellani, Giampiero Giovacchini, Lorenzo Stefano Maffioli & Bruno Alfano, "Global and regional accuracy of deep learning-based tumor segmentation from whole-body [18F]fluorodeoxyglucose PET/CT images," *EJNMMI Research*, vol. 16, no. 3, 2026.
- [63] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [64] Hao Guan, Mingxia Liu, "Domain Adaptation for Medical Image Analysis: A Survey," *IEEE Trans Biomed Eng*, vol. 69, no. 3, pp. 1173-1185, 2022.
- [65] H. Q. Tuan, T. T. Duong, B. N. Ha et al, "Development of a Vietnamese PET/CT Dataset for Machine Learning-Based Analysis of Non-Small Cell Lung Cancer Images," *Atom Indonesia*, vol. 51, no. 2, pp. 137-143, 2025.
- [66] Simon K Warfield, Kelly H Zou, William M Wells, "Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation," *IEEE Trans Med Imaging*, vol. 23, no. 7, pp. 903-921, 2004.
- [67] Anne-Sophie Dewalle-Vignion, Nacim Betrouni, Clio Baillet, Maximilien Vermandel, "Is STAPLE algorithm confident to assess segmentation methods in PET imaging?," *Phys Med Biol*, vol. 60, no. 24, pp. 9473-9491, 2015.
- [68] Fabian Isensee, Tassilo Wald, Constantin Ulrich, Michael Baumgartner, Saikat Roy, Klaus Maier-Hein & Paul F. Jäger, "nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation," in *Medical*

- [69] Jakob Dexe, Sergios Gatidis, Marcel Früh, Katharina Jeblick et al, "AutoPET Challenge on Fully Automated Lesion Segmentation in Oncologic PET/CT Imaging, Part 2: Domain Generalization," *J Nucl Med*, vol. 67, no. 3, pp. 481-488, 2026.
- [70] "LUNG-PET-CT-Dx Dataset," The Cancer Imaging Archive (TCIA)., 2024. [Online]. Available: <https://www.cancerimagingarchive.net/>.
- [71] Lena Maier-Hein, Annika Reinke, Patrick Godau et al, "Metrics reloaded: recommendations for image analysis validation," *Nat Methods*, vol. 21, no. 2, pp. 195-212, 2024.
- [72] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox & Olaf Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Athens, Greece, 2016.
- [73] M. Jorge Cardoso, Wenqi Li, Richard Brown et al., "MONAI: An open-source framework for deep learning in healthcare," 2022. [Online]. Available: <https://arxiv.org/abs/2211.02701>.
- [74] Ronald Boellaard et al, "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0," *Eur J Nucl Med Mol Imaging*, vol. 42, no. 2, pp. 328-54, 2015.
- [75] Mahul B Amin et al, "The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging," *CA Cancer J Clin*, vol. 67, no. 2, pp. 93-99, 2017.
- [76] Heber MacMahon et al, "Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017.," *Radiology*, vol. 284 number 1, 2017.
- [77] Andriy Fedorov et al, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magn Reson Imaging*, vol. 30, no. 9, pp. 1323-41, 2012.
- [78] Goram Mufarah Alshmrani, Qiang Ni, Richard Jiang and Nada Muhammed, "Hyper-Dense_Lung_Seg: Multimodal-Fusion-Based Modified U-Net for

Lung Tumour Segmentation Using Multimodality of CT-PET Scans," *Diagnostics*, vol. 13, no. 22, 2023.

- [79] Robert J Gillies, Paul E Kinahan, Hedvig Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, no. 2, pp. 563-77, 2016.
- [80] Kenneth Clark et al, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *J Digit Imaging*, vol. 26, no. 6, pp. 1045-1057, 2013.
- [81] Hugo J W L Aerts et al, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat Commun*, vol. 5:4006, 2014.
- [82] Nicholas E Protonotarios, Iason Katsamenis, Stavros Sykiotis, Nikolaos Dikaos 3, George A Kastis 3, Sofia N Chatziioannou, Marinos Metaxas, Nikolaos Doulamis, Anastasios Doulamis, "A few-shot U-Net deep learning model for lung cancer lesion segmentation via PET/CT imaging," *Biomed Phys Eng Express*, vol. 8, no. 2, 2022.
- [83] Mary L McHugh, "Interrater reliability: The kappa statistic," *Biochem Med (Zagreb)*, vol. 22, no. 3, pp. 276-282, 2012.
- [84] Vladimir Vezhnevets. Vadim Konushin, "GrowCut: Interactive multi-label N-D image segmentation by cellular automata.," in *Proceedings of Graphicon*, 2005.
- [85] Chintan Parmar et al, "Machine learning methods for quantitative radiomic biomarkers.," *Scientific Reports*, vol. 5, 2015.
- [86] Alex Zwanenburg et al, "The Image Biomarker Standardization Initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328-338, 2020.
- [87] Philippe Lambin et al, "Radiomics: The bridge between medical imaging and personalized medicine," *Nat Rev Clin Oncol*, vol. 14, no. 12, pp. 749-762, 2017.
- [88] YUANDA ZHENG et al, "FDG-PET/CT imaging for tumor staging and definition of tumor volumes in radiation treatment planning in non-small cell lung cancer," *Oncol Lett*, vol. 7, no. 4, pp. 1015-1020, 2014.
- [89] B. N. Ha, N. T. Kien, T. T. Duong, T. N. Toan, H. Q. Tuan, "Development of Brain MRI Image Segmentation program using UNET++ network for

radiosurgery planning," *Nuclear science and technology*, vol. 13, no. 1, pp. 1-9, 2023.

- [90] Olaf Ronneberger, Philipp Fischer & Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Munich, Germany, 2015.
- [91] Lu Zhou, Chaoyong Wu, Yiheng Chen and Zhicheng Zhang, "Multitask connected U-Net: automatic lung cancer segmentation from CT images using PET knowledge guidance," *Frontiers in Artificial Intelligence*, vol. 7, 2024.
- [92] Fabian Isensee, Paul F Jaeger, Simon A A Kohl, Jens Petersen, Klaus H Maier-Hein, "nnunet: a self-configuring method for deep learning-based biomedical image segmentation," *Nat Methods*, vol. 18, no. 2, pp. 203-211, 2021.
- [93] Matteo Ferrante, Lisa Rinaldi, Francesca Botta, Xiaobin Hu, Andreas Dolp, Marta Minotti, Francesca De Piano, Gianluigi Funicelli, Stefania Volpe, Federica Bellerba, Paolo De Marco, Sara Raimondi, Stefania Rizzo, Kuangyu Shi, "Application of nnu-net for automatic segmentation of lung lesions on ct images and its implication for radiomic models," *J Clin Med*, vol. 11, no. 24, 2022.
- [94] Konstantinos Kamnitsas, Christian Ledig, Virginia F J Newcombe 3, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, Ben Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med Image Anal*, vol. 36, pp. 61-78, 2017.
- [95] Jonathan Long, Evan Shelhamer, Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [96] Andriy Myronenko, "3D MRI Brain Tumor Segmentation Using Autoencoder Regularization," in *MICCAI 2018*, Granada, Spain, 2018.
- [97] Chenhong Zhou, Shengcong Chen, Changxing Ding, Dacheng Tao, "Learning Contextual and Attentive Information for Brain Tumor Segmentation," in *4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018*, Granada, Spain, 2018.
- [98] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby, "An Image is

Worth 16x16 Words: Transformers for Image Recognition at Scale," 3rd June 2021. [Online]. Available: <https://arxiv.org/abs/2010.11929>.

- [99] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao et al, "Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [100] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei et al, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.
- [101] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," 8 Feb 2021. [Online]. Available: <https://arxiv.org/abs/2102.04306>.
- [102] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger Roth, Daguang Xu, "UNETR: Transformers for 3D Medical Image Segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021.
- [103] Wenxuan Wang, Chen Chen, Meng Ding, Hong Yu, Sen Zha, Jiangyun Li, "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021: 24th International Conference*, Strasbourg, France, 2021.
- [104] Yutong Xie, Jianpeng Zhang, Chunhua Shen, Yong Xia, "CoTr: Efficiently Bridging CNN and Transformer for 3D Medical Image Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, 2021.
- [105] Michela Antonelli, Annette Kopp-Schneider et al, "The Medical Segmentation Decathlon," *NATURE COMMUNICATIONS*, vol. 13, no. 4128, 2022.
- [106] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, "Rethinking the Inception Architecture for Computer Vision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [107] Sergey Ioffe, Christian Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

- [108] Mojtaba Seyedhosseini, Mehdi Sajjadi, Tolga Tasdizen, "Image Segmentation with Cascaded Hierarchical Models and Logistic Disjunctive Normal Networks," in *Proc IEEE Int Conf Comput Vis*, 2013.
- [109] Bharath Hariharan, Pablo Arbeláez, Ross Girshick, Jitendra Malik, "Hypercolumns for Object Segmentation and Fine-grained Localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [110] Fausto Milletaria, Seyed-Ahmad Ahmadi et al, "Hough-CNN: Deep Learning for Segmentation of Deep Brain Regions in MRI and Ultrasound," *Computer Vision and Image Understanding*, vol. 164, pp. 92-102, 2017.
- [111] Jens Kleesiek, Gregor Urban, Alexander Hubert, Daniel Schwarz, Klaus Maier-Hein, Martin Bendszus, Armin Biller, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *Neuroimage*, vol. 1, no. 129, pp. 160-169, 2016.
- [112] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri, "Deep End2End Voxel2Voxel Prediction," 2015. [Online]. Available: <http://arxiv.org/abs/1511.06681>.
- [113] Fabian Isensee, Klaus H.Maier-Hein, "Look Ma, no code: fine tuning nnU-Net for the AutoPET II challenge by only adjusting its JSON plans," 2023. [Online]. Available: <https://arxiv.org/abs/2309.13747>.
- [114] J M Bland, D G Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, no. 8476, pp. 307-310, 1986.
- [115] Hiroaki Nomori et al, "Positron emission tomography in lung cancer," *General Thoracic and Cardiovascular Surgery*, vol. 57, pp. 184-191, 2009.
- [116] Sheila Rankin, "PET/CT for staging and monitoring non small cell lung cancer. Cancer Imaging," *Cancer Imaging*, vol. 8, pp. S27-S31, 2008.
- [117] "Nghiên cứu thuật toán định vị và phân đoạn khối u phổi từ các thể tích lồng ngực PET/CT dựa trên phân tích đặc điểm hình ảnh," *Tạp chí y học Việt Nam*, vol. 533, no. Số đặc biệt tháng 12, 2023.