

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

.....***.....



Nguyễn Thị Trang

**NGHIÊN CỨU PHƯƠNG PHÁP HỢP NHẤT ĐẶC TRƯNG
ĐA NGUỒN CHO PHÂN LOẠI ẢNH TỔN THƯƠNG DA**

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Hà Nội - Năm 2026

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

.....***.....



Nguyễn Thị Trang

**NGHIÊN CỨU PHƯƠNG PHÁP HỢP NHẤT ĐẶC TRƯNG
ĐA NGUỒN CHO PHÂN LOẠI ẢNH TỔN THƯƠNG DA**

LUẬN ÁN TIẾN SĨ MÁY TÍNH

NGÀNH: KHOA HỌC MÁY TÍNH

MÃ SỐ: 9 48 01 01

**Xác nhận của Học viện
Khoa học và Công nghệ**

Người hướng dẫn 1
(Ký, ghi rõ họ tên)

Người hướng dẫn 2
(Ký, ghi rõ họ tên)

Hà Nội – Năm 2026

LỜI CAM ĐOAN

Tôi xin cam đoan rằng luận án này là công trình nghiên cứu do tôi thực hiện dưới sự hướng dẫn khoa học của PGS.TS. Nguyễn Việt Anh và PGS.TS. Vũ Văn Hiệu.

Các kết quả nghiên cứu, số liệu thực nghiệm và nội dung trình bày trong luận án là trung thực, phản ánh đúng quá trình nghiên cứu của tác giả và chưa từng được công bố trong bất kỳ công trình nào khác dưới danh nghĩa một luận án hoặc luận văn để bảo vệ học vị, trừ những nội dung đã được trích dẫn, kế thừa và ghi rõ nguồn tài liệu tham khảo theo đúng quy định.

Tôi xin cam đoan rằng mọi sự giúp đỡ về chuyên môn, học thuật và kỹ thuật trong quá trình thực hiện luận án đều đã được trân trọng ghi nhận; mọi tài liệu, kết quả, số liệu và nhận định của các tác giả khác được sử dụng trong luận án đều được trích dẫn đầy đủ, chính xác và minh bạch.

Tôi hoàn toàn chịu trách nhiệm trước pháp luật và quy định của cơ sở đào tạo về tính trung thực, chính xác và nguyên gốc của các nội dung trình bày trong luận án này.

Hà Nội, ngày 12 tháng 9 năm 2026

Tác giả luận án

Nguyễn Thị Trang

LỜI CẢM ƠN

Trước hết, tác giả xin bày tỏ lòng biết ơn sâu sắc và trân trọng tới PGS.TS. Nguyễn Việt Anh và PGS.TS. Vũ Văn Hiệu, là những người thầy hướng dẫn khoa học đã tận tình định hướng, chỉ bảo và hỗ trợ tác giả trong suốt quá trình học tập, nghiên cứu và hoàn thành luận án. Những ý kiến chuyên môn sâu sắc, sự nghiêm túc trong nghiên cứu khoa học và sự động viên kịp thời của các thầy là nguồn động lực quan trọng giúp tác giả vượt qua những khó khăn trong quá trình thực hiện luận án.

Tác giả xin trân trọng cảm ơn Ban Giám đốc, các thầy cô giáo, các nhà khoa học và cán bộ của Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tạo điều kiện thuận lợi về học thuật, chuyên môn và môi trường nghiên cứu trong suốt thời gian học tập và thực hiện luận án.

Tác giả xin chân thành cảm ơn Ban Giám hiệu, lãnh đạo đơn vị, cùng các đồng nghiệp tại Trường Đại học Hạ Long, nơi tác giả đang công tác, đã luôn quan tâm, tạo điều kiện thuận lợi và hỗ trợ tác giả trong quá trình học tập, nghiên cứu và hoàn thiện luận án.

Tác giả cũng xin chân thành cảm ơn các đồng nghiệp, bạn bè và những người đã luôn quan tâm, chia sẻ, động viên và hỗ trợ trong suốt quá trình học tập, nghiên cứu và hoàn thiện luận án.

Cuối cùng, tác giả xin bày tỏ lòng biết ơn sâu sắc tới gia đình và người thân đã luôn ở bên cạnh, động viên, chia sẻ và tạo mọi điều kiện thuận lợi cả về tinh thần lẫn vật chất để tác giả yên tâm học tập, nghiên cứu và hoàn thành luận án này.

Tác giả xin trân trọng cảm ơn.

Hà Nội, ngày 12 tháng 9 năm 2026

Tác giả luận án

Nguyễn Thị Trang

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT	vi
DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ	xiii
DANH MỤC CÁC BẢNG	xv
DANH MỤC CÁC ĐỊNH NGHĨA	xvi
 MỞ ĐẦU	 1
 CHƯƠNG 1 TỔNG QUAN VỀ PHÂN LOẠI TỔN THƯƠNG DA VÀ CƠ SỞ LÝ THUYẾT	 8
1.1 Tổng quan bài toán phân loại ảnh tổn thương da	8
1.2 Biểu diễn ảnh và các loại đặc trưng	9
1.2.1 Lựa chọn đặc trưng có giám sát bằng tiêu chí Chi-square	9
1.2.2 Đặc trưng thủ công: màu sắc, kết cấu, hình dạng	11
1.2.3 Đặc trưng học sâu trong biểu diễn ảnh tổn thương da	14
1.2.4 Các hướng khai thác và lựa chọn bản đồ đặc trưng học sâu	15
1.3 Hợp nhất đặc trưng và học đa nguồn	17
1.3.1 Hợp nhất đặc trưng sâu và đặc trưng thủ công	18
1.3.2 Hợp nhất ảnh và dữ liệu mô tả lâm sàng	18
1.3.3 Các phương pháp hợp nhất phổ biến	18
1.3.4 Hạn chế của các phương pháp hợp nhất trực tiếp	20
1.4 Entropy trong học biểu diễn và đánh giá độ tin cậy	20
1.4.1 Entropy trên phân phối xác suất	20
1.4.2 Entropy trên biểu diễn đặc trưng	21
1.4.3 Vai trò của entropy	23
1.5 Học trên đồ thị và học bán giám sát	25
1.5.1 Biểu diễn dữ liệu trên đồ thị	25
1.5.2 Học bán giám sát	27
1.5.3 Rủi ro của gán nhãn giả	27
1.6 Đánh giá mô hình phân loại ảnh y tế	28
1.6.1 Các chỉ số theo lớp trong dữ liệu mất cân bằng	28
1.6.2 ROC, AUC và các độ đo tổng hợp	29
1.7 Cơ sở hình thành hướng tiếp cận của luận án	31
1.7.1 Các vấn đề nghiên cứu cần tiếp tục giải quyết	31
1.7.2 Định hướng tiếp cận của luận án	33
Kết luận chương	33
 CHƯƠNG 2 HỢP NHẤT ĐẶC TRƯNG ĐA NGUỒN DỰA TRÊN CHỌN LỘC ENTROPY CHO PHÂN LOẠI ẢNH TỔN THƯƠNG DA	 35
2.1 Phát biểu bài toán	36
2.2 Ý tưởng phương pháp đề xuất	38
2.3 Phương pháp BlockNet đề xuất	41

2.3.1	Kiến trúc tổng quan của BlockNet	41
2.3.2	Trích xuất và biểu diễn bản đồ đặc trưng từ CNN	43
2.3.3	Xác định khoảng entropy phù hợp	46
2.3.4	Chọn lọc bản đồ đặc trưng dựa trên entropy	51
2.3.5	Hợp nhất đặc trưng đa nguồn	55
2.3.6	Cân bằng dữ liệu trong không gian đặc trưng	57
2.3.7	Bộ phân loại và hàm mục tiêu	60
2.4	Thuật toán đề xuất và quy trình thực hiện	61
2.5	Thiết lập thực nghiệm	65
2.5.1	Bộ dữ liệu	65
2.5.2	Các mô hình so sánh	69
2.5.3	Cấu hình thực nghiệm	70
2.5.4	Các độ đo đánh giá	72
2.6	Kết quả thực nghiệm	72
2.6.1	Đánh giá ảnh hưởng của chọn lọc bản đồ đặc trưng dựa trên entropy	72
2.6.2	Đánh giá kết quả hợp nhất đặc trưng đa nguồn	76
2.7	Phân tích và thảo luận kết quả thực nghiệm	81
2.7.1	Giải thích sự chênh lệch giữa hai nhóm kết quả trên HAM10000	82
2.7.2	Ảnh hưởng của chọn lọc bản đồ đặc trưng dựa trên entropy	83
2.7.3	Ảnh hưởng của các block CNN	83
2.7.4	Hiệu quả của hợp nhất đặc trưng đa nguồn	84
2.7.5	Ảnh hưởng của cân bằng trong không gian đặc trưng	84
2.7.6	Một số nhận xét tổng quát	85
	Kết luận chương	85

CHƯƠNG 3 MÔ HÌNH PHÂN LOẠI TỔN THƯƠNG DA DỰA TRÊN HỌC ĐA NGUỒN VÀ ĐỒ THỊ

		87
3.1	Phát biểu bài toán	89
3.2	Cơ sở đề xuất phương pháp	92
3.3	Phương pháp đề xuất	94
3.3.1	Trích xuất đặc trưng đa nguồn	95
3.3.2	Hợp nhất đặc trưng đa nguồn có xét metadata	97
3.3.3	Xây dựng đồ thị từ biểu diễn hợp nhất	98
3.3.4	Pha 1: Huấn luyện GNN có giám sát	101
3.3.5	Pha 2: Gán nhãn giả có chọn lọc dựa trên độ tin cậy	103
3.4	Thuật toán và quy trình thực hiện	109
3.5	Thiết lập thực nghiệm	113
3.5.1	Bộ dữ liệu ISIC 2020	114
3.5.2	Metadata lâm sàng	115
3.5.3	Lựa chọn đặc trưng	117
3.5.4	Chỉ số đánh giá	119
3.5.5	Các mô hình so sánh	120
3.5.6	Thiết lập môi trường thực nghiệm	120
3.5.7	Thiết lập tham số mô hình	121
3.6	Kết quả thực nghiệm	122
3.6.1	Kết quả của pha huấn luyện có giám sát	122
3.6.2	Kết quả của pha gán nhãn giả dựa trên độ tin cậy	125

3.6.3	Kết quả phân tích theo mức độ tin cậy	129
3.7	Phân tích và thảo luận kết quả thực nghiệm	130
	Kết luận chương	132
	KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TƯƠNG LAI	133
	TÀI LIỆU THAM KHẢO	137

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

1. Các ký hiệu

Ký hiệu	Tiếng Anh	Tiếng Việt
$\mathcal{X}$	Input space	Không gian đầu vào
$\mathcal{Y}$	Label space	Không gian nhãn
D	Dataset	Tập dữ liệu
D_L	Labeled dataset	Tập dữ liệu có nhãn
D_U	Unlabeled dataset	Tập dữ liệu chưa gán nhãn
D_{val}	Validation dataset	Tập dữ liệu kiểm định
$D_{\text{pseudo}}^{\text{High}}$	High trust pseudo labeled dataset	Tập nhãn giả có độ tin cậy cao
D_{new}	Extended training dataset	Tập huấn luyện mở rộng
$D_{\text{balanced}}^{(b)}$	Balanced feature dataset at block b	Tập dữ liệu đặc trưng đã cân bằng tại block b
x_i	Input sample	Mẫu đầu vào thứ i
y_i	Reference label	Nhãn tham chiếu của mẫu thứ i
$\hat{y}_i$	Predicted label	Nhãn dự đoán của mẫu thứ i
$\hat{\mathbf{y}}_i$	Predicted probability vector	Vector xác suất dự đoán của mẫu thứ i
N	Number of samples	Số lượng mẫu
C	Number of classes	Số lớp
$\mathcal{B}$	Set of convolutional blocks	Tập các block tích chập
b	Block index	Chỉ số block tích chập
$F_m^{(b)}(x_i)$	Feature map m at block b	Bản đồ đặc trưng thứ m tại block b
$\mathcal{F}^{(b)}(x_i)$	Feature map set at block b	Tập bản đồ đặc trưng tại block b
$\mathcal{F}^{*(b)}(x_i)$	Selected feature map set	Tập bản đồ đặc trưng được giữ lại
M_b	Number of feature maps at block b	Số bản đồ đặc trưng tại block b
$H_{\text{act}}(F)$	Activation entropy	Entropy kích hoạt của bản đồ đặc trưng
$\theta^{(b)}$	Minimum entropy threshold	Ngưỡng entropy tối thiểu tại block b
$e_{\min}^{(b)}, e_{\max}^{(b)}$	Valid entropy bounds	Cận dưới và cận trên của khoảng entropy phù hợp

$z_i^{(d,b)}$	Deep feature vector at block b	Vector đặc trưng học sâu tại block b
$z_i^{(h)}$	Handcrafted feature vector	Vector đặc trưng thủ công
$z_i^{(b)}$	Fused feature vector at block b	Vector đặc trưng hợp nhất tại block b
$n_k^{(b)}$	Number of feature vectors in class k	Số vector đặc trưng của lớp k tại block b
$m_k^{(b)}$	Feature balancing multiplier	Hệ số tăng cường đặc trưng của lớp k tại block b
$n_k'^{(b)}$	Balanced class size	Số vector đặc trưng của lớp k sau cân bằng tại block b
$\Delta_k^{(b)}$	Number of additional feature vectors	Số vector đặc trưng cần bổ sung cho lớp k tại block b
$\psi(\cdot)$	Feature fusion operator	Toán tử hợp nhất đặc trưng
$g_{\omega}^{(b)}$	Classifier at block b	Bộ phân loại tại block b
$\omega^{(b)}$	Classifier parameters	Tham số bộ phân loại tại block b
$\mathbf{x}_i^{deep}$	Deep feature vector	Vector đặc trưng học sâu
$\mathbf{x}_i^{hand}$	Handcrafted feature vector	Vector đặc trưng thủ công
$\mathbf{x}_i^{meta}$	Clinical metadata vector	Vector metadata lâm sàng
$\mathbf{x}_i^{img}$	Image representation vector	Vector biểu diễn ảnh
$\mathbf{h}_i$	Fused node representation	Biểu diễn nút hợp nhất
$\mathbf{z}_i$	Graph learned node representation	Biểu diễn nút sau khi học trên đồ thị
$\Phi(\cdot)$	Metadata aware fusion function	Hàm hợp nhất có xét metadata
α_i^{img}	Image branch attention weight	Trọng số chú ý của đường ảnh
α_i^{meta}	Metadata branch attention weight	Trọng số chú ý của đường metadata
$\mathbf{W}_{img}$	Image projection matrix	Mã trận chiếu của biểu diễn ảnh
$\mathbf{W}_{meta}$	Metadata projection matrix	Mã trận chiếu của metadata
$G = (V, E)$	Graph	Đồ thị
V	Node set	Tập nút
E	Edge set	Tập cạnh

$\mathcal{N}_k(v_i)$	Set of k nearest neighbors	Tập k nút láng giềng gần nhất của nút v_i
s_{ij}	Cosine similarity	Độ tương đồng cosine giữa hai nút
f_θ	Classification model	Mô hình phân loại
θ	Model parameters	Tham số mô hình
$\mathcal{L}_{sup}$	Supervised loss	Hàm mất mát có giám sát
$H_{\text{pred}}(\hat{y}_i)$	Predictive entropy	Entropy dự đoán của mẫu thứ i
q_1, q_2	Entropy quantiles	Các phân vị entropy
α_1, α_2	Entropy thresholds	Các ngưỡng phân tầng độ tin cậy
TP	True positive	Dương tính đúng
TN	True negative	Âm tính đúng
FP	False positive	Dương tính giả
FN	False negative	Âm tính giả

2. Các chữ viết tắt

Chữ viết tắt	Tiếng Anh	Tiếng Việt
ACC	Accuracy	Độ chính xác
AUC	Area Under the Curve	Diện tích dưới đường cong
ROC	Receiver Operating Characteristic	Đường cong đặc trưng hoạt động của bộ thu
F1	F1 score	Điểm F1
CNN	Convolutional Neural Network	Mạng nơ ron tích chập
DNN	Deep Neural Network	Mạng nơ ron sâu
MLP	Multilayer Perceptron	Mạng perceptron nhiều lớp
GNN	Graph Neural Network	Mạng nơ ron đồ thị
GCN	Graph Convolutional Network	Mạng tích chập đồ thị
GAT	Graph Attention Network	Mạng chú ý đồ thị
GIN	Graph Isomorphism Network	Mạng đẳng cấu đồ thị
GINE	Graph Isomorphism Network with Edge features	Mạng đẳng cấu đồ thị có đặc trưng cạnh
GraphSAGE	Graph Sample and Aggregate	Mô hình lấy mẫu và tổng hợp trên đồ thị
SVM	Support Vector Machine	Máy vector hỗ trợ

KNN	K Nearest Neighbors	Phương pháp k láng giềng gần nhất
RF	Random Forest	Rừng ngẫu nhiên
RGB	Red Green Blue	Không gian màu đỏ, lục và lam
HSV	Hue Saturation Value	Không gian màu sắc độ, độ bão hòa và độ sáng
GLCM	Gray Level Cooccurrence Matrix	Ma trận đồng xuất hiện mức xám
HOG	Histogram of Oriented Gradients	Histogram hướng gradient
LBP	Local Binary Pattern	Mẫu nhị phân cục bộ
SE	Squeeze and Excitation	Cơ chế nén và kích thích
CBAM	Convolutional Block Attention Module	Mô đun chú ý cho block tích chập
SMOTE	Synthetic Minority Oversampling Technique	Kỹ thuật tăng mẫu tổng hợp cho lớp thiểu số
ECE	Expected Calibration Error	Sai số hiệu chỉnh kỳ vọng
ISIC	International Skin Imaging Collaboration	Tổ chức hợp tác quốc tế về ảnh da
HAM10000	Human Against Machine with 10000 training images	Bộ dữ liệu ảnh tổn thương da HAM10000
AKIEC	Actinic Keratoses and Intraepithelial Carcinoma	Dày sừng ánh sáng và ung thư biểu mô trong biểu bì
BCC	Basal Cell Carcinoma	Ung thư biểu mô tế bào đáy
BKL	Benign Keratosis Like Lesions	Tổn thương dạng dày sừng lành tính
DF	Dermatofibroma	U xơ da
MEL	Melanoma	U hắc tố
MM	Malignant Melanoma	U hắc tố ác tính
NV	Melanocytic Nevus	Nốt ruồi sắc tố
SCC	Squamous Cell Carcinoma	Ung thư biểu mô tế bào vảy
VASC	Vascular Lesions	Tổn thương mạch máu
CT1	Publication 1	Công trình công bố thứ nhất của luận án
CT2	Publication 2	Công trình công bố thứ hai của luận án
CT3	Publication 3	Công trình công bố thứ ba của luận án

CT4	Publication 4	Công trình công bố thứ tư của luận án
-----	---------------	--

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

Hình 1	Khoảng trống nghiên cứu của hệ thống phân loại ảnh tổn thương da. Các vùng khoanh đỏ thể hiện những hạn chế chính liên quan đến chất lượng biểu diễn đầu vào, mất cân bằng dữ liệu, độ ổn định của mô hình và khai thác chưa đầy đủ thông tin đa nguồn.	4
Hình 1.1	Minh họa biến thiên nội lớp, tương đồng liên lớp và sai khác nguồn thu nhận trong ảnh tổn thương da.	10
Hình 1.2	Minh họa kiến trúc VGG16 dưới dạng các block tích chập.	15
Hình 1.3	Bản đồ đặc trưng của ảnh tại lớp <code>block1_conv1</code>	16
Hình 1.4	Minh họa hợp nhất đặc trưng bằng phép cộng từng phần tử.	19
Hình 1.5	Minh họa hợp nhất đặc trưng bằng phép nhân từng phần tử.	19
Hình 1.6	Minh họa hợp nhất đặc trưng bằng phép ghép nối vector.	20
Hình 1.7	Minh họa pipeline học trên đồ thị cho bài toán phân loại ảnh tổn thương da. Các đặc trưng ảnh và metadata được sử dụng để xây dựng đồ thị, trong đó mỗi nút biểu diễn một mẫu và các cạnh phản ánh mức độ tương đồng giữa các mẫu. Mô hình GNN lan truyền thông tin giữa các nút lân cận nhằm học biểu diễn có tính ngữ cảnh, tạo cơ sở cho việc khai thác quan hệ giữa các mẫu trong các thiết lập học bán giám sát.	26
Hình 1.8	Đường cong ROC và AUC	30
Hình 2.1	Ý tưởng phương pháp đề xuất: chọn lọc bản đồ đặc trưng dựa trên entropy và hợp nhất đặc trưng đa nguồn.	39
Hình 2.2	Ví dụ các bản đồ đặc trưng được trích xuất từ các block khác nhau của mạng CNN.	40
Hình 2.3	Kiến trúc tổng quan của mô hình BlockNet.	41
Hình 2.4	Cấu trúc phân cấp các block tích chập của VGG16 được sử dụng để trích xuất bản đồ đặc trưng.	44
Hình 2.5	Ví dụ các bản đồ đặc trưng được trích xuất từ các lớp tích chập của VGG16. Mỗi dòng biểu diễn một lớp tích chập, gồm một số bản đồ đặc trưng đầu và cuối; dấu “...” biểu diễn các bản đồ đặc trưng trung gian không hiển thị.	45
Hình 2.6	Hai bản đồ đặc trưng minh họa cho quá trình tính entropy.	47
Hình 2.7	Minh họa một số bản đồ đặc trưng đầu và cuối được trích xuất từ lớp <code>block1_conv1</code> của VGG16 đối với một ảnh đầu vào.	49
Hình 2.8	Phân bố entropy của các bản đồ đặc trưng tại lớp <code>block1_conv1</code> và các ngưỡng dùng để xác định khoảng entropy phù hợp.	50
Hình 2.9	Ba bản đồ đặc trưng đại diện theo các mức entropy khác nhau: thấp, nằm trong khoảng phù hợp và cao.	50
Hình 2.10	Minh họa chọn lọc bản đồ đặc trưng dựa trên entropy tại lớp <code>block1_conv1</code> . Các bản đồ đặc trưng có viền xanh được giữ lại do nằm trong khoảng entropy phù hợp; các bản đồ đặc trưng có viền đỏ bị loại bỏ.	54
Hình 2.11	Các bản đồ đặc trưng hợp lệ được giữ lại từ lớp <code>block1_conv1</code>	56
Hình 2.12	Một số ảnh mẫu trong bộ dữ liệu VNCancer với các lớp tổn thương da khác nhau.	66

Hình 2.13 Một số ảnh mẫu trong bộ dữ liệu HAM10000 thuộc các lớp tổn thương da khác nhau.	67
Hình 2.14 So sánh ROC trên bộ dữ liệu VNCancer trước và sau khi sử dụng chọn lọc entropy.	75
Hình 2.15 So sánh ROC trên bộ dữ liệu HAM10000 trước và sau khi sử dụng chọn lọc entropy.	76
Hình 2.16 Động học huấn luyện của BlockNet trên bộ dữ liệu VNCancer. . .	77
Hình 2.17 Phân tích chi tiết cấu hình BlockNet cao nhất trên bộ dữ liệu VNCancer.	79
Hình 2.18 Động học huấn luyện của BlockNet trên bộ dữ liệu HAM10000. .	81
Hình 2.19 Phân tích chi tiết cấu hình BlockNet cao nhất trên bộ dữ liệu HAM10000.	82
 Hình 3.1 Bối cảnh và động cơ nghiên cứu.	 88
Hình 3.2 Sơ đồ minh họa quá trình biểu diễn dữ liệu đa nguồn, xây dựng đồ thị và học bán giám sát cho phân loại ảnh tổn thương da.	90
Hình 3.3 Cơ sở tổng quát của phương pháp đề xuất.	92
Hình 3.4 Tổng quan mô hình đề xuất.	94
Hình 3.5 Các nguồn đặc trưng đa nguồn và biểu diễn hợp nhất được sử dụng trong phương pháp đề xuất.	95
Hình 3.6 Hợp nhất đặc trưng đa nguồn theo cơ chế chú ý có xét metadata. .	97
Hình 3.7 Biểu diễn nút đa nguồn được xây dựng từ đặc trưng hợp nhất cho mô hình mạng nơ-ron đồ thị.	99
Hình 3.8 Đồ thị k-NN với $k = 5$ được xây dựng từ biểu diễn hợp nhất dựa trên độ tương đồng cosine.	100
Hình 3.9 Cơ chế lan truyền thông tin trong mô hình mạng nơ-ron đồ thị. . .	101
Hình 3.10 Pha 1: Huấn luyện có giám sát mô hình mạng nơ-ron đồ thị trên đồ thị k-NN.	103
Hình 3.11 Ví dụ quá trình sinh xác suất dự đoán và tính entropy cho các mẫu dữ liệu chưa gán nhãn.	105
Hình 3.12 Ví dụ xác định ngưỡng entropy và phân tầng độ tin cậy của dự đoán. .	107
Hình 3.13 Ví dụ tính entropy cho một mẫu dữ liệu chưa gán nhãn.	108
Hình 3.14 Ví dụ lựa chọn mẫu có độ tin cậy cao cho quá trình gán nhãn giả. .	109
Hình 3.15 Mở rộng tập huấn luyện sử dụng các mẫu nhãn giả thuộc nhóm độ tin cậy cao.	110
Hình 3.16 Ví dụ các mẫu tổn thương lành tính và ác tính trong bộ dữ liệu ISIC 2020.	114
Hình 3.17 Phân bố lớp lành tính và ác tính trong bộ dữ liệu ISIC 2020. . . .	115
Hình 3.18 Đánh giá mức độ của các nhóm đặc trưng bằng phân tích Chi-square. .	118
Hình 3.19 Ma trận nhầm lẫn chuẩn hóa của các mô hình học máy truyền thống ở pha huấn luyện có giám sát. Các mô hình này có Specificity cao nhưng còn bỏ sót mẫu ác tính ở một số cấu hình.	124
Hình 3.20 Ma trận nhầm lẫn chuẩn hóa của các kiến trúc nền GNN có tích hợp metadata ở pha huấn luyện có giám sát. Một số mô hình GNN làm thay đổi cân bằng giữa khả năng nhận diện lớp lành tính và ác tính. . . .	125
Hình 3.21 Diễn biến độ chính xác trên tập huấn luyện của các kiến trúc nền GNN ở pha huấn luyện có giám sát.	126

Hình 3.22	Diễn biến hàm mất mát trên tập huấn luyện của các kiến trúc nền GNN ở pha huấn luyện có giám sát.	126
Hình 3.23	Diễn biến độ chính xác trên tập xác thực của các kiến trúc nền GNN ở pha huấn luyện có giám sát.	126
Hình 3.24	Diễn biến hàm mất mát trên tập xác thực của các kiến trúc nền GNN ở pha huấn luyện có giám sát.	126
Hình 3.25	So sánh các chỉ số phân loại của mô hình GINE trước và sau pha gán nhãn giả dựa trên độ tin cậy.	127
Hình 3.26	Diễn biến độ chính xác (trái) và hàm mất mát (phải) của mô hình GINE ở pha huấn luyện có giám sát và pha gán nhãn giả dựa trên độ tin cậy.	128
Hình 3.27	Ma trận nhầm lẫn của mô hình GINE trước và sau pha gán nhãn giả dựa trên độ tin cậy. Pha hai làm thay đổi số lượng dự đoán sai ở lớp thiếu số và biên quyết định giữa hai lớp.	128
Hình 3.28	Phân bố số lượng mẫu theo ba mức độ tin cậy trên tập chưa gán nhãn D_U (ISIC 2020 Test, 10,982 mẫu).	129

DANH MỤC CÁC BẢNG

Bảng 1.1	So sánh hướng tiếp cận của BlockNet với các nhóm phương pháp gần nhất	17
Bảng 1.2	Ví dụ entropy của một số bản đồ đặc trưng của ảnh “ISIC_0024345.jpg” tại lớp block1_conv1 của VGG16	24
Bảng 1.3	Tổng hợp các hướng học đa nguồn, học trên đồ thị và khai thác dữ liệu chưa gán nhãn	32
Bảng 2.1	Các block tích chập của VGG16 được sử dụng để trích xuất bản đồ đặc trưng.	44
Bảng 2.2	Các lớp tổn thương da và phân bố số lượng trong bộ dữ liệu VNCancer	65
Bảng 2.3	Các lớp tổn thương da và phân bố số lượng trong bộ dữ liệu HAM10000	67
Bảng 2.4	Các mô hình trích rút đặc trưng được sử dụng trong thực nghiệm .	70
Bảng 2.5	Cấu hình thực nghiệm chính	72
Bảng 2.6	Kết quả trên bộ dữ liệu VNCancer khi không sử dụng chọn lọc entropy	73
Bảng 2.7	Kết quả trên bộ dữ liệu VNCancer khi sử dụng chọn lọc entropy .	74
Bảng 2.8	Kết quả trên bộ dữ liệu HAM10000 khi không sử dụng chọn lọc entropy	75
Bảng 2.9	Kết quả trên bộ dữ liệu HAM10000 khi sử dụng chọn lọc entropy .	76
Bảng 2.10	So sánh hiệu năng trên bộ dữ liệu VNCancer sử dụng kiểm định chéo 5 folds. Kết quả được trình bày dưới dạng trung bình $\pm$ độ lệch chuẩn (%).	78
Bảng 2.11	So sánh hiệu năng trên bộ dữ liệu HAM10000 sử dụng kiểm định chéo 5 folds. Kết quả được trình bày dưới dạng trung bình $\pm$ độ lệch chuẩn (%).	80
Bảng 2.12	So sánh giữa BlockNet và một số nghiên cứu gần đây trên HAM10000.	81
Bảng 3.1	Phân bố dữ liệu có nhãn và chưa gán nhãn trong ISIC 2020	115
Bảng 3.2	Thông kê metadata thiếu trong bộ dữ liệu ISIC 2020	116
Bảng 3.3	Ví dụ metadata lâm sàng thô trong bộ dữ liệu ISIC 2020	117
Bảng 3.4	Metadata lâm sàng sau chuẩn hóa và one-hot encoding	117
Bảng 3.5	Các nhóm đặc trưng được lựa chọn trong nghiên cứu	119
Bảng 3.6	Cấu hình môi trường thực nghiệm	121
Bảng 3.7	Cấu hình xây dựng đồ thị	121
Bảng 3.8	Các tham số huấn luyện của mô hình GNN	122
Bảng 3.9	Các tham số của mô hình học máy truyền thống	122
Bảng 3.10	So sánh kết quả pha huấn luyện có giám sát giữa các mô hình học máy truyền thống và các kiến trúc nền GNN trên bộ dữ liệu ISIC 2020. Các giá trị được báo cáo dưới dạng trung bình $\pm$ độ lệch chuẩn trên các lần kiểm chứng lặp phân tầng.	123
Bảng 3.11	So sánh kết quả của mô hình GINE trước và sau pha gán nhãn giả dựa trên độ tin cậy. Các giá trị được báo cáo dưới dạng trung bình $\pm$ độ lệch chuẩn trên các lần kiểm chứng.	127

Bảng 3.12	Kết quả kiểm định cặp giữa mô hình GINE ở pha huấn luyện có giám sát và mô hình GINE sau pha gán nhãn giả dựa trên độ tin cậy trên 15 lần kiểm chứng. Hiệu trung bình được tính theo dạng (Pha 2 – Pha 1).	129
Bảng 3.13	Phân bố số lượng mẫu theo mức độ tin cậy trên tập chưa gán nhãn D_U	130
Bảng 3.14	So sánh mô hình đề xuất với một số phương pháp liên quan trên ISIC 2020 theo chỉ số Accuracy	131

DANH MỤC CÁC ĐỊNH NGHĨA

1.1	Định nghĩa (Entropy dự đoán)	20
1.2	Định nghĩa (Entropy của bản đồ đặc trưng)	22
2.1	Định nghĩa (BlockNet [CT3])	37
2.2	Định nghĩa (Entropy của bản đồ đặc trưng [CT3])	46
2.3	Định nghĩa (Khoảng entropy phù hợp [CT3])	48
2.4	Định nghĩa (Bản đồ đặc trưng hợp lệ [CT3])	52
2.5	Định nghĩa (Trích rút đặc trưng từ BlockNet [CT3])	55
2.6	Định nghĩa (Hàm phân loại)	60
3.1	Định nghĩa (Đồ thị đa nguồn dữ liệu [CT4])	99
3.2	Định nghĩa (Entropy dự đoán [CT4])	103

MỞ ĐẦU

1. Tính cấp thiết của luận án

Trong những năm gần đây, trí tuệ nhân tạo, đặc biệt là học sâu, đã được ứng dụng ngày càng rộng rãi trong phân tích ảnh y tế và hỗ trợ chẩn đoán. Trong đó, phân loại ảnh tổn thương da là một bài toán có ý nghĩa khoa học và thực tiễn rõ rệt, do có khả năng hỗ trợ phát hiện sớm các tổn thương nghi ngờ, cung cấp thêm thông tin tham khảo cho bác sĩ và góp phần nâng cao hiệu quả sàng lọc bệnh da liễu. Việc phát hiện sớm các tổn thương nguy hiểm như ung thư hắc tố da (melanoma) có vai trò quan trọng trong nâng cao hiệu quả điều trị và giảm nguy cơ tử vong của người bệnh. Tuy nhiên, dữ liệu ảnh tổn thương da trong thực tế thường có sự biến thiên lớn về màu sắc, kết cấu, hình dạng và kích thước giữa các lớp cũng như trong cùng một lớp tổn thương. Nhiều loại tổn thương có biểu hiện thị giác tương đồng nhưng khác biệt về mức độ nguy hiểm, trong khi cùng một loại tổn thương lại có thể xuất hiện dưới nhiều dạng khác nhau do điều kiện thu nhận ảnh, thiết bị chụp hoặc đặc điểm sinh học của từng bệnh nhân. Bên cạnh đó, dữ liệu ảnh tổn thương da thường mất cân bằng lớp, trong đó số lượng mẫu lành tính lớn hơn đáng kể so với các lớp ác tính hoặc hiếm gặp. Dữ liệu gán nhãn trong lĩnh vực y tế cũng thường hạn chế do cần chuyên gia da liễu và chi phí gán nhãn cao, trong khi dữ liệu chưa gán nhãn có thể tồn tại với số lượng lớn. Những đặc điểm này làm cho bài toán phân loại ảnh tổn thương da vẫn còn nhiều thách thức và cần tiếp tục được nghiên cứu [1–3].

Ở trong nước, một số hướng nghiên cứu liên quan đã được triển khai theo các trục chuyên biệt như phân loại ảnh quy mô lớn, nâng cao chất lượng ảnh y sinh, độ tin cậy của mô hình học sâu và ứng dụng kỹ thuật hình ảnh trong da liễu [4–7]. Các nghiên cứu này cho thấy tiềm năng ứng dụng của thị giác máy tính và học sâu trong các bài toán ảnh y tế. Tuy nhiên, số lượng nghiên cứu tiếp cận bài toán phân loại ảnh tổn thương da theo một khung tích hợp, trong đó biểu diễn đặc trưng, hợp nhất đa nguồn, mô hình hóa quan hệ giữa các mẫu, khai thác dữ liệu chưa gán nhãn và lượng hóa độ tin cậy được xử lý thống nhất, vẫn còn hạn chế. Điều này cho thấy cần có một hướng nghiên cứu liên mạch hơn, vừa bám sát tiến bộ quốc tế, vừa phù hợp với đặc thù dữ liệu ảnh y tế và mục tiêu ứng dụng trong thực tế.

Một hướng nghiên cứu nền tảng trong phân loại ảnh tổn thương da là sử dụng các đặc trưng thủ công như màu sắc, kết cấu và hình dạng. Các đặc trưng này có ưu điểm là dễ diễn giải, phản ánh được một số dấu hiệu thị giác quen thuộc trong phân tích ảnh da liễu và từng được sử dụng rộng rãi trong các hệ thống phân loại tổn thương da [8–10]. Về mặt ý nghĩa miền, đặc trưng màu sắc có thể phản ánh sự phân bố sắc tố, đặc trưng kết cấu có thể mô tả bề mặt tổn thương, còn đặc trưng hình dạng có thể hỗ trợ nhận diện tính bất đối xứng hoặc mức độ không đều của vùng tổn thương. Tuy nhiên, các đặc trưng thủ công phụ thuộc nhiều vào kinh nghiệm thiết kế, khó bao quát hết các biến thiên phức tạp của ảnh y tế và thường suy giảm hiệu quả khi dữ liệu thay đổi về điều kiện chụp, nhiễu hoặc phân bố lớp. Vì vậy, mặc dù đặc trưng thủ công vẫn có giá trị bổ trợ, việc chỉ dựa vào nhóm đặc trưng này là chưa đủ để xây dựng một hệ phân loại ổn định.

Hướng nghiên cứu tiếp theo khai thác đặc trưng học sâu thông qua các mô hình CNN, học chuyển giao và các kiến trúc học sâu hiện đại. Các nghiên cứu theo hướng

này cho thấy đặc trưng học sâu có khả năng biểu diễn mạnh hơn so với đặc trưng thủ công, đặc biệt trong các bài toán có cấu trúc hình ảnh phức tạp [11–13]. Tuy nhiên, nhiều phương pháp vẫn sử dụng trực tiếp toàn bộ đặc trưng hoặc toàn bộ feature map được trích xuất từ mạng mà chưa có cơ chế đánh giá mức độ hữu ích của từng feature map đối với bài toán phân loại. Trong khi đó, không phải mọi feature map đều mang thông tin hữu ích; một số feature map có thể chứa nhiễu, thông tin dư thừa, phản hồi yếu hoặc phản hồi không ổn định. Nếu các thành phần này được giữ lại và đưa trực tiếp vào bước hợp nhất, chúng có thể làm suy giảm chất lượng và khả năng phân biệt của biểu diễn cuối cùng. Vì vậy, cần có cơ chế đánh giá và chọn lọc đặc trưng sâu ở mức biểu diễn trước khi đưa vào hợp nhất và phân loại.

Bên cạnh việc khai thác từng nguồn đặc trưng riêng lẻ, hợp nhất đặc trưng đa nguồn là một hướng tiếp cận quan trọng nhằm tận dụng tính bổ trợ giữa các dạng biểu diễn. Trong bài toán phân loại ảnh tổn thương da, đặc trưng học sâu có thể mô tả các mẫu hình phức tạp từ ảnh, trong khi đặc trưng thủ công có thể bổ sung các mô tả trực quan về màu sắc, kết cấu hoặc hình dạng [9, 10]. Ngoài ra, metadata lâm sàng như tuổi, giới tính và vị trí tổn thương có thể cung cấp thông tin ngữ cảnh hữu ích, góp phần cải thiện hiệu quả phân loại trong một số thiết lập [14–16]. Các nghiên cứu gần đây tiếp tục mở rộng hợp nhất ảnh và dữ liệu mô tả lâm sàng bằng attention, cross-attention và Transformer, cho thấy xu hướng chuyển từ ghép nối trực tiếp sang học tương tác giữa các nguồn [14, 17–19]. Tuy nhiên, nhiều phương pháp hợp nhất hiện nay chủ yếu giải quyết tương tác giữa các nguồn trong phạm vi từng mẫu, trong khi việc đồng thời lựa chọn biểu diễn đầu vào, mô hình hóa quan hệ liên mẫu và khai thác dữ liệu chưa gán nhãn có kiểm soát độ tin cậy vẫn chưa được nghiên cứu đầy đủ. Do đó, vấn đề quan trọng không chỉ là hợp nhất nhiều nguồn đặc trưng, mà là tổ chức, chọn lọc và hợp nhất các nguồn biểu diễn một cách có kiểm soát trong một quy trình học thống nhất.

Một vấn đề khác cần được chú ý là phần lớn các mô hình phân loại ảnh truyền thống xử lý từng ảnh như một mẫu độc lập. Trong thực tế, các ảnh tổn thương da có thể tồn tại quan hệ tương đồng về màu sắc, kết cấu, hình dạng hoặc đặc điểm lâm sàng. Việc khai thác quan hệ giữa các mẫu có thể bổ sung thêm thông tin cho quá trình phân loại, nhất là trong điều kiện dữ liệu có nhãn hạn chế. Các mô hình học trên đồ thị mở ra khả năng biểu diễn mỗi mẫu dữ liệu như một nút và khai thác quan hệ giữa các mẫu thông qua các cạnh của đồ thị [20–22]. Tuy nhiên, việc kết hợp học đa nguồn với học trên đồ thị trong bài toán phân loại ảnh tổn thương da vẫn cần được nghiên cứu sâu hơn, đặc biệt khi đồng thời xét đến đặc trưng học sâu, đặc trưng thủ công và metadata.

Ngoài dữ liệu có nhãn, dữ liệu chưa gán nhãn là một nguồn thông tin tiềm năng trong ảnh y tế. Các phương pháp học bán giám sát và gán nhãn giả đã được nghiên cứu nhằm tận dụng nguồn dữ liệu này [23–25]. Tuy nhiên, nếu mô hình dự đoán sai nhưng vẫn có độ tự tin cao, các nhãn giả sai có thể tiếp tục được đưa ngược vào quá trình huấn luyện và gây lan truyền sai lệch. Rủi ro này đặc biệt đáng chú ý trong dữ liệu y tế mất cân bằng, khi các lớp ác tính hoặc hiếm gặp thường khó nhận diện hơn. Vì vậy, việc khai thác dữ liệu chưa gán nhãn cần đi kèm cơ chế đánh giá độ tin cậy của dự đoán trước khi sử dụng cho tái huấn luyện.

Từ các phân tích trên, có thể thấy bài toán phân loại ảnh tổn thương da không chỉ cần một mô hình học sâu mạnh hơn, mà cần một cách tiếp cận toàn diện hơn đối với biểu diễn dữ liệu. Các vấn đề cốt lõi bao gồm: lựa chọn feature map có giá trị thông tin phù hợp, hợp nhất hiệu quả đặc trưng học sâu với đặc trưng thủ công và metadata, mô hình

hóa quan hệ giữa các mẫu, và khai thác dữ liệu chưa gán nhãn một cách đáng tin cậy. Trên cơ sở đó, luận án lựa chọn hướng nghiên cứu hợp nhất đặc trưng đa nguồn cho phân loại ảnh tổn thương da. Hướng tiếp cận của luận án không tập trung vào việc xây dựng một backbone học sâu hoàn toàn mới, mà tập trung vào cách tổ chức, chọn lọc, hợp nhất và khai thác biểu diễn dữ liệu trong điều kiện dữ liệu có biến thiên lớn, mất cân bằng lớp, hạn chế nhãn và tồn tại nhiều nguồn thông tin bổ trợ.

Các nghiên cứu ban đầu của Nghiên cứu sinh đã tiếp cận bài toán phân loại ảnh theo hai hướng hỗ trợ cho luận án. Công trình [CT1] khảo sát khả năng hợp nhất đặc trưng thủ công và đặc trưng học sâu trong phân loại ảnh, cho thấy các nhóm đặc trưng này có tính bổ trợ và có thể cải thiện hiệu quả phân loại khi được kết hợp phù hợp. Công trình [CT2] tiếp tục đánh giá các mô hình CNN và mô hình ensemble trên dữ liệu HAM10000, qua đó làm rõ vai trò của các backbone học sâu trong phân loại tổn thương da. Tuy nhiên, [CT1] và [CT2] chủ yếu đóng vai trò là các nghiên cứu nền tảng và định hướng. Các công trình này chưa giải quyết trực tiếp các vấn đề cốt lõi của luận án như lựa chọn feature map theo entropy (đại lượng đo mức độ phân tán của thông tin trong một phân bố), cân bằng biểu diễn đặc trưng, mô hình hóa quan hệ giữa các mẫu bằng đồ thị và khai thác dữ liệu chưa gán nhãn có kiểm soát độ tin cậy. Vì vậy, luận án tiếp tục phát triển hai hướng đóng góp chính được trình bày trong [CT3] và [CT4].

2. Khoảng trống nghiên cứu

Các kết quả nghiên cứu đã công bố cho thấy mạng học sâu có khả năng tạo ra biểu diễn mạnh cho ảnh tổn thương da. Tuy nhiên, tại mỗi tầng tích chập, mô hình thường sinh ra nhiều bản đồ đặc trưng, trong khi mức độ đóng góp của từng bản đồ đối với nhiệm vụ phân loại không đồng đều. Việc sử dụng toàn bộ các bản đồ đặc trưng mà không đánh giá trước có thể duy trì các thành phần dư thừa, phản hồi yếu hoặc không ổn định, từ đó làm giảm chất lượng biểu diễn đưa vào hợp nhất và phân loại. Entropy có thể hỗ trợ định lượng thông tin của bản đồ đặc trưng; tuy nhiên, entropy cao không mặc nhiên đồng nghĩa với khả năng phân biệt tốt, vì một bản đồ có entropy cao vẫn có thể chứa nhiễu. Do đó, việc lựa chọn cần dựa trên một khoảng entropy phù hợp và được kiểm chứng bằng hiệu quả phân loại, thay vì chỉ ưu tiên các giá trị entropy lớn.

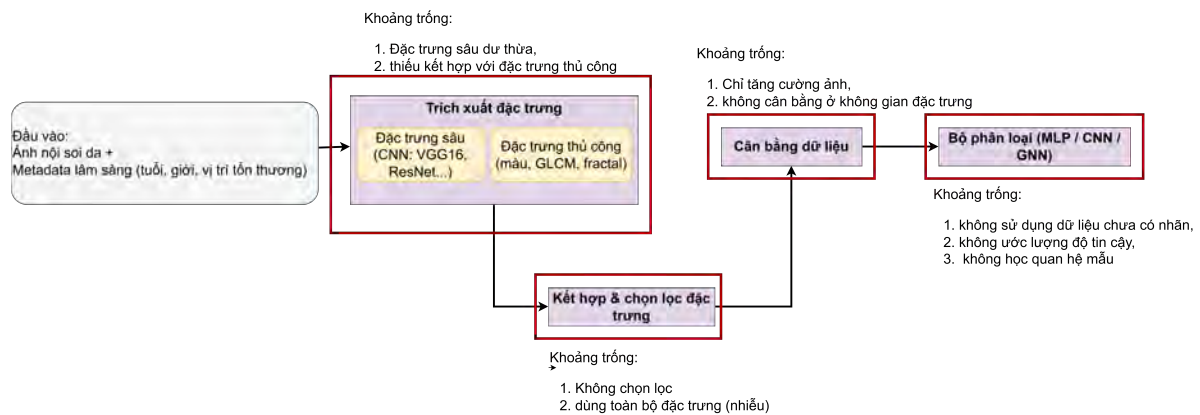
Sau bước chọn lọc, hiệu quả của biểu diễn còn phụ thuộc vào cách kết hợp các nguồn đặc trưng. Đặc trưng học sâu có khả năng mô tả các mẫu hình phức tạp, trong khi đặc trưng thủ công cung cấp các mô tả dựa trên màu sắc, kết cấu và hình dạng. Hai nhóm đặc trưng có tính bổ trợ, nhưng phép ghép nối trực tiếp chưa bảo đảm loại bỏ nhiễu, giảm dư thừa hoặc xử lý sự khác biệt về số chiều và phân bố giữa các nguồn. Đồng thời, mất cân bằng lớp có thể tiếp tục làm cho biểu diễn của các lớp hiếm chưa được học đầy đủ. Vì vậy, chọn lọc bản đồ đặc trưng, cân bằng biểu diễn và hợp nhất với đặc trưng thủ công cần được xem là các thành phần liên kết trong cùng một quy trình, thay vì những thao tác tách rời.

Ở mạch nghiên cứu mở rộng, ảnh tổn thương da không chỉ cung cấp thông tin thị giác mà còn có thể đi kèm dữ liệu mô tả lâm sàng như tuổi, giới tính và vị trí tổn thương. Các nguồn này khác nhau về bản chất nhưng có thể bổ trợ cho nhau. Tuy vậy, phần lớn mô hình vẫn xử lý mỗi ảnh như một mẫu độc lập nên chưa khai thác đầy đủ quan hệ tương đồng giữa các mẫu. Biểu diễn dữ liệu dưới dạng đồ thị, trong đó mỗi mẫu là một nút và quan hệ tương đồng giữa các mẫu là một cạnh, cho phép đưa thông tin liên mẫu vào quá trình học. Hiệu quả của cách tiếp cận này còn phụ thuộc vào việc tổ chức biểu

diễn đa nguồn và xây dựng quan hệ phù hợp, bởi các liên kết thiếu tin cậy có thể truyền thông tin sai giữa các mẫu.

Bên cạnh đó, học bán giám sát, phương thức học đồng thời từ dữ liệu có nhãn và dữ liệu chưa gán nhãn, là một hướng phù hợp khi nhãn chuyên gia khan hiếm. Gán nhãn giả, tức sử dụng dự đoán của mô hình làm nhãn tạm thời cho mẫu chưa gán nhãn, có thể mở rộng tập huấn luyện nhưng cũng tạo nguy cơ khuếch đại sai lệch nếu dự đoán không chính xác được tái sử dụng. Vì thế, quá trình này cần một cơ chế kiểm soát độ tin cậy, chẳng hạn sử dụng entropy dự đoán để chỉ lựa chọn những nhãn giả có mức bất định thấp trước khi tái huấn luyện.

Từ các lập luận trên, luận án xác định hai khoảng trống nghiên cứu có quan hệ kế tiếp nhau. Thứ nhất, còn thiếu một quy trình thống nhất có khả năng lựa chọn các bản đồ đặc trưng theo khoảng entropy phù hợp, sử dụng các biểu diễn đã chọn để hỗ trợ cân bằng lớp và hợp nhất chúng với đặc trưng thủ công nhằm nâng cao chất lượng biểu diễn cho phân loại ảnh tổn thương da. Khoảng trống này được luận án giải quyết bằng mô hình BlockNet và được kiểm chứng trong công trình [CT3]. Thứ hai, còn thiếu một khung học đa nguồn bán giám sát có khả năng kết hợp ảnh, đặc trưng thủ công và dữ liệu mô tả lâm sàng, đồng thời mô hình hóa quan hệ giữa các mẫu bằng đồ thị và khai thác dữ liệu chưa gán nhãn thông qua gán nhãn giả có kiểm soát độ tin cậy. Khoảng trống này được giải quyết trong hướng nghiên cứu của Chương 3 và được kiểm chứng trong công trình [CT4]. Hình 1 minh họa trực quan các khoảng trống nghiên cứu đó trong chuỗi xử lý của hệ thống phân loại ảnh tổn thương da.



Hình 1: Khoảng trống nghiên cứu của hệ thống phân loại ảnh tổn thương da. Các vùng khoanh đỏ thể hiện những hạn chế chính liên quan đến chất lượng biểu diễn đầu vào, mất cân bằng dữ liệu, độ ổn định của mô hình và khai thác chưa đầy đủ thông tin đa nguồn.

3. Mục tiêu nghiên cứu của luận án

Trên cơ sở hai khoảng trống nghiên cứu đã xác định, mục tiêu tổng quát của luận án là phát triển các phương pháp hợp nhất đặc trưng đa nguồn nhằm nâng cao hiệu quả phân loại ảnh tổn thương da trong điều kiện dữ liệu biến thiên lớn, mất cân bằng lớp, hạn chế nhãn và có nhiều nguồn thông tin bổ trợ.

Để đạt được mục tiêu tổng quát, luận án tập trung vào hai mục tiêu cụ thể:

Thứ nhất, đề xuất mô hình BlockNet có khả năng lựa chọn các bản đồ đặc trưng theo khoảng entropy phù hợp, hỗ trợ cân bằng lớp trong không gian đặc trưng và hợp

nhất đặc trưng học sâu với đặc trưng thủ công, qua đó nâng cao chất lượng biểu diễn và hiệu quả phân loại ảnh tổn thương da.

Thứ hai, phát triển khung học đa nguồn bán giám sát dựa trên đồ thị, kết hợp đặc trưng học sâu, đặc trưng thủ công và dữ liệu mô tả lâm sàng, đồng thời khai thác quan hệ giữa các mẫu và dữ liệu chưa gán nhãn bằng cơ chế gán nhãn giả có kiểm soát độ tin cậy, nhằm nâng cao hiệu quả phân loại và hạn chế sự lan truyền của nhãn sai.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của luận án là các phương pháp hợp nhất đặc trưng đa nguồn cho bài toán phân loại ảnh tổn thương da. Trong đó, luận án tập trung vào cách trích xuất, chọn lọc, tổ chức, hợp nhất và khai thác các nguồn thông tin khác nhau gồm đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng nhằm nâng cao hiệu quả phân loại.

Phạm vi nghiên cứu của luận án được xác định như sau:

- **Về bài toán nghiên cứu:** Luận án tập trung vào bài toán phân loại ảnh tổn thương da. Các bài toán khác như phân đoạn tổn thương, phát hiện vùng tổn thương, định vị tổn thương, truy vấn ảnh tương tự hoặc xây dựng hệ thống chẩn đoán lâm sàng hoàn chỉnh không thuộc phạm vi chính của luận án.
- **Về dữ liệu nghiên cứu:** Luận án sử dụng các bộ dữ liệu ảnh tổn thương da phục vụ đánh giá thực nghiệm, trong đó có các dữ liệu phổ biến như HAM10000, ISIC và dữ liệu VNCancer. Các bộ dữ liệu này được sử dụng nhằm kiểm chứng hiệu quả của phương pháp trong điều kiện dữ liệu có biến thiên hình thái, mất cân bằng lớp và khác biệt về nguồn thu nhận.
- **Về nguồn thông tin đầu vào:** Luận án xem xét ba nhóm thông tin chính: đặc trưng học sâu trích xuất từ ảnh, đặc trưng thủ công mô tả màu sắc, kết cấu, hình dạng, và metadata lâm sàng như tuổi, giới tính, vị trí tổn thương khi dữ liệu có cung cấp. Các nguồn dữ liệu phi cấu trúc như bệnh án văn bản tự do, dữ liệu xét nghiệm hoặc dữ liệu di truyền không thuộc phạm vi nghiên cứu.
- **Về phương pháp:** Luận án tập trung vào các kỹ thuật chọn lọc feature map theo entropy, tăng cường/cân bằng biểu diễn, hợp nhất đặc trưng, học đa nguồn, học trên đồ thị và học bán giám sát dựa trên gán nhãn giả có kiểm soát độ tin cậy. Luận án không đặt mục tiêu xây dựng một backbone học sâu hoàn toàn mới.
- **Về đánh giá:** Luận án đánh giá mô hình thông qua các chỉ số phân loại phù hợp với dữ liệu mất cân bằng như Accuracy, Precision, Recall, F1-score, AUC, Sensitivity và Specificity. Việc đánh giá tập trung vào hiệu năng phân loại, độ ổn định và khả năng khai thác dữ liệu chưa gán nhãn, không bao gồm thử nghiệm lâm sàng trên người bệnh.

Như vậy, phạm vi của luận án được giới hạn ở nghiên cứu phương pháp và kiểm chứng thực nghiệm cho bài toán phân loại ảnh tổn thương da, với trọng tâm là cải thiện chất lượng biểu diễn, khai thác dữ liệu đa nguồn và sử dụng dữ liệu chưa gán nhãn một cách đáng tin cậy.

5. Phương pháp nghiên cứu

Để đạt được các mục tiêu nghiên cứu đã đặt ra, luận án sử dụng kết hợp các phương pháp nghiên cứu phù hợp với đặc thù của lĩnh vực khoa học máy tính, bao gồm nghiên cứu lý thuyết, mô hình hóa thuật toán, thực nghiệm tính toán và phân tích đánh giá định lượng. Cách trình bày này nhằm làm rõ phương pháp nghiên cứu ở cấp độ học thuật, đồng thời phân biệt với các kỹ thuật cụ thể được sử dụng trong quá trình xây dựng mô hình.

Trước hết, luận án sử dụng phương pháp nghiên cứu lý thuyết thông qua khảo sát, phân tích và tổng hợp các công trình nghiên cứu liên quan đến phân loại ảnh tổn thương da, trích xuất đặc trưng thủ công, đặc trưng học sâu, hợp nhất đặc trưng, học đa nguồn, học trên đồ thị, học bán giám sát và đánh giá độ tin cậy của mô hình. Việc tổng hợp này nhằm xác định các hạn chế còn tồn tại và làm cơ sở hình thành các câu hỏi nghiên cứu của luận án.

Tiếp theo, luận án sử dụng phương pháp mô hình hóa và xây dựng thuật toán để phát triển các thành phần đề xuất. Ở hướng nghiên cứu thứ nhất, luận án xây dựng mô hình BlockNet dựa trên cơ chế chọn lọc feature map theo entropy, tăng cường/cân bằng biểu diễn và hợp nhất với đặc trưng thủ công. Ở hướng nghiên cứu thứ hai, luận án xây dựng khung học đa nguồn bán giám sát trên đồ thị, trong đó đặc trưng học sâu, đặc trưng thủ công và metadata được hợp nhất để tạo biểu diễn nút, sau đó các mẫu được liên kết thành đồ thị và huấn luyện bằng các mô hình mạng nơ-ron đồ thị.

Luận án sử dụng phương pháp thực nghiệm tính toán để kiểm chứng hiệu quả của các mô hình đề xuất. Các thực nghiệm được tiến hành trên các bộ dữ liệu ảnh tổn thương da, với các thiết lập so sánh giữa mô hình đề xuất và các mô hình cơ sở. Quá trình thực nghiệm bao gồm tiền xử lý dữ liệu, trích xuất đặc trưng, xây dựng tập huấn luyện và tập kiểm tra, huấn luyện mô hình, đánh giá kết quả và phân tích lỗi.

Đối với dữ liệu chưa gán nhãn, luận án triển khai quy trình học bán giám sát dựa trên gán nhãn giả có chọn lọc. Mô hình được huấn luyện ban đầu trên dữ liệu có nhãn, sau đó dự đoán nhãn cho dữ liệu chưa gán nhãn. Entropy của phân phối dự đoán được sử dụng để đánh giá độ tin cậy; các mẫu có entropy thấp, tương ứng với độ tin cậy cao, được lựa chọn để bổ sung vào tập huấn luyện trong pha tái huấn luyện.

Cuối cùng, luận án sử dụng phương pháp phân tích định lượng và so sánh thực nghiệm để đánh giá hiệu quả mô hình. Các chỉ số như Accuracy, Precision, Recall, F1-score, AUC, Sensitivity và Specificity được sử dụng nhằm phản ánh hiệu năng phân loại trong điều kiện dữ liệu mất cân bằng. Ngoài ra, luận án phân tích ma trận nhầm lẫn, sự khác biệt giữa các lớp, ảnh hưởng của entropy, hiệu quả của gán nhãn giả và vai trò của các nguồn đặc trưng để làm rõ ưu điểm cũng như hạn chế của các phương pháp đề xuất.

Như vậy, phương pháp nghiên cứu của luận án kết hợp giữa khảo sát lý thuyết, mô hình hóa thuật toán, thực nghiệm trên dữ liệu thực và phân tích đánh giá kết quả, nhằm bảo đảm các đóng góp của luận án được kiểm chứng một cách hệ thống và phù hợp với bài toán phân loại ảnh tổn thương da.

6. Đóng góp của luận án

Luận án có hai đóng góp chính tương ứng với hai mục tiêu nghiên cứu:

Thứ nhất, luận án đề xuất mô hình BlockNet lựa chọn các bản đồ đặc trưng theo khoảng entropy phù hợp, sử dụng các biểu diễn được chọn để hỗ trợ cân bằng lớp trong

không gian đặc trưng và hợp nhất chúng với đặc trưng thủ công. Mô hình góp phần giảm ảnh hưởng của các phản hồi dư thừa, yếu hoặc không ổn định, đồng thời khai thác tính bổ trợ giữa đặc trưng học sâu và tri thức miền để nâng cao chất lượng biểu diễn cho phân loại ảnh tổn thương da. Đóng góp này được trình bày và kiểm chứng trong [CT3].

Thứ hai, luận án phát triển khung học đa nguồn bán giám sát dựa trên đồ thị, trong đó đặc trưng học sâu, đặc trưng thủ công và dữ liệu mô tả lâm sàng được hợp nhất để biểu diễn mẫu; quan hệ tương đồng giữa các mẫu được mô hình hóa bằng đồ thị; dữ liệu chưa gán nhãn được khai thác bằng cơ chế gán nhãn giả có kiểm soát độ tin cậy dựa trên entropy dự đoán. Khung đề xuất cho phép khai thác đồng thời thông tin đa nguồn, quan hệ liên mẫu và dữ liệu chưa gán nhãn, đồng thời hạn chế nguy cơ lan truyền nhãn sai trong quá trình tái huấn luyện. Đóng góp này được trình bày và kiểm chứng trong [CT4].

7. Cấu trúc của luận án

Ngoài phần Mở đầu và Kết luận, luận án gồm ba chương nghiên cứu chính, được tổ chức theo trình tự từ tổng quan cơ sở lý thuyết đến các phương pháp đề xuất và thực nghiệm đánh giá.

Chương 1 trình bày tổng quan về bài toán phân loại ảnh tổn thương da và các cơ sở lý thuyết liên quan. Nội dung chương tập trung vào các dạng biểu diễn đặc trưng ảnh, bao gồm đặc trưng thủ công, đặc trưng học sâu và metadata; các phương pháp hợp nhất dữ liệu đa nguồn; vai trò của entropy trong học biểu diễn và đánh giá độ tin cậy; các mô hình học trên đồ thị và học bán giám sát; cùng với các chỉ số đánh giá mô hình trong bối cảnh dữ liệu mất cân bằng.

Chương 2 đề xuất mô hình BlockNet dựa trên cơ chế chọn lọc feature map theo entropy và hợp nhất đặc trưng cho phân loại ảnh tổn thương da. Chương này trình bày phương pháp lựa chọn các feature map có giá trị thông tin phù hợp, sử dụng các feature map được chọn để tăng cường và cân bằng biểu diễn, sau đó kết hợp với đặc trưng thủ công nhằm nâng cao chất lượng biểu diễn đầu vào và cải thiện hiệu năng phân loại.

Chương 3 đề xuất mô hình học đa nguồn bán giám sát dựa trên đồ thị với cơ chế đánh giá độ tin cậy dựa trên entropy cho phân loại ảnh tổn thương da. Nội dung chương tập trung vào việc kết hợp đặc trưng học sâu, đặc trưng thủ công và metadata trong một không gian biểu diễn chung, xây dựng đồ thị quan hệ giữa các mẫu và khai thác dữ liệu chưa gán nhãn thông qua gán nhãn giả có kiểm soát độ tin cậy để tái huấn luyện mô hình.

Phần Kết luận tổng hợp các kết quả đạt được của luận án, chỉ ra những hạn chế còn tồn tại và đề xuất các hướng nghiên cứu tiếp theo.

CHƯƠNG 1: TỔNG QUAN VỀ PHÂN LOẠI TỔN THƯƠNG DA VÀ CƠ SỞ LÝ THUYẾT

Mở đầu

Chương này trình bày tổng quan về bài toán phân loại ảnh tổn thương da và cơ sở lý thuyết cho các phương pháp đề xuất. Nội dung tập trung vào biểu diễn ảnh, hợp nhất đặc trưng (feature fusion), entropy, học trên đồ thị, học bán giám sát và các độ đo đánh giá trong bối cảnh dữ liệu mất cân bằng.

Các nội dung được tổ chức theo hai hướng nghiên cứu: chọn lọc và hợp nhất đặc trưng ở mức biểu diễn; học đa nguồn (multimodal learning) bán giám sát trên đồ thị có kiểm soát độ tin cậy. Từ lần xuất hiện tiếp theo, luận án thống nhất sử dụng các thuật ngữ tiếng Việt đã được xác lập.

1.1. Tổng quan bài toán phân loại ảnh tổn thương da

Ảnh tổn thương da là ảnh ghi nhận vùng da có biểu hiện bất thường về màu sắc, hình dạng, kết cấu hoặc cấu trúc bề mặt. Trong luận án, hai dạng ảnh được xem xét gồm ảnh lâm sàng chụp bằng điện thoại và ảnh soi da (dermoscopy), tức ảnh được thu nhận bằng thiết bị chuyên dụng có khả năng quan sát rõ hơn các cấu trúc dưới bề mặt da. So với ảnh tự nhiên, đối tượng cần nhận dạng trong ảnh tổn thương da thường không có hình dạng cố định, ranh giới với vùng da lành có thể không rõ và dấu hiệu phân biệt giữa các bệnh có thể chỉ nằm ở những biến đổi nhỏ về sắc tố, đường biên hoặc kết cấu.

Phân loại ảnh tổn thương da là bài toán gán mỗi ảnh đầu vào vào một lớp bệnh lý hoặc một nhóm chẩn đoán xác định trước. Trong học có giám sát, tức cách học mô hình từ các mẫu đã có nhãn, mô hình được huấn luyện trên tập ảnh đã được chuyên gia xác định lớp và sau đó dự đoán nhãn cho ảnh mới. Các lớp có thể gồm u hắc tố ác tính (Melanoma, MEL), ung thư biểu mô tế bào đáy (Basal Cell Carcinoma, BCC), dày sừng ánh sáng và ung thư biểu mô tế bào trong da (Actinic Keratoses and Intraepithelial Carcinoma, AKIEC), tổn thương dạng dày sừng lành tính (Benign Keratosis-like Lesions, BKL) cùng các lớp khác tùy theo bộ dữ liệu. Bài toán đã được nghiên cứu trên các bộ dữ liệu như HAM10000 và ISIC, đồng thời là một hướng ứng dụng của học sâu trong hỗ trợ chẩn đoán da liễu [1, 3, 26, 27].

Về hình thức, giả sử tập dữ liệu có nhãn được ký hiệu theo Công thức (1.1)

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N, \quad (1.1)$$

trong đó x_i là ảnh tổn thương da thứ i , $y_i \in \mathcal{Y}$ là nhãn tương ứng, $\mathcal{Y} = \{1, 2, \dots, C\}$ là tập nhãn gồm C lớp, và N là số lượng mẫu có nhãn.

Mục tiêu của mô hình phân loại là học một ánh xạ như trong Công thức (1.2)

$$f_{\theta} : \mathcal{X} \rightarrow \mathcal{Y}, \quad (1.2)$$

trong đó $\mathcal{X}$ là không gian ảnh đầu vào và θ là tập tham số của mô hình.

Trong các mô hình học sâu, đầu ra thường là một phân phối xác suất trên các lớp theo Công thức (1.3).

$$p_{\theta}(y | x) = [p_{\theta}(1 | x), p_{\theta}(2 | x), \dots, p_{\theta}(C | x)], \quad (1.3)$$

với $p_\theta(c | x)$ là xác suất mẫu x thuộc lớp c . Nhân dự đoán được xác định theo Công thức (1.4).

$$\hat{y} = \arg \max_{c \in \mathcal{Y}} p_\theta(c | x). \quad (1.4)$$

Trong bài toán phân loại ảnh tổn thương da, dữ liệu đầu vào có thể gồm ảnh và các nguồn thông tin bổ trợ như đặc trưng thủ công hoặc dữ liệu mô tả lâm sàng (metadata). Do đó, một mẫu dữ liệu có thể được mô tả dưới dạng đa nguồn theo Công thức (1.5).

$$x_i = \left(x_i^{(img)}, x_i^{(h)}, x_i^{(m)} \right), \quad (1.5)$$

trong đó $x_i^{(img)}$ biểu diễn thông tin ảnh, $x_i^{(h)}$ biểu diễn các đặc trưng thủ công như màu sắc, kết cấu hoặc hình dạng, và $x_i^{(m)}$ biểu diễn dữ liệu mô tả lâm sàng như tuổi, giới tính hoặc vị trí tổn thương nếu dữ liệu có cung cấp. Nguồn thông tin này cung cấp ngữ cảnh bổ trợ cho các biểu hiện thị giác trên ảnh [14, 28, 29].

Bài toán có một số đặc thù so với phân loại ảnh tự nhiên. Biến thiên nội lớp, tức mức khác biệt lớn giữa các ảnh cùng một lớp, thể hiện qua sự thay đổi về kích thước, màu sắc, hình dạng, kết cấu và điều kiện thu nhận. Ngược lại, tương đồng liên lớp, tức sự giống nhau về biểu hiện thị giác giữa các lớp khác nhau, khiến ranh giới phân loại trở nên không rõ. Dữ liệu còn thường mất cân bằng lớp, nghĩa là số mẫu giữa các lớp chênh lệch lớn. Tóc hoặc lông, phản xạ ánh sáng, vùng nền, màu da, thiết bị và khoảng cách chụp có thể che khuất hoặc làm thay đổi dấu hiệu bệnh lý. Việc xác lập nhãn cần chuyên gia y tế nên dữ liệu có nhãn thường hạn chế, trong khi dữ liệu chưa gán nhãn có thể nhiều hơn [30–34].

Sáu ảnh trong Hình 1.1 minh họa các khó khăn trên. Ba ảnh BCC khác nhau rõ rệt dù cùng nhãn; trong đó ảnh VNCancer do bác sĩ chụp bằng điện thoại khác về trường nhìn, nền ảnh và điều kiện chiếu sáng so với hai ảnh soi da của HAM10000. Ba ảnh AKIEC, BKL và MEL lại có một số biểu hiện gần nhau về vùng sắc tố, đường biên không đều hoặc kết cấu bề mặt. Hình này chỉ minh họa đặc điểm thị giác và sai khác nguồn thu nhận, không thay thế kết luận chẩn đoán của bác sĩ.

Những đặc thù này đặt ra yêu cầu mô hình phải học được biểu diễn có khả năng phân biệt, khai thác hợp lý thông tin bổ trợ và xử lý tình trạng thiếu nhãn. Vì vậy, hướng thứ nhất của luận án sử dụng entropy kích hoạt như một tiêu chí sàng lọc không giám sát đối với các bản đồ đặc trưng trước khi hợp nhất với đặc trưng thủ công; hiệu quả phân loại của tập bản đồ được chọn được đánh giá bằng thực nghiệm. Hướng thứ hai biểu diễn quan hệ giữa các mẫu bằng đồ thị, hợp nhất nhiều nguồn dữ liệu và khai thác dữ liệu chưa gán nhãn bằng cơ chế gán nhãn giả có kiểm soát độ tin cậy. Các phần tiếp theo trình bày cơ sở lý thuyết tương ứng cho hai hướng nghiên cứu này.

1.2. Biểu diễn ảnh và các loại đặc trưng

1.2.1. Lựa chọn đặc trưng có giám sát bằng tiêu chí Chi-square

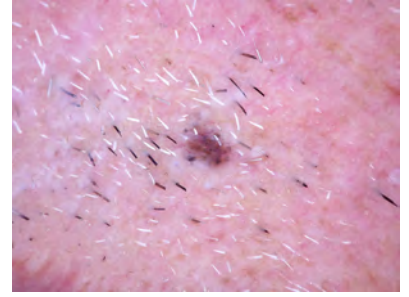
Chi-square là một tiêu chí thống kê có giám sát dùng để đánh giá mức độ phụ thuộc giữa đặc trưng và nhãn lớp [35]. Với một đặc trưng đã được biểu diễn phù hợp và nhãn lớp, dữ liệu được tổ chức theo các giá trị quan sát O_{rc} và các giá trị kỳ vọng E_{rc} ứng với



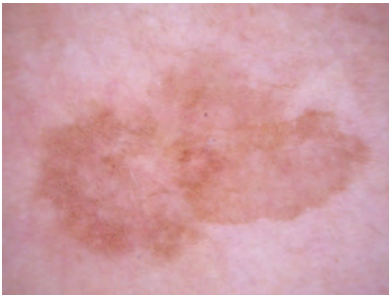
(a) BCC, VNCancer, ảnh điện thoại



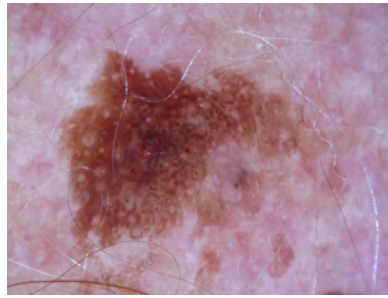
(b) BCC, HAM10000, ISIC_0024345



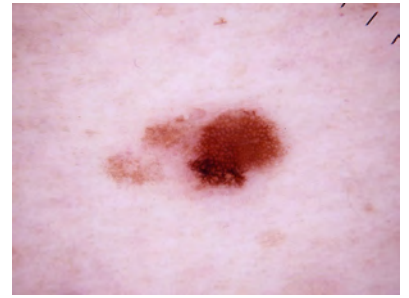
(c) BCC, HAM10000, ISIC_0024411



(d) AKIEC, HAM10000, ISIC_0033869



(e) BKL, HAM10000, ISIC_0034235



(f) MEL, HAM10000, ISIC_0024313

Hình 1.1: Minh họa biến thiên nội lớp, tương đồng liên lớp và sai khác nguồn thu nhận trong ảnh tổn thương da.

mức r của đặc trưng và lớp c . Thống kê Chi-square được xác định bởi

$$\chi^2 = \sum_r \sum_c \frac{(O_{rc} - E_{rc})^2}{E_{rc}}, \quad (1.6)$$

trong đó E_{rc} là tần suất kỳ vọng khi đặc trưng và nhãn độc lập. Giá trị χ^2 lớn cho thấy phân bố của đặc trưng giữa các lớp khác biệt rõ hơn, do đó đặc trưng có mức liên quan thống kê cao hơn với nhãn. Tiêu chí này cung cấp thứ hạng về mức liên quan, nhưng không tự khẳng định quan hệ nhân quả hoặc bảo đảm rằng đặc trưng có điểm cao sẽ luôn cải thiện mô hình khi kết hợp với các đặc trưng khác.

Khi áp dụng cho lựa chọn đặc trưng, các giá trị đầu vào cần có biểu diễn phù hợp với phép kiểm định, chẳng hạn các đại lượng không âm hoặc dữ liệu được phân khoảng trước khi lập bảng tần suất. Việc xác định cách biểu diễn phải được thực hiện nhất quán giữa các tập dữ liệu. Đồng thời, điểm Chi-square và quyết định lựa chọn đặc trưng chỉ được ước lượng từ tập huấn luyện; quy tắc đã xác định sau đó mới được áp dụng cho tập kiểm định và tập kiểm tra. Nguyên tắc này ngăn thông tin nhãn ngoài tập huấn luyện tham gia vào quá trình lựa chọn và hạn chế rò rỉ dữ liệu.

Trong luận án, Chi-square được sử dụng ở Chương 3 để so sánh mức liên quan của các nhóm đặc trưng thị giác ứng viên với nhãn phân loại. Điểm của từng thành phần được tổng hợp theo nhóm để hỗ trợ xếp hạng các loại đặc trưng. Metadata lâm sàng được xử

lý ở một nhánh riêng và không thuộc bước xếp hạng các nhóm đặc trưng thị giác bằng Chi-square. Chi tiết áp dụng và các nhóm đặc trưng được lựa chọn được trình bày tại Mục 3.5.3.

1.2.2. Đặc trưng thủ công: màu sắc, kết cấu, hình dạng

Đặc trưng thủ công là các đại lượng được thiết kế dựa trên tri thức miền để mô tả trực tiếp màu sắc, kết cấu và hình dạng của ảnh [8–10]. Đặc trưng màu phản ánh phân bố sắc tố; đặc trưng kết cấu mô tả quan hệ cường độ và cấu trúc bề mặt; đặc trưng hình dạng phản ánh hướng cạnh, đường biên và mức độ không đều của tổn thương [36–39]. Trong luận án, các đặc trưng này cung cấp thông tin bổ trợ cho đặc trưng học sâu, không được xem là phương án thay thế đặc trưng học sâu.

Với ảnh gồm N điểm ảnh, histogram chuẩn hóa của kênh màu c tại khoảng thứ b được xác định bởi

$$h_{c,b} = \frac{1}{N} \sum_{n=1}^N \mathbb{I}(q_c(x_n) = b), \quad b = 1, \dots, B_c, \quad (1.7)$$

trong đó $q_c(x_n)$ là chỉ số khoảng sau khi lượng tử hóa giá trị kênh c của điểm ảnh x_n , B_c là số khoảng của kênh và $\mathbb{I}(\cdot)$ là hàm chỉ thị. Véc-tơ histogram màu được tạo bằng cách ghép histogram của các kênh:

$$\mathbf{h} = [\mathbf{h}_1; \mathbf{h}_2; \dots; \mathbf{h}_C], \quad d_{\text{hist}} = \sum_{c=1}^C B_c. \quad (1.8)$$

Nếu xây dựng riêng histogram cho từng kênh rồi ghép nối thì số chiều được xác định bằng tổng số bin của các kênh. Nếu xây dựng histogram liên hợp nhiều chiều thì số chiều được xác định bằng số tổ hợp bin của các kênh [37, 38].

Đối với histogram HSV, sắc độ H phản ánh loại màu, độ bão hòa S phản ánh độ thuần của màu và độ sáng V phản ánh mức sáng. Ba thành phần được lượng tử hóa đồng thời để tạo histogram liên hợp ba chiều [38, 40]. Gọi q_H , q_S và q_V lần lượt là các hàm lượng tử hóa kênh H , S và V . Phần tử histogram tại ô (a, b, c) được xác định bởi

$$h_{a,b,c}^{\text{HSV}} = \frac{1}{N} \sum_{n=1}^N \mathbb{I}(q_H(H_n) = a, q_S(S_n) = b, q_V(V_n) = c), \quad B_H = 8, \quad B_S = 2, \quad B_V = 2. \quad (1.9)$$

Sau khi trải phẳng histogram ba chiều theo một thứ tự cố định, mỗi tổ hợp khoảng (a, b, c) tạo một thành phần. Do đó, số chiều của véc-tơ histogram HSV là

$$d_{\text{HSV}} = B_H B_S B_V = 8 \times 2 \times 2 = 32. \quad (1.10)$$

Việc dành nhiều mức lượng tử hơn cho kênh H giúp giữ chi tiết biến đổi sắc độ; hai mức của S và V biểu diễn mức thấp và mức cao của độ bão hòa và độ sáng. Mỗi tổ hợp (a, b, c) là một bin màu của histogram liên hợp HSV. Vì vậy, số chiều được tính bằng tích số bin của ba kênh.

Đặc trưng kết cấu được tính từ ma trận đồng xuất hiện mức xám GLCM. Với khoảng cách δ và hướng θ , phần tử $P_{\delta,\theta}(i, j)$ biểu diễn xác suất mức xám i xuất hiện đồng thời với mức xám j . Bốn đại lượng được sử dụng gồm

$$\text{Contrast} = \sum_{i,j} (i - j)^2 P_{\delta,\theta}(i, j), \quad (1.11)$$

$$\text{Correlation} = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)}{\sigma_i \sigma_j} P_{\delta, \theta}(i, j), \quad (1.12)$$

$$\text{Energy} = \sum_{i,j} P_{\delta, \theta}(i, j)^2, \quad (1.13)$$

$$\text{Homogeneity} = \sum_{i,j} \frac{P_{\delta, \theta}(i, j)}{1 + |i - j|}. \quad (1.14)$$

Khi mỗi đại lượng được lấy trung bình trên các hướng khảo sát, véc-tơ GLCM gồm bốn thành phần, do đó

$$d_{\text{GLCM}} = 4. \quad (1.15)$$

Đặc trưng HOG mô tả phân bố hướng gradient [39]. Giả sử ảnh có kích thước $W \times H$, mỗi ô có kích thước $p_x \times p_y$, mỗi khối gồm $b_x \times b_y$ ô, các khối dịch chuyển với bước một ô và histogram gồm K hướng. Số khối theo hai chiều là

$$N_x = \left\lfloor \frac{W}{p_x} \right\rfloor - b_x + 1, \quad N_y = \left\lfloor \frac{H}{p_y} \right\rfloor - b_y + 1, \quad (1.16)$$

và số chiều HOG trước bước rút gọn hoặc lựa chọn là

$$d_{\text{HOG}} = N_x N_y b_x b_y K. \quad (1.17)$$

Đặc trưng fractal mô tả mức độ phức tạp và không đều của cấu trúc tổn thương [41, 42]. Theo phương pháp đếm hộp, gọi $N(\varepsilon)$ là số hộp cạnh ε cần để phủ cấu trúc đang xét, chiều fractal được xác định bởi

$$D_f = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}. \quad (1.18)$$

Để biểu diễn sự không đều của toàn bộ đường biên bằng một véc-tơ có độ dài cố định, đường biên tổn thương được tham số hóa theo chiều dài cung và nội suy thành Q điểm cách đều $\{p_q\}_{q=1}^Q$. Tại mỗi điểm p_q , xét một lân cận đường biên $\mathcal{C}_q$ và đếm số hộp $N_q(\varepsilon_s)$ giao với lân cận này tại các tỉ lệ ε_s , $s = 1, \dots, S$. Thành phần thứ q của chữ ký fractal là hệ số góc ước lượng bằng bình phương tối thiểu:

$$D_q = \frac{\sum_{s=1}^S (u_s - \bar{u})(v_{q,s} - \bar{v}_q)}{\sum_{s=1}^S (u_s - \bar{u})^2}, \quad u_s = \log(1/\varepsilon_s), \quad v_{q,s} = \log N_q(\varepsilon_s). \quad (1.19)$$

Các giá trị D_q mô tả mức độ phức tạp cục bộ quanh các vị trí liên tiếp trên đường biên. Véc-tơ fractal được xác định bởi

$$\mathbf{x}^{(f)} = [D_1, D_2, \dots, D_Q], \quad d_f = Q. \quad (1.20)$$

Khi đường biên được chuẩn hóa thành $Q = 128$ điểm, mỗi điểm tạo một thành phần fractal cục bộ nên véc-tơ mô tả hình dạng có $d_f = 128$ chiều. Giá trị này được viết tường minh dưới dạng

$$d_f = Q = 128. \quad (1.21)$$

Đặc trưng LBP mô tả cấu trúc vi mô quanh mỗi điểm ảnh [43]. Với P điểm lân cận

trên bán kính R , mã LBP tại vị trí (u, v) được xác định bởi

$$\text{LBP}_{P,R}(u, v) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \quad s(a) = \begin{cases} 1, & a \geq 0, \\ 0, & a < 0, \end{cases} \quad (1.22)$$

trong đó g_c là mức xám của điểm trung tâm và g_p là mức xám của điểm lân cận thứ p . Histogram chuẩn hóa của các mã LBP tạo véc-tơ đặc trưng; nếu histogram gồm B_{LBP} khoảng thì $d_{\text{LBP}} = B_{\text{LBP}}$.

Bộ lọc Gabor ở tần số f và hướng θ được biểu diễn bởi [44]

$$g_{f,\theta}(x, y) = \exp\left(-\frac{x_\theta^2 + \gamma^2 y_\theta^2}{2\sigma^2}\right) \cos(2\pi f x_\theta + \phi), \quad (1.23)$$

trong đó $x_\theta = x \cos \theta + y \sin \theta$ và $y_\theta = -x \sin \theta + y \cos \theta$. Với mỗi cặp (f, θ) , giá trị trung bình và phương sai của đáp ứng lọc được giữ lại. Nếu sử dụng N_f tần số và N_θ hướng thì

$$d_{\text{Gabor}} = 2N_f N_\theta. \quad (1.24)$$

Với ba tần số và bốn hướng, $d_{\text{Gabor}} = 2 \times 3 \times 4 = 24$.

Đặc trưng wavelet được tạo bằng biến đổi wavelet rời rạc hai chiều [45]. Với mỗi nhóm hệ số W_k , véc-tơ giữ lại giá trị trung bình và phương sai:

$$\mu_k = \frac{1}{|W_k|} \sum_{w \in W_k} w, \quad \sigma_k^2 = \frac{1}{|W_k|} \sum_{w \in W_k} (w - \mu_k)^2. \quad (1.25)$$

Nếu có K_w nhóm hệ số được tổng hợp thì $d_{\text{wavelet}} = 2K_w$.

Đặc trưng biên được tính sau khi phát hiện biên Canny và xác định các đường bao [46]. Gọi L_r là độ dài đường bao thứ r , $r = 1, \dots, N_c$. Hai thành phần được sử dụng là

$$\mu_L = \frac{1}{N_c} \sum_{r=1}^{N_c} L_r, \quad \sigma_L = \sqrt{\frac{1}{N_c} \sum_{r=1}^{N_c} (L_r - \mu_L)^2}, \quad (1.26)$$

do đó véc-tơ $[\mu_L, \sigma_L]$ có $d_{\text{edge}} = 2$ chiều.

Đặc trưng tương quan màu đo quan hệ tuyến tính giữa từng cặp kênh RGB bằng hệ số tương quan Pearson [38, 47]:

$$\rho_{ab} = \frac{\sum_{n=1}^N (a_n - \bar{a})(b_n - \bar{b})}{\sqrt{\sum_{n=1}^N (a_n - \bar{a})^2} \sqrt{\sum_{n=1}^N (b_n - \bar{b})^2}}. \quad (1.27)$$

Ba cặp kênh (R, G) , (R, B) và (G, B) tạo véc-tơ $[\rho_{RG}, \rho_{RB}, \rho_{GB}]$ có $d_{\text{corr}} = 3$ chiều.

Sau khi hình thành, các thành phần đặc trưng có thể được đánh giá bổ trợ bằng Chi-square theo Mục 1.2.1. Công thức và tham số trích rút quyết định số chiều ban đầu, còn Chi-square chỉ đánh giá mức liên quan với nhãn. Khi nhiều bộ mô tả cùng phản ánh một nhóm thông tin, các thành phần được chọn hoặc rút gọn được ghép thành ba nhóm màu, kết cấu và hình dạng:

$$\mathbf{x}^{(h)} = [\mathbf{x}^{(c)}; \mathbf{x}^{(t)}; \mathbf{x}^{(s)}], \quad d_{\text{handcrafted}} = d_c + d_t + d_s. \quad (1.28)$$

Các giá trị d_c , d_t và d_s được xác định bởi loại bộ mô tả, tham số trích rút và phương pháp rút gọn hoặc lựa chọn đặc trưng; chúng không được suy ra từ điểm Chi-square. Chẳng

hạn, Công thức (1.10) xác định véc-tơ màu HSV có 32 chiều, còn Công thức (1.15) xác định véc-tơ kết cấu GLCM có 4 chiều. Histogram màu phản ánh phân bố cường độ hoặc sắc độ nhưng không biểu diễn đầy đủ kết cấu và hình dạng của đối tượng. Vì vậy, các nhóm màu sắc, kết cấu và hình dạng thường được kết hợp để tạo biểu diễn bổ trợ và đầy đủ hơn [48, 49].

1.2.3. Đặc trưng học sâu trong biểu diễn ảnh tổn thương da

Đặc trưng học sâu là các đặc trưng được học tự động từ dữ liệu thông qua các mô hình mạng nơ-ron, đặc biệt là mạng nơ-ron tích chập (CNN). Khác với đặc trưng thủ công, đặc trưng học sâu không cần thiết kế trực tiếp từng thuộc tính thị giác, mà được hình thành thông qua quá trình lan truyền ảnh qua nhiều tầng tích chập. Các tầng đầu của CNN thường học các đặc trưng mức thấp như biên, hướng cạnh, vùng màu và kết cấu đơn giản; trong khi các tầng sâu hơn có xu hướng học các mẫu hình phức tạp hơn liên quan đến cấu trúc tổn thương [1, 50–53].

Sự phát triển của các kiến trúc CNN hiện đại như VGGNet, ResNet, DenseNet và EfficientNet đã giúp cải thiện hiệu năng phân loại ảnh y tế nói chung và ảnh tổn thương da nói riêng [27, 51, 52, 54]. Gần đây, EfficientNetV2 tiếp tục cải thiện hiệu quả huấn luyện và hiệu quả tham số của các backbone CNN, trong khi các mô hình Transformer thị giác như Vision Transformer và Swin Transformer cho phép khai thác quan hệ không gian ở phạm vi rộng hơn thông qua cơ chế self-attention [55–57]. Trong lĩnh vực ảnh da liễu, các mô hình định hướng tổn thương da như DermViT cũng cho thấy tiềm năng của Transformer trong phân loại ảnh tổn thương da [33, 58]. Tuy nhiên, mục tiêu của luận án không phải là xây dựng một backbone học sâu hoàn toàn mới, mà tập trung vào cách đánh giá, chọn lọc và hợp nhất các biểu diễn đặc trưng được trích xuất từ mạng học sâu. Các mô hình học sâu hiện đại thường sinh ra số lượng lớn bản đồ đặc trưng ở nhiều block khác nhau của mạng. Không phải mọi bản đồ đặc trưng đều mang giá trị thông tin như nhau đối với bài toán phân loại tổn thương da; một số bản đồ có thể chứa phản hồi yếu, dư thừa, ít liên quan hoặc không ổn định. Nếu các thành phần này đều được giữ lại trước khi hợp nhất, chúng có thể gây nhiễu biểu diễn và làm giảm khả năng phân biệt của mô hình trong điều kiện dữ liệu y tế hạn chế hoặc mất cân bằng.

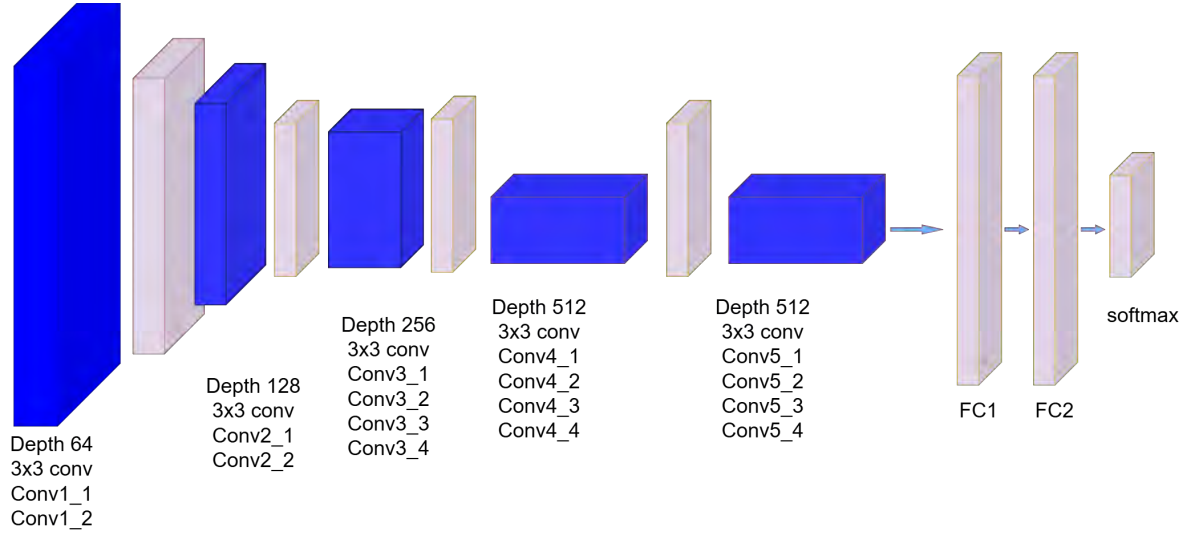
Một kiến trúc CNN điển hình thường được tổ chức thành nhiều khối xử lý (block), trong đó mỗi khối gồm một hoặc nhiều lớp tích chập, hàm kích hoạt và có thể đi kèm lớp pooling. Hình 1.2 minh họa kiến trúc VGG16 dưới dạng các block liên tiếp. Trong VGG16, các block đầu xử lý ảnh ở độ phân giải cao và sinh ra các bản đồ đặc trưng có kích thước không gian lớn; các block sau giảm dần kích thước không gian nhưng tăng số lượng kênh đặc trưng. Cách tổ chức này cho phép mô hình học biểu diễn phân cấp từ các tín hiệu thị giác đơn giản đến các đặc trưng ngữ nghĩa phức tạp hơn.

Về hình thức, giả sử ảnh x_i được đưa qua một mạng CNN và xét tại lớp tích chập thứ l . Đầu ra tại lớp này là một tập các bản đồ đặc trưng:

$$F_i^{(l)} = \{f_{i,1}^{(l)}, f_{i,2}^{(l)}, \dots, f_{i,K_l}^{(l)}\}, \quad (1.29)$$

trong đó mỗi bản đồ đặc trưng $f_{i,k}^{(l)}$ là phản ứng của một bộ lọc tích chập đối với một kiểu mẫu thị giác nhất định trong ảnh.

Không chỉ đóng vai trò là biểu diễn trung gian, các bản đồ đặc trưng đã được nhiều nghiên cứu khai thác để phân tích và giải thích mô hình học sâu. Zeiler và Fergus [59] đã đề xuất phương pháp trực quan hóa để quan sát cách các lớp CNN phản ứng với dữ liệu



Hình 1.2: Minh họa kiến trúc VGG16 dưới dạng các block tích chập.

đầu vào. Các phương pháp như Class Activation Mapping (CAM) [60] và Grad-CAM [61] tiếp tục sử dụng bản đồ đặc trưng để xác định các vùng ảnh quan trọng đối với quyết định phân loại. Trong lĩnh vực ảnh y tế, các phương pháp giải thích dựa trên bản đồ đặc trưng ngày càng được quan tâm nhằm tăng khả năng diễn giải và độ tin cậy của mô hình [62, 63].

Để minh họa cụ thể, Hình 1.3 trình bày toàn bộ 64 bản đồ đặc trưng thu được khi đưa ảnh ISIC_0024345 qua lớp block1_conv1 của VGG16. Đây là lớp tích chập đầu tiên, do đó các bản đồ đặc trưng chủ yếu phản ánh các đặc trưng mức thấp như biên, vùng chuyển màu và kết cấu.

Trong lĩnh vực phân loại ảnh tổn thương da, bản đồ đặc trưng cũng được khai thác để hỗ trợ giải thích và cải thiện hiệu quả mô hình. Van Molle và Cộng sự [64] sử dụng trực quan hóa bản đồ đặc trưng để hỗ trợ quyết định lâm sàng, trong khi các phương pháp dựa trên CAM được áp dụng để xác định vùng tổn thương quan trọng [65]. Ngoài ra, các đặc trưng học sâu trích xuất từ các tầng CNN cũng được sử dụng trực tiếp như vector đặc trưng cho các mô hình phân loại [66].

Các bản đồ đặc trưng sau đó có thể được tổng hợp để tạo thành vector đặc trưng:

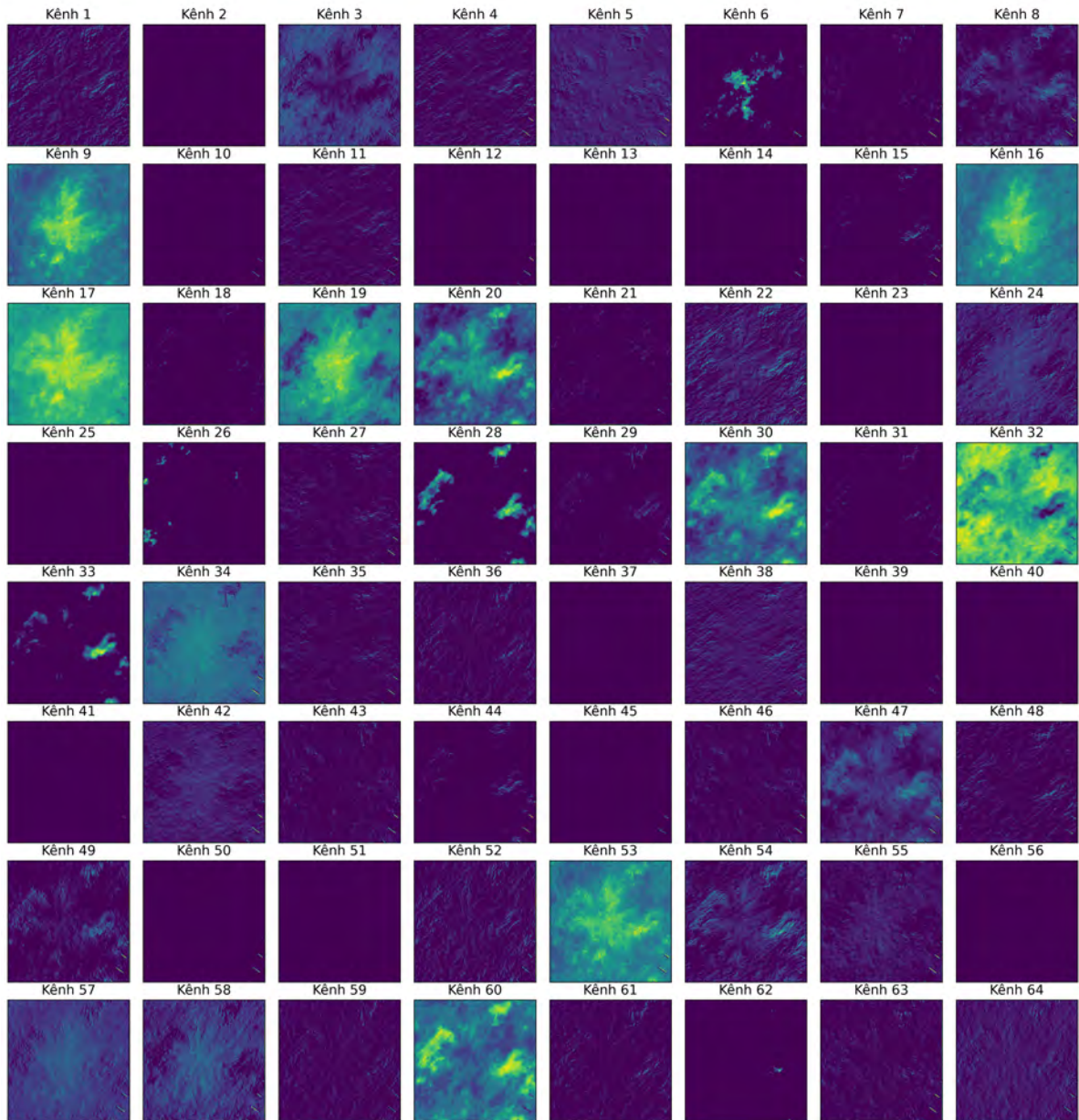
$$x_i^{(d)} = [d_1, d_2, \dots, d_{n_d}], \quad (1.30)$$

Tuy nhiên, các bản đồ đặc trưng có kiểu phân bố kích hoạt khác nhau. Một số bản đồ phản ánh rõ cấu trúc của ảnh, trong khi một số khác chứa phản hồi yếu, dư thừa, ít liên quan hoặc không ổn định. Việc giữ lại đồng thời các thành phần này có thể làm giảm chất lượng biểu diễn trước bước hợp nhất và phân loại. Điều đó đặt ra nhu cầu xây dựng một tiêu chí sàng lọc bản đồ đặc trưng và đánh giá tác động của tập biểu diễn được chọn bằng kết quả phân loại, là cơ sở cho hướng nghiên cứu được phát triển trong Chương 2.

1.2.4. Các hướng khai thác và lựa chọn bản đồ đặc trưng học sâu

Trong phân loại ảnh tổn thương da, một hướng phổ biến là trích xuất đặc trưng từ các mạng CNN đã được huấn luyện trước, sau đó hợp nhất đặc trưng lấy từ nhiều tầng hoặc nhiều kiến trúc để tăng tính bổ trợ của biểu diễn. Mahbod và Cộng sự hợp nhất các đặc trưng học sâu từ nhiều mạng đã tinh chỉnh, còn Benyahia và Cộng sự khai thác đồng

Feature maps tại lớp block1_conv1 (64/64 kênh, kích thước 224×224)

*Hình 1.3: Bản đồ đặc trưng của ảnh tại lớp block1_conv1.*

thời nhiều biểu diễn học sâu cho phân loại tổn thương da [11, 66]. Các phương pháp này cho thấy lợi ích của việc kết hợp nhiều biểu diễn, nhưng trọng tâm là lựa chọn tầng, lựa chọn mô hình hoặc ghép nối các véc-tơ đặc trưng đầu ra; chúng chưa đánh giá riêng mức đóng góp của từng bản đồ đặc trưng trong mỗi block trước khi hợp nhất với đặc trưng thủ công.

Một hướng khác là học trọng số cho các kênh đặc trưng ngay trong mạng. Khối Squeeze-and-Excitation học quan hệ giữa các kênh và tái hiệu chỉnh đáp ứng của từng kênh bằng trọng số phụ thuộc dữ liệu [67]. CBAM mở rộng cơ chế này bằng cách lần lượt tạo attention theo kênh và theo không gian để tăng cường các đáp ứng quan trọng [68]. Hai phương pháp đều là cơ chế gán trọng số mềm, được tối ưu cùng mô hình theo hàm mất mát có sử dụng nhãn. Chúng không loại bỏ trực tiếp một tập bản đồ dựa trên thống kê phân bố kích hoạt và cũng không nhằm tạo tập đặc trưng đã sàng lọc để hợp

nhất với đặc trưng thủ công bên ngoài mạng.

Entropy của bản đồ đặc trưng cũng đã được sử dụng để chọn lọc bộ lọc nhằm nén CNN. Wang và Cộng sự xây dựng tiêu chí tầm quan trọng dựa trên entropy của bản đồ đặc trưng để loại các bộ lọc ít đóng góp, trong khi Shao và Cộng sự đo lường thông tin của bản đồ để đánh giá tầm quan trọng của bộ lọc và giảm chi phí của mạng [69, 70]. Các nghiên cứu này xác nhận rằng thống kê trên bản đồ đặc trưng có thể hỗ trợ đánh giá bộ lọc, nhưng mục tiêu chính là nén kiến trúc và giảm số phép tính. Bài toán của luận án khác ở chỗ không thay đổi backbone CNN, mà đánh giá tương đối các bản đồ trong từng block, chọn biểu diễn trước khi hợp nhất với đặc trưng thủ công và kiểm chứng tập đặc trưng thu được trên dữ liệu tổn thương da mất cân bằng.

Do đó, điểm cần tiếp tục nghiên cứu không phải là khẳng định chưa từng có phương pháp lựa chọn hoặc gán trọng số cho kênh CNN. Vấn đề cụ thể là xây dựng cơ chế sàng lọc bản đồ theo từng block, không phụ thuộc nhãn tại bước tính tiêu chí, phù hợp với quá trình hợp nhất đặc trưng học sâu và đặc trưng thủ công. Trong cơ chế này, entropy chỉ cung cấp một tiêu chí hỗ trợ đánh giá về mức phân tán của kích hoạt. Giá trị đối với phân loại phải được xác nhận bằng kết quả thực nghiệm, thay vì suy ra trực tiếp từ độ lớn entropy.

Bảng 1.1: So sánh hướng tiếp cận của BlockNet với các nhóm phương pháp gần nhất

Tiêu chí	Hợp nhất trực tiếp	SE	CBAM	Chọn lọc bằng entropy để nén mạng	BlockNet
Đối tượng xử lý	Véc-tơ đặc trưng	Kênh đặc trưng	Kênh và vị trí không gian	Bộ lọc hoặc bản đồ đặc trưng	Bản đồ đặc trưng trong từng block
Tiêu chí lựa chọn hoặc điều tiết	Không lựa chọn từng bản đồ	Trọng số kênh học được	Attention theo kênh và không gian	Entropy hoặc lượng thông tin	Khoảng entropy tương đối trong từng block
Sử dụng nhãn tại bước lựa chọn	Không áp dụng	Gián tiếp qua hàm mất mát	Gián tiếp qua hàm mất mát	Không trực tiếp	Không
Phạm vi xử lý	Tầng hoặc mô hình được chọn	Các block có mô-đun SE	Các mô-đun CBAM	Nhiều lớp trong mạng	Độc lập theo từng block
Phương thức hợp nhất	Ghép nối, cộng hoặc kết hợp tuyến tính	Tái hiệu chỉnh kênh trong CNN	Tái hiệu chỉnh kênh và không gian	Không phải mục tiêu chính	Ghép nối đặc trưng học sâu đã chọn với đặc trưng thủ công
Xử lý mất cân bằng	Không tích hợp cố định	Không tích hợp	Không tích hợp	Không tích hợp	Lấy mẫu lặp trong không gian đặc trưng

So sánh trong Bảng 1.1 cho thấy đóng góp của BlockNet không nằm ở việc tạo backbone CNN mới, cũng không nằm riêng ở phép ghép nối. Điểm khác biệt là chuỗi xử lý thống nhất gồm chọn lọc bản đồ theo từng block bằng một tiêu chí không sử dụng nhãn tại bước lựa chọn, tạo biểu diễn học sâu từ tập bản đồ được giữ lại, hợp nhất với đặc trưng thủ công và giảm mất cân bằng trong không gian đặc trưng. Entropy chỉ đóng vai trò tiêu chí chọn lọc; hiệu quả của toàn bộ chuỗi xử lý được kiểm chứng bằng thực nghiệm thay vì suy ra trực tiếp từ entropy.

1.3. Hợp nhất đặc trưng và học đa nguồn

Trong bài toán phân loại ảnh tổn thương da, một mẫu có thể gồm ảnh, đặc trưng thủ công và dữ liệu mô tả lâm sàng. Học đa nguồn khai thác tính bổ trợ giữa các nguồn này để tạo biểu diễn phục vụ phân loại [14, 15, 48, 49, 71, 72].

1.3.1. Hợp nhất đặc trưng sâu và đặc trưng thủ công

Đặc trưng học sâu biểu diễn các mẫu phức tạp được học từ dữ liệu, còn đặc trưng thủ công mô tả màu sắc, kết cấu hoặc hình dạng dựa trên tri thức miền. Hợp nhất hai loại đặc trưng giúp khai thác tính bổ trợ giữa khả năng học biểu diễn và khả năng diễn giải, nhất là khi dữ liệu tổn thương da có quy mô hạn chế hoặc mất cân bằng [8–11, 66].

Về hình thức, giả sử $x_i^{(d)}$ là đặc trưng học sâu và $x_i^{(h)}$ là đặc trưng thủ công của ảnh x_i , biểu diễn hợp nhất có thể được viết:

$$x_i^{(dh)} = \mathcal{F} \left(x_i^{(d)}, x_i^{(h)} \right), \quad (1.31)$$

trong đó $\mathcal{F}(\cdot)$ là hàm hợp nhất đặc trưng.

Các nghiên cứu gần đây cho thấy việc kết hợp đặc trưng sâu và đặc trưng thủ công có thể giúp cải thiện độ ổn định của mô hình trong điều kiện dữ liệu mất cân bằng hoặc dữ liệu có sự biến thiên lớn giữa các mẫu [32, 73]. Tuy nhiên, hiệu quả của quá trình hợp nhất phụ thuộc vào chất lượng biểu diễn của từng nguồn đặc trưng trước khi kết hợp. Nếu đặc trưng đầu vào chứa nhiều thành phần dư thừa hoặc nhiễu, quá trình hợp nhất có thể làm tăng độ phức tạp của biểu diễn mà không cải thiện tương ứng khả năng phân loại.

1.3.2. Hợp nhất ảnh và dữ liệu mô tả lâm sàng

Dữ liệu mô tả lâm sàng như tuổi, giới tính và vị trí tổn thương cung cấp ngữ cảnh bổ trợ cho ảnh. Các mô hình đa nguồn có thể ghép nối trực tiếp hoặc học tương tác giữa đặc trưng ảnh và nguồn dữ liệu này bằng attention, cross-attention hay Transformer [14, 15, 17–19, 28, 29]. Mục đích là điều tiết đóng góp của từng nguồn thay vì xem dữ liệu mô tả lâm sàng như một véc-tơ phụ độc lập.

Về hình thức, biểu diễn đa nguồn của một mẫu có thể được viết:

$$x_i = \mathcal{F} \left(x_i^{(d)}, x_i^{(h)}, x_i^{(m)} \right), \quad (1.32)$$

trong đó $x_i^{(m)}$ là véc-tơ dữ liệu mô tả lâm sàng và $\mathcal{F}(\cdot)$ là hàm hợp nhất các nguồn đặc trưng.

Mặc dù dữ liệu mô tả lâm sàng thường có số chiều nhỏ hơn đặc trưng ảnh, việc hợp nhất vẫn phải xét sự khác biệt phân phối và mức đóng góp của từng nguồn trong biểu diễn chung.

1.3.3. Các phương pháp hợp nhất phổ biến

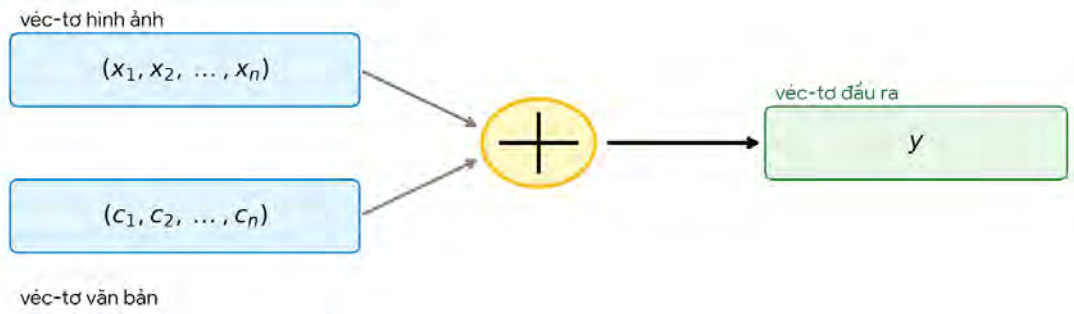
Hợp nhất đặc trưng kết hợp các véc-tơ từ nhiều nguồn thành một biểu diễn chung. Các phép cộng, nhân và ghép nối là những cách hợp nhất trực tiếp, đơn giản và thường được dùng làm phương pháp đối chứng [48, 49, 72]. Các phương pháp dựa trên attention, cross-attention và Transformer có thể học tương tác giữa các nguồn nhưng phức tạp hơn và đòi hỏi nhiều dữ liệu hơn [29, 74]. Vì vậy, hiệu quả không chỉ phụ thuộc số nguồn mà còn phụ thuộc cách lựa chọn và điều tiết biểu diễn trước khi hợp nhất.

Giả sử $x = (x_1, x_2, \dots, x_n)$ là vector đặc trưng ảnh và $c = (c_1, c_2, \dots, c_n)$ là vector đặc trưng bổ trợ, các phương pháp hợp nhất cơ bản có thể được mô tả như sau:

Cộng từng phần tử (element-wise sum):

$$y = x + c = (x_1 + c_1, x_2 + c_2, \dots, x_n + c_n), \quad (1.33)$$

trong đó mỗi thành phần đầu ra là tổng của các thành phần tương ứng. Phương pháp này không làm tăng số chiều biểu diễn nhưng giả định rằng các đặc trưng tương ứng có ý nghĩa tương đồng như minh họa trong Hình 1.4.

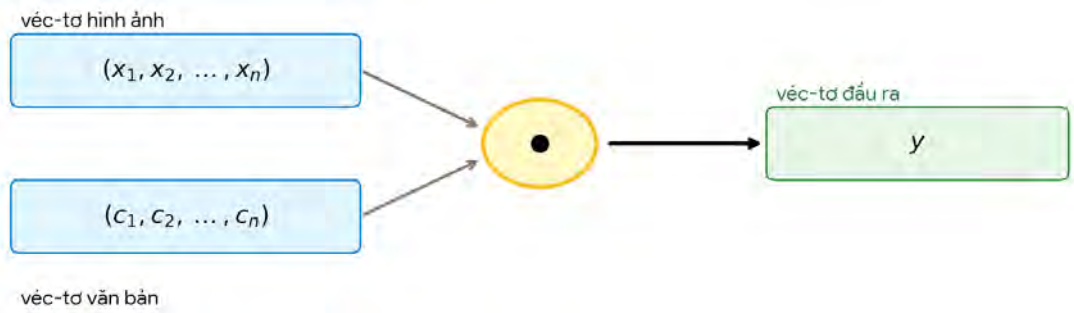


Hình 1.4: Minh họa hợp nhất đặc trưng bằng phép cộng từng phần tử.

Nhân từng phần tử (element-wise product):

$$y = x \odot c = (x_1 c_1, x_2 c_2, \dots, x_n c_n), \quad (1.34)$$

trong đó $\odot$ là phép nhân Hadamard. Phương pháp này có thể được xem như một cơ chế điều biến, trong đó một nguồn đặc trưng có thể làm nổi bật hoặc làm suy giảm ảnh hưởng của nguồn còn lại như minh họa trong Hình 1.5.



Hình 1.5: Minh họa hợp nhất đặc trưng bằng phép nhân từng phần tử.

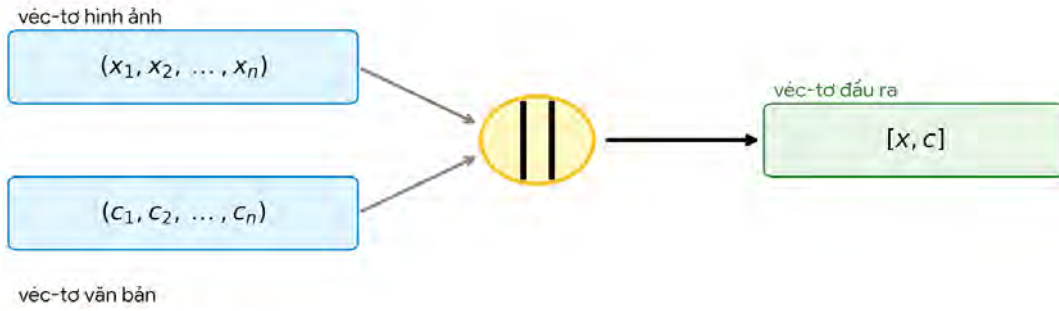
Ghép nối vector (concatenation):

$$y = [x; c], \quad (1.35)$$

trong đó vector đầu ra được tạo bằng cách nối các vector đầu vào. Đây là phương pháp linh hoạt và được sử dụng rộng rãi do không yêu cầu các vector phải có cùng số chiều như minh họa trong Hình 1.6.

Mặc dù phép ghép nối vector có ưu điểm đơn giản và giữ lại toàn bộ thông tin của các nguồn đặc trưng, phương pháp này cũng làm tăng số chiều của biểu diễn đầu ra. Khi số lượng nguồn dữ liệu hoặc số chiều đặc trưng tăng lên, biểu diễn hợp nhất có thể trở nên dư thừa và làm tăng nguy cơ quá khớp, trong các bài toán y tế có số lượng mẫu huấn luyện hạn chế. Ngoài ra, phép ghép nối trực tiếp không phân biệt được mức độ quan trọng của từng nguồn dữ liệu, do đó các đặc trưng ít hữu ích hoặc chứa nhiễu vẫn có thể ảnh hưởng đến biểu diễn cuối cùng.

Hợp nhất có trọng số và các cơ chế attention điều tiết đóng góp của từng nguồn hoặc từng thành phần đặc trưng [17, 75, 76]. Các mô hình Transformer và phương pháp



Hình 1.6: Minh họa hợp nhất đặc trưng bằng phép ghép nối vector.

dựa trên lý thuyết bằng chứng còn học tương tác hoặc mức tin cậy giữa các nguồn [18, 19, 29, 77]. Đối với ảnh tổn thương da, các mô hình này có tiềm năng khai thác tốt hơn ảnh và dữ liệu mô tả lâm sàng, nhưng độ phức tạp và nhu cầu dữ liệu khiến các phép hợp nhất trực tiếp vẫn cần thiết làm cơ sở đối chứng [15, 28].

1.3.4. Hạn chế của các phương pháp hợp nhất trực tiếp

Hợp nhất trực tiếp có ba hạn chế chính. Thứ nhất, các phép cộng, nhân hoặc ghép nối chưa mô hình hóa đầy đủ tương tác giữa đặc trưng ảnh, đặc trưng thủ công và dữ liệu mô tả lâm sàng [17–19, 75]. Thứ hai, ghép nối có thể làm tăng số chiều, giữ lại thành phần dư thừa và tăng nguy cơ quá khớp khi dữ liệu y tế hạn chế [49]. Thứ ba, các phép hợp nhất này không đánh giá chất lượng từng nguồn trước khi kết hợp. Do đó, cần lựa chọn và điều tiết biểu diễn trước hợp nhất để hạn chế nhiễu, đồng thời khai thác quan hệ giữa các mẫu trong các chương tiếp theo.

1.4. Entropy trong học biểu diễn và đánh giá độ tin cậy

Việc học biểu diễn đặc trưng và hợp nhất đa nguồn đặt ra hai vấn đề quan trọng: (i) đánh giá chất lượng của đặc trưng và (ii) đo lường độ tin cậy của dự đoán. Entropy, xuất phát từ lý thuyết thông tin [78, 79], được sử dụng trong luận án trên hai phân phối khác nhau. Entropy kích hoạt được tính từ tần suất của các mức giá trị xuất hiện trong bản đồ đặc trưng bằng logarit cơ số 2 và có đơn vị bit. Entropy dự đoán được tính từ phân phối xác suất lớp bằng logarit tự nhiên và có đơn vị nat. Hai đại lượng khác đối tượng tính, cơ sở và mục đích sử dụng nên không được diễn giải thay thế cho nhau.

1.4.1. Entropy trên phân phối xác suất

Trong bài toán phân loại, đầu ra của mô hình thường được biểu diễn dưới dạng phân phối xác suất trên các lớp. Entropy của phân phối này phản ánh mức độ không chắc chắn của mô hình đối với dự đoán. Khi phân phối xác suất tập trung mạnh vào một lớp, entropy thường thấp và cho thấy mô hình có xu hướng tự tin hơn với dự đoán. Ngược lại, khi xác suất phân bố tương đối đều giữa nhiều lớp, entropy cao hơn và phản ánh mức độ không chắc chắn lớn hơn [80, 81].

ĐỊNH NGHĨA 1.1 [Entropy dự đoán] Cho mô hình phân loại f_θ với tập tham số θ . Với một mẫu đầu vào x , mô hình sinh ra phân phối xác suất dự đoán trên (C) lớp: $p_\theta(y|x) = [p_\theta(1|x), \dots, p_\theta(C|x)]$. Entropy dự đoán của mô hình đối với mẫu (x) được xác định theo logarit tự nhiên bởi:

$$H_{\text{pred}}(p_\theta(x)) = - \sum_{c=1}^C p_\theta(c|x) \ln p_\theta(c|x). \quad (1.36)$$

Trong đó, $p_{\theta}(c|x)$ là xác suất dự đoán mẫu x thuộc lớp c . Entropy dự đoán sử dụng logarit tự nhiên, có đơn vị nat và được dùng để so sánh mức độ chắc chắn giữa các dự đoán. Với bài toán nhị phân, giá trị lớn nhất là $\ln 2$, đạt được khi hai lớp cùng có xác suất 0,5.

Giá trị entropy thấp cho thấy phân phối xác suất dự đoán tập trung vào một hoặc một số ít lớp, tương ứng với mức độ không chắc chắn thấp hơn. Ngược lại, giá trị entropy cao cho thấy xác suất dự đoán phân tán hơn trên nhiều lớp, phản ánh mức độ không chắc chắn cao hơn của mô hình.

Trong nhiều nghiên cứu gần đây về học sâu có kiểm soát độ tin cậy, entropy dự đoán được sử dụng như một chỉ số phản ánh mức độ không chắc chắn của mô hình, qua đó hỗ trợ phát hiện các mẫu khó phân loại hoặc các dự đoán có độ tin cậy thấp [82–84]. Bên cạnh đó, entropy cũng được khai thác để lựa chọn nhãn giả trong học bán giám sát hoặc đánh giá mức độ tin cậy của mô hình khi gặp dữ liệu có phân bố khác với dữ liệu huấn luyện [25, 85]. Trong phạm vi luận án này, entropy dự đoán được sử dụng như một tiêu chí định lượng để đánh giá mức độ không chắc chắn của mô hình, làm cơ sở cho việc lựa chọn các mẫu chưa gán nhãn có độ tin cậy cao trong quá trình gán nhãn giả ở Chương 3.

VÍ DỤ 1.1 [Entropy trên phân phối xác suất] Xét bài toán hai lớp.

Trường hợp 1 (entropy thấp):

$$p = [0.95, 0.05]$$

$$H = -(0.95 \ln 0.95 + 0.05 \ln 0.05) = 0.1985$$

Trường hợp 2 (entropy cao):

$$p = [0.50, 0.50]$$

$$H = -(0.5 \ln 0.5 + 0.5 \ln 0.5) = 0.6931$$

Kết quả cho thấy entropy trong trường hợp thứ hai lớn hơn, phản ánh mức độ không chắc chắn cao hơn của mô hình. Các giá trị trong ví dụ được tính bằng logarit tự nhiên và có đơn vị nat.

1.4.2. Entropy trên biểu diễn đặc trưng

Bên cạnh entropy dự đoán, entropy có thể được tính trên histogram giá trị kích hoạt của từng bản đồ đặc trưng CNN. Đại lượng này mô tả mức phân tán của đáp ứng trong bản đồ: phân bố tập trung tạo entropy thấp, còn phân bố rộng hơn tạo entropy cao. Vì phép tính không sử dụng nhãn lớp, entropy kích hoạt không trực tiếp đo khả năng phân biệt, mức liên quan đến nhãn hoặc giá trị chẩn đoán của bản đồ đặc trưng.

Để định lượng mức độ đa dạng thông tin của bản đồ đặc trưng, entropy có thể được tính dựa trên histogram của các giá trị trong bản đồ đặc trưng [78, 79]. Các nghiên cứu gần đây cũng cho thấy entropy có thể hỗ trợ đánh giá và lựa chọn đặc trưng trong các mô hình học sâu [84, 86]. Trong mục này, entropy được trình bày ở mức tổng quan như một thước đo phản ánh sự phân bố thông tin của bản đồ đặc trưng; cách sử dụng entropy để chọn lọc bản đồ đặc trưng trong mô hình đề xuất sẽ được trình bày cụ thể ở Chương 2.

Trong luận án, entropy kích hoạt không được sử dụng như một tiêu chí tuyệt đối để khẳng định bản đồ đặc trưng tốt hay xấu. Entropy thấp chỉ cho biết phân bố kích hoạt tập trung hơn; entropy cao chỉ cho biết phân bố rộng hơn và trong một số trường hợp có thể

đi kèm nhiều. BlockNet sử dụng một khoảng entropy tương đối trong từng block như giả thuyết sàng lọc không giám sát. Tính hữu ích đối với phân loại của tập bản đồ được giữ lại được kiểm chứng bằng các thí nghiệm ở Chương 2, thay vì suy ra trực tiếp từ entropy của từng bản đồ.

Entropy được tính từ histogram của một bản đồ đặc trưng là entropy biên của phân bố kích hoạt. Đại lượng này không sử dụng nhãn lớp và không trực tiếp đo khả năng phân biệt giữa các lớp. Vì vậy, một bản đồ có entropy cao vẫn có thể chủ yếu phản ánh nhiễu hoặc biến thiên không liên quan; ngược lại, một bản đồ entropy thấp vẫn có thể hữu ích nếu phản ứng chọn lọc với một dấu hiệu quan trọng của tổn thương. Cách diễn giải phù hợp là entropy mô tả mức phân tán của đáp ứng, không phải thước đo tuyệt đối về giá trị chẩn đoán của bản đồ đặc trưng.

Về bản chất, thông tin tương hỗ (mutual information) đánh giá mức phụ thuộc giữa đặc trưng và nhãn; Fisher score so sánh mức phân tách giữa các lớp với mức phân tán trong từng lớp; còn Chi-square đánh giá mức phụ thuộc thống kê giữa đặc trưng và nhãn sau khi biểu diễn dữ liệu phù hợp. Đây là các tiêu chí có giám sát, khác với entropy biên được tính trực tiếp trên từng bản đồ mà không cần nhãn. Vì các đại lượng đo những thuộc tính khác nhau, không thể mặc nhiên suy ra tương quan đơn điệu giữa entropy với thông tin tương hỗ, Fisher score hoặc Chi-square. Trong luận án, entropy được dùng làm tiêu chí sàng lọc không giám sát; tính hữu ích của tập bản đồ được chọn phải được đánh giá bằng hiệu năng phân loại và các thí nghiệm đối chứng trình bày ở Chương 2.

ĐỊNH NGHĨA 1.2 [Entropy của bản đồ đặc trưng] Cho bản đồ đặc trưng f gồm N_f phần tử. Gọi K_f là số mức giá trị kích hoạt khác nhau thực sự xuất hiện trong f , n_r là số lần xuất hiện của mức thứ r và $p_r = n_r/N_f$. Trong đó entropy kích hoạt được xác định bởi:

$$H_{\text{act}}(f) = - \sum_{r=1}^{K_f} p_r \log_2 p_r. \quad (1.37)$$

Entropy kích hoạt có đơn vị bit và thỏa mãn:

$$0 \leq H_{\text{act}}(f) \leq \log_2 K_f \leq \log_2 N_f. \quad (1.38)$$

Giá trị $H_{\text{act}}(f)$ phản ánh mức độ phân tán của các giá trị kích hoạt trong bản đồ đặc trưng. Entropy thấp cho thấy các giá trị tập trung vào một số ít mức, còn entropy cao cho thấy các giá trị phân bố rộng hơn trên nhiều mức. Vì $H_{\text{act}}(f)$ không sử dụng nhãn lớp, đại lượng này không trực tiếp đo khả năng phân biệt giữa các lớp. Trong luận án, $H_{\text{act}}(f)$ được sử dụng như một tiêu chí sàng lọc không giám sát theo nghĩa tương đối trong từng block.

VÍ DỤ 1.2 [Entropy của nhiều bản đồ đặc trưng trong cùng một block] Xét ảnh ISIC_0024345 tại lớp block1_conv1 của VGG16. Giả sử chọn 5 bản đồ đặc trưng với các phân phối histogram:

$$p^{(1)} = [0.70, 0.20, 0.08, 0.02], \quad p^{(2)} = [0.55, 0.25, 0.15, 0.05],$$

$$p^{(3)} = [0.40, 0.30, 0.20, 0.10], \quad p^{(4)} = [0.30, 0.25, 0.25, 0.20],$$

$$p^{(5)} = [0.25, 0.25, 0.25, 0.25]$$

Entropy tương ứng:

$$H(f_1) = 0.8503, \quad H(f_2) = 1.1217, \quad H(f_3) = 1.2799,$$

$$H(f_4) = 1.3762, \quad H(f_5) = 1.3863$$

Kết quả cho thấy bản đồ đặc trưng có phân phối càng tập trung thì entropy càng thấp, trong khi phân phối càng đồng đều thì entropy càng cao. Trong cùng một block, entropy được sử dụng để so sánh tương đối mức phân tán của phân bố kích hoạt giữa các bản đồ đặc trưng.

Trong cùng một lớp tích chập, entropy của các bản đồ đặc trưng có thể được so sánh theo phân bố chung của toàn bộ các kênh. Cụ thể, gọi H_k là entropy của bản đồ đặc trưng thứ k , và $\bar{H}$ là giá trị entropy trung bình của tất cả các bản đồ đặc trưng trong lớp. Khi đó:

- Bản đồ đặc trưng có H_k nhỏ hơn so với $\bar{H}$ được xem là **entropy thấp**, thường phản ánh phân bố giá trị tập trung và có thể chứa ít biến thiên thông tin.
- Bản đồ đặc trưng có H_k xấp xỉ $\bar{H}$ được xem là **entropy trung bình**, thường phản ánh mức độ phân bố thông tin cân bằng hơn trong cùng lớp tích chập.
- Bản đồ đặc trưng có H_k lớn hơn so với $\bar{H}$ được xem là **entropy cao**, thường phản ánh phân bố giá trị phân tán mạnh và có thể chứa nhiều hoặc biểu diễn kém ổn định.

Trong nghiên cứu này, việc phân loại entropy thấp, trung bình và cao được hiểu theo nghĩa tương đối trong cùng một lớp hoặc cùng một block, thay vì sử dụng một ngưỡng tuyệt đối cố định cho mọi trường hợp. Cách phân loại chỉ so sánh mức phân tán kích hoạt giữa các bản đồ; nó không xếp hạng trực tiếp khả năng phân biệt lớp. Cơ chế sàng lọc theo khoảng entropy và cách đánh giá tập bản đồ được giữ lại được trình bày trong Chương 2.

Bảng 1.2 minh họa một số bản đồ đặc trưng được chọn từ lớp block1_conv1. Các bản đồ đặc trưng có entropy thấp, trung bình và cao được xác định tương đối theo phân bố entropy của toàn bộ các kênh trong cùng một lớp.

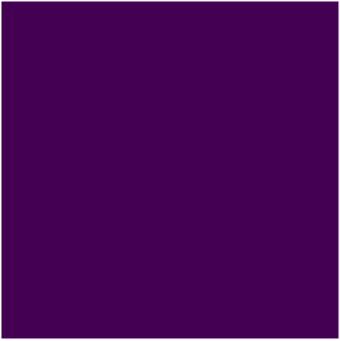
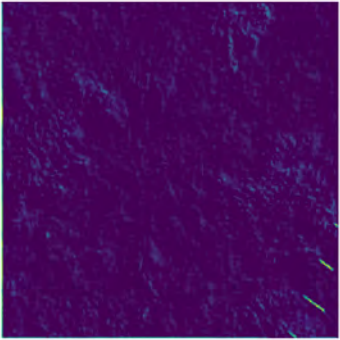
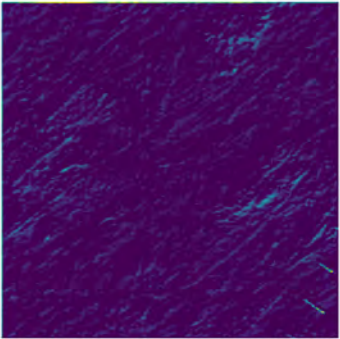
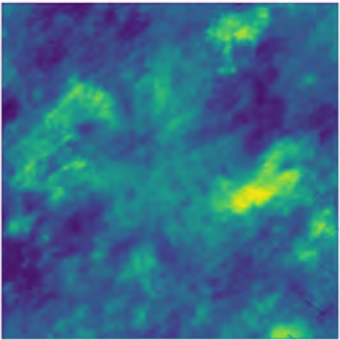
1.4.3. Vai trò của entropy

Luận án sử dụng hai đại lượng cần được phân biệt. Ở Chương 2, entropy kích hoạt $H_{\text{act}}(F_m)$ được tính trên histogram của bản đồ đặc trưng và mô tả mức phân tán kích hoạt. Ở Chương 3, entropy dự đoán $H_{\text{pred}}(\hat{y}_i)$ được tính trên phân phối xác suất lớp và được sử dụng như tiêu chí hỗ trợ đánh giá gần đúng về mức độ không chắc chắn. Entropy kích hoạt không trực tiếp đo khả năng phân biệt lớp; entropy dự đoán thấp cũng không bảo đảm dự đoán đúng nếu mô hình chưa được hiệu chỉnh xác suất tốt.

Trong các mô hình học sâu hiện đại, entropy ngày càng được sử dụng rộng rãi để đánh giá mức độ không chắc chắn, chất lượng hiệu chỉnh xác suất và độ tin cậy của mô hình [81, 87, 88]. Đặc biệt trong lĩnh vực ảnh y tế, entropy đã được ứng dụng để:

- phát hiện các mẫu khó phân loại,
- đánh giá độ tin cậy của dự đoán,
- chọn lọc nhãn giả trong học bán giám sát,

Bảng 1.2: Ví dụ entropy của một số bản đồ đặc trưng của ảnh “ISIC_0024345.jpg” tại lớp *block1_conv1* của VGG16

Kênh	Bản đồ đặc trưng	Nhóm entropy
	H = 0.0000	
2		Thấp
	H = 7.4503	
43		Trung bình
	H = 7.8394	
48		Trung bình
	H = 15.6095	
30		Cao

- và phân tích chất lượng biểu diễn đặc trưng.

Các nghiên cứu gần đây tiếp tục khai thác entropy và các thước đo không chắc chắn

để nâng cao độ tin cậy của mô hình học sâu trong ảnh y tế [82–86].

Trong luận án, entropy được sử dụng theo hai hướng độc lập về đối tượng đo: (i) sàng lọc không giám sát các bản đồ theo mức phân tán kích hoạt tương đối trong từng block ở Chương 2; và (ii) phân tầng mức độ tự tin tương đối của dự đoán để hỗ trợ chọn nhãn giả ở Chương 3. Hiệu quả của hướng thứ nhất được đánh giá bằng kết quả phân loại của tập biểu diễn sau sàng lọc; độ tin cậy của hướng thứ hai được diễn giải trong giới hạn của entropy dự đoán và không thay thế việc đánh giá chất lượng hiệu chỉnh xác suất.

1.5. Học trên đồ thị và học bán giám sát

Trong các bài toán phân loại ảnh, trong lĩnh vực ảnh y tế, dữ liệu thường tồn tại dưới dạng các mẫu độc lập và số lượng dữ liệu có nhãn thường hạn chế. Điều này làm giảm khả năng tổng quát hóa của mô hình học sâu nếu chỉ dựa trên dữ liệu có nhãn. Do đó, việc khai thác mối quan hệ giữa các mẫu và tận dụng dữ liệu chưa gán nhãn trở thành một hướng tiếp cận quan trọng. Hai hướng nghiên cứu tiêu biểu là biểu diễn dữ liệu dưới dạng đồ thị và học bán giám sát [20–22, 89].

Các nghiên cứu gần đây cho thấy mạng nơ-ron đồ thị (GNN) có khả năng khai thác hiệu quả cấu trúc quan hệ giữa các mẫu dữ liệu thông qua cơ chế lan truyền thông tin trên đồ thị [90, 91]. Đồng thời, học bán giám sát ngày càng được quan tâm trong ảnh y tế do khả năng tận dụng dữ liệu chưa gán nhãn để cải thiện hiệu quả mô hình mà không cần chi phí gán nhãn lớn [92, 93].

1.5.1. Biểu diễn dữ liệu trên đồ thị

Trong các phương pháp học truyền thống, các mẫu dữ liệu thường được xem là độc lập. Tuy nhiên, trong nhiều bài toán thực tế, đặc biệt là phân loại ảnh tổn thương da, các mẫu có thể tồn tại những mối quan hệ tiềm ẩn. Do đó, một cách tiếp cận hiệu quả là biểu diễn dữ liệu dưới dạng đồ thị $G = (V, E)$, trong đó mỗi nút $v_i \in V$ tương ứng với một mẫu dữ liệu (ví dụ: một ảnh), và mỗi cạnh $e_{ij} \in E$ biểu diễn mối quan hệ giữa hai mẫu.

Mối quan hệ giữa các mẫu thường được xác định dựa trên độ tương đồng trong không gian đặc trưng. Ma trận kề của đồ thị được xây dựng như sau:

$$A_{ij} = \text{sim}(x_i, x_j), \quad (1.39)$$

trong đó $\text{sim}(\cdot)$ có thể là cosine similarity hoặc khoảng cách Euclidean. Trong thực tế, để đảm bảo tính hiệu quả và tránh nhiễu, đồ thị thường được xây dựng bằng phương pháp k-nearest neighbors (k-NN), trong đó mỗi nút chỉ kết nối với một số lượng hữu hạn các nút lân cận có độ tương đồng cao [89, 90]. Tham số k trong đồ thị k-NN không nên được hiểu là một hằng số cố định cho mọi dữ liệu, mà là một siêu tham số cần được lựa chọn theo đặc điểm của tập dữ liệu, mật độ mẫu và mục tiêu cân bằng giữa khả năng khai thác quan hệ lân cận và nguy cơ đưa thêm cạnh nhiễu vào đồ thị.

Trên cơ sở đồ thị này, các mô hình học sâu trên đồ thị, đặc biệt là mạng nơ-ron đồ thị (mạng nơ-ron đồ thị – GNN), được sử dụng để khai thác cấu trúc quan hệ giữa các mẫu. Thông qua cơ chế lan truyền và tổng hợp thông tin giữa các nút lân cận, GNN cho phép mỗi nút không chỉ học từ đặc trưng của chính nó mà còn từ các nút xung quanh, từ đó hình thành các biểu diễn đặc trưng có tính ngữ cảnh [20–22, 91].

Điều này đặc biệt phù hợp trong bài toán phân loại ảnh tổn thương da, nơi các mẫu có thể có biểu hiện thị giác tương tự nhưng thuộc các lớp khác nhau, hoặc ngược lại. Việc khai thác quan hệ giữa các mẫu trong không gian đặc trưng giúp mô hình tận dụng

thông tin bổ trợ từ các mẫu lân cận, qua đó cải thiện khả năng phân biệt giữa các lớp và tăng tính ổn định của biểu diễn.

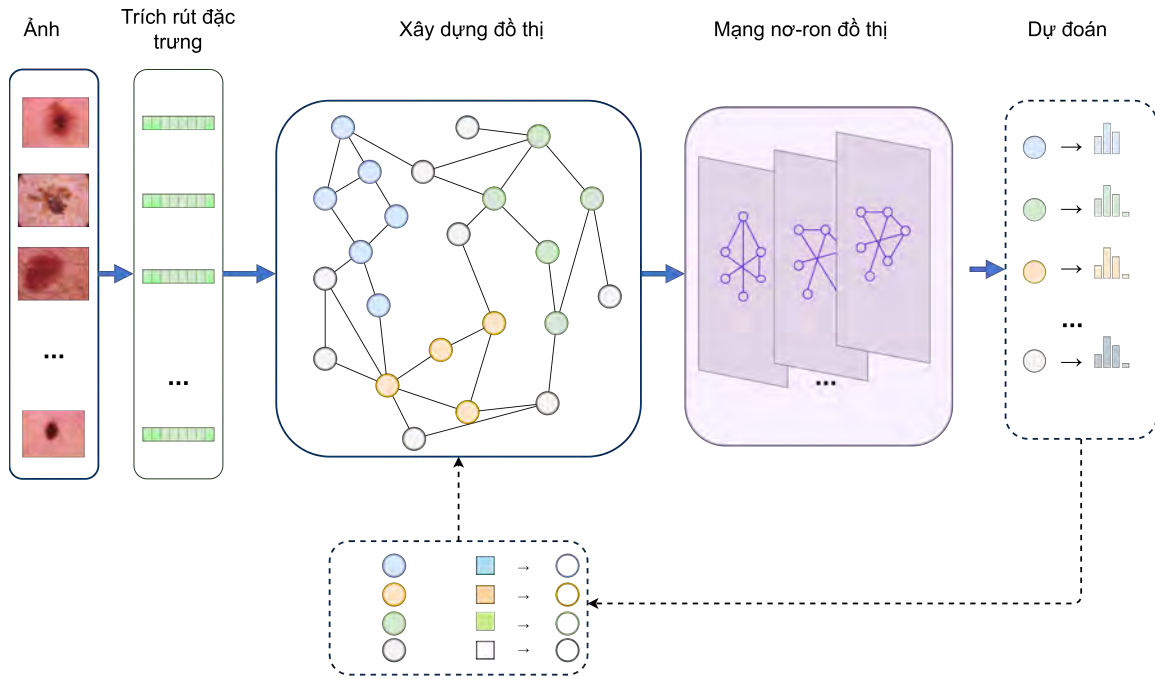
Một bước cập nhật đặc trưng điển hình trong GNN có thể được mô tả như sau:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} W^{(l)} h_j^{(l)} \right), \quad (1.40)$$

trong đó $h_i^{(l)}$ là vector đặc trưng của nút i tại lớp l , $\mathcal{N}(i)$ là tập các nút lân cận, và $W^{(l)}$ là tham số cần học [20–22].

Ngoài các mô hình GCN cơ bản, nhiều biến thể GNN khác cũng đã được đề xuất nhằm cải thiện khả năng biểu diễn và lan truyền thông tin trên đồ thị, chẳng hạn như Graph Attention Network (GAT), GraphSAGE và Graph Isomorphism Network (GIN) [21, 22, 91]. Các mô hình này cho phép học biểu diễn linh hoạt hơn thông qua attention hoặc cơ chế tổng hợp đặc trưng thích nghi.

Hình 1.7 minh họa cách dữ liệu ảnh được chuyển đổi thành đồ thị và được xử lý bởi các mô hình học trên đồ thị. Trong đó, các ảnh đầu vào được trích xuất thành các vector đặc trưng, sau đó được biểu diễn dưới dạng các nút trong đồ thị. Các cạnh giữa các nút phản ánh mức độ tương đồng giữa các mẫu, trong khi các lớp GNN thực hiện việc lan truyền thông tin để học các biểu diễn đặc trưng có tính ngữ cảnh.



Hình 1.7: Minh họa pipeline học trên đồ thị cho bài toán phân loại ảnh tổn thương da. Các đặc trưng ảnh và metadata được sử dụng để xây dựng đồ thị, trong đó mỗi nút biểu diễn một mẫu và các cạnh phản ánh mức độ tương đồng giữa các mẫu. Mô hình GNN lan truyền thông tin giữa các nút lân cận nhằm học biểu diễn có tính ngữ cảnh, tạo cơ sở cho việc khai thác quan hệ giữa các mẫu trong các thiết lập học bán giám sát.

Từ hình có thể thấy rằng mỗi ảnh được biểu diễn như một nút trong đồ thị, trong đó các cạnh thể hiện mối quan hệ tương đồng giữa các mẫu. Quá trình lan truyền thông tin trong GNN giúp các nút cập nhật biểu diễn của mình dựa trên ngữ cảnh lân cận, từ đó nâng cao khả năng phân biệt giữa các lớp dữ liệu.

Trong lĩnh vực ảnh y tế, nhiều nghiên cứu đã ứng dụng GNN cho các bài toán phân loại bệnh, phân tích ảnh mô bệnh học và phân loại tổn thương da [89, 94, 95]. Các nghiên cứu này cho thấy GNN có khả năng tận dụng hiệu quả mối quan hệ giữa các mẫu dữ liệu trong không gian đặc trưng, trong các thiết lập dữ liệu hạn chế nhãn.

1.5.2. Học bán giám sát

Trong nhiều bài toán thực tế, dữ liệu có nhãn thường khan hiếm trong khi dữ liệu chưa gán nhãn lại rất phong phú. Học bán giám sát (semi-supervised learning) được phát triển nhằm tận dụng cả hai loại dữ liệu này để nâng cao hiệu quả của mô hình [96, 97].

Trong bối cảnh học trên đồ thị, GNN có khả năng lan truyền thông tin từ các nút có nhãn sang các nút chưa có nhãn thông qua cấu trúc kết nối của đồ thị. Điều này giúp cải thiện khả năng học biểu diễn và phân loại ngay cả khi số lượng nhãn ban đầu hạn chế [20, 93].

Một kỹ thuật phổ biến trong học bán giám sát là gán nhãn giả, trong đó mô hình được sử dụng để dự đoán nhãn cho các mẫu chưa gán nhãn, sau đó các nhãn dự đoán này được sử dụng như nhãn giả để mở rộng tập dữ liệu huấn luyện [23–25]. Về bản chất, gán nhãn giả cho phép mô hình khai thác thêm thông tin từ dữ liệu chưa gán nhãn mà không cần chi phí chú thích thủ công. Khi các dự đoán đủ chính xác, phương pháp này có thể giúp mở rộng tập huấn luyện hiệu quả và cải thiện khả năng tổng quát hóa của mô hình.

Tuy nhiên, gán nhãn giả cũng tiềm ẩn rủi ro lan truyền sai lệch nếu mô hình sử dụng các dự đoán thiếu tin cậy để tái huấn luyện. Trong trường hợp mô hình dự đoán sai nhưng vẫn có độ tự tin cao, các nhãn giả sai có thể tiếp tục được đưa ngược vào quá trình học, làm khuếch đại sai số qua nhiều vòng lặp huấn luyện. Hiện tượng này thường được gọi là confirmation bias [25]. Rủi ro này đặc biệt đáng chú ý trong các bài toán ảnh y tế và dữ liệu mất cân bằng, nơi các lớp hiếm thường khó dự đoán hơn và dễ bị chi phối bởi các lớp phổ biến.

Do đó, trong các nghiên cứu gần đây, vấn đề quan trọng không chỉ nằm ở việc tạo nhãn giả, mà còn ở cách đánh giá độ tin cậy của nhãn giả trước khi sử dụng cho tái huấn luyện. Nhiều hướng tiếp cận đã đề xuất các cơ chế có xét đến mức độ không chắc chắn hoặc có kiểm soát độ tin cậy nhằm lựa chọn các mẫu có độ tin cậy cao [84, 85, 98]. Một số nghiên cứu sử dụng entropy hoặc ước lượng mức độ không chắc chắn để đánh giá mức độ chắc chắn của mô hình trước khi sinh nhãn giả [80, 82]. Đây cũng là cơ sở để luận án tiếp cận bài toán gán nhãn giả theo hướng có kiểm soát độ tin cậy, trong đó entropy dự đoán được sử dụng để đánh giá độ tin cậy và chọn lọc các nhãn giả trước khi tái huấn luyện trong Chương 3.

Các phương pháp học bán giám sát hiện đại như FixMatch, consistency regularization hoặc self-training đã cho thấy hiệu quả trong nhiều bài toán thị giác máy tính và ảnh y tế [24, 99]. Đặc biệt, trong ảnh y tế, học bán giám sát được xem là hướng tiếp cận phù hợp do quá trình gán nhãn thường yêu cầu chuyên gia và chi phí cao [92].

1.5.3. Rủi ro của gán nhãn giả

Mặc dù gán nhãn giả mang lại nhiều lợi ích, phương pháp này cũng tiềm ẩn những rủi ro. Nếu mô hình đưa ra dự đoán sai và các nhãn sai này được sử dụng trong quá trình huấn luyện tiếp theo, sai số có thể bị khuếch đại qua các vòng lặp, dẫn đến hiện tượng lan truyền sai lệch nhãn, thường được gọi là confirmation bias [25].

Rủi ro này đặc biệt nghiêm trọng trong các bài toán dữ liệu mất cân bằng, khi các lớp chiếm ưu thế có thể chi phối quá trình học và làm suy giảm khả năng nhận diện các

lớp hiếm. Do đó, vấn đề cốt lõi không chỉ nằm ở việc sử dụng nhãn giả, mà ở cách lựa chọn và kiểm soát chất lượng của các nhãn giả được đưa vào huấn luyện.

Nhiều nghiên cứu gần đây đã đề xuất các cơ chế có xét đến mức độ không chắc chắn hoặc có kiểm soát độ tin cậy nhằm giảm rủi ro của gán nhãn giả bằng cách chỉ lựa chọn các mẫu có độ tin cậy cao [84, 85]. Một số phương pháp còn kết hợp entropy hoặc ước lượng mức độ không chắc chắn để đánh giá mức độ chắc chắn của mô hình trước khi sinh nhãn giả [80, 82].

Trong bối cảnh ảnh tổn thương da, việc kết hợp học bán giám sát với mô hình học trên đồ thị và cơ chế đánh giá độ tin cậy được xem là hướng nghiên cứu tiềm năng nhằm tận dụng dữ liệu chưa gán nhãn đồng thời hạn chế nguy cơ lan truyền sai lệch nhãn.

1.6. Đánh giá mô hình phân loại ảnh y tế

Đánh giá mô hình là một bước quan trọng trong các bài toán phân loại ảnh y tế nhằm xác định khả năng tổng quát hóa, độ ổn định và mức độ tin cậy của mô hình trên dữ liệu chưa quan sát. Trong các bài toán phân loại tổn thương da, dữ liệu thường mất cân bằng mạnh giữa các lớp, do đó việc chỉ sử dụng accuracy là chưa đủ để phản ánh đầy đủ chất lượng của mô hình [100–104].

Các nghiên cứu gần đây trong lĩnh vực ảnh y tế thường sử dụng đồng thời nhiều độ đo như precision, recall, sensitivity, specificity, F1-score và AUC nhằm đánh giá mô hình từ nhiều góc nhìn khác nhau [105–108]. Ngoài ra, trong các thiết lập dữ liệu mất cân bằng, các độ đo macro-average ngày càng được sử dụng rộng rãi nhằm giảm ảnh hưởng chi phối của các lớp lớn [102–104].

1.6.1. Các chỉ số theo lớp trong dữ liệu mất cân bằng

Trong các bài toán phân loại với dữ liệu mất cân bằng, việc đánh giá mô hình chỉ bằng accuracy là chưa đủ, vì mô hình có thể đạt độ chính xác cao nhưng vẫn hoạt động kém trên các lớp hiếm. Do đó, các chỉ số theo lớp như precision, recall, sensitivity, specificity và F1-score được sử dụng để phản ánh đầy đủ hơn hành vi của bộ phân loại [100–104].

Accuracy được xác định bởi Công thức (1.41).

$$\text{Acc} := \frac{TP + TN}{TP + TN + FP + FN}. \quad (1.41)$$

Trong đó, Acc là độ chính xác tổng thể; TP là số mẫu dương tính đúng; TN là số mẫu âm tính đúng; FP là số mẫu âm tính bị dự đoán sai thành dương tính; và FN là số mẫu dương tính bị dự đoán sai thành âm tính.

Precision của lớp c được xác định bởi Công thức (1.42).

$$\text{Prec}_c := \frac{TP_c}{TP_c + FP_c}. \quad (1.42)$$

Trong đó, Prec_c là precision của lớp c ; TP_c là số mẫu được dự đoán đúng thuộc lớp c ; và FP_c là số mẫu không thuộc lớp c nhưng bị dự đoán nhầm là lớp c . Precision đặc biệt hữu ích khi cần kiểm soát số lượng dự đoán dương tính giả [100, 102–104].

Recall của lớp c được xác định bởi Công thức (1.43).

$$\text{Rec}_c := \frac{TP_c}{TP_c + FN_c}. \quad (1.43)$$

Trong đó, Rec_c là recall của lớp c ; TP_c là số mẫu dương tính đúng của lớp c ; và FN_c là số mẫu thuộc lớp c nhưng bị dự đoán sai. Recall phản ánh khả năng phát hiện đúng các mẫu thuộc lớp quan tâm [100, 102–104].

Sensitivity được xác định bởi Công thức (1.44).

$$Sens := \frac{TP}{TP + FN}. \quad (1.44)$$

Trong đó, Sens là độ nhạy; TP là số mẫu dương tính đúng; và FN là số mẫu dương tính bị dự đoán sai. Trong các bài toán y sinh, sensitivity đặc biệt quan trọng vì phản ánh khả năng phát hiện đúng các trường hợp bệnh [101, 103–105].

Specificity được xác định bởi Công thức (1.45).

$$Spec := \frac{TN}{TN + FP}. \quad (1.45)$$

Trong đó, Spec là độ đặc hiệu; TN là số mẫu âm tính đúng; và FP là số mẫu âm tính bị dự đoán sai thành dương tính. Chỉ số này phản ánh khả năng loại trừ đúng các mẫu không thuộc lớp bệnh [101, 103–105].

F1-score được xác định bởi Công thức (1.46).

$$F1 := \frac{2 \text{Prec} \text{Rec}}{\text{Prec} + \text{Rec}}. \quad (1.46)$$

Trong đó, F1 là trung bình điều hòa giữa precision và recall; Prec là precision; và Rec là recall. F1-score phù hợp khi cần đánh giá đồng thời khả năng phát hiện đúng và khả năng hạn chế dự đoán nhầm trong điều kiện dữ liệu lệch lớp [100, 102–104].

Ngoài các chỉ số trên, giá trị loss cũng thường được theo dõi trong quá trình huấn luyện để đánh giá khả năng hội tụ và độ ổn định của quá trình tối ưu. Tuy nhiên, loss chủ yếu phản ánh hành vi tối ưu hóa, trong khi accuracy, precision, recall, sensitivity, specificity và F1-score phản ánh trực tiếp chất lượng phân loại trên dữ liệu đánh giá.

1.6.2. ROC, AUC và các độ đo tổng hợp

Để đánh giá khả năng phân biệt của mô hình trên toàn bộ miền ngưỡng quyết định, đường cong ROC và diện tích dưới đường cong AUC được sử dụng rộng rãi [101, 103–105].

True Positive Rate được xác định bởi Công thức (1.47).

$$TPR := \frac{TP}{TP + FN}. \quad (1.47)$$

Trong đó, TPR là tỉ lệ dương tính đúng; TP là số mẫu dương tính đúng; và FN là số mẫu dương tính bị dự đoán sai.

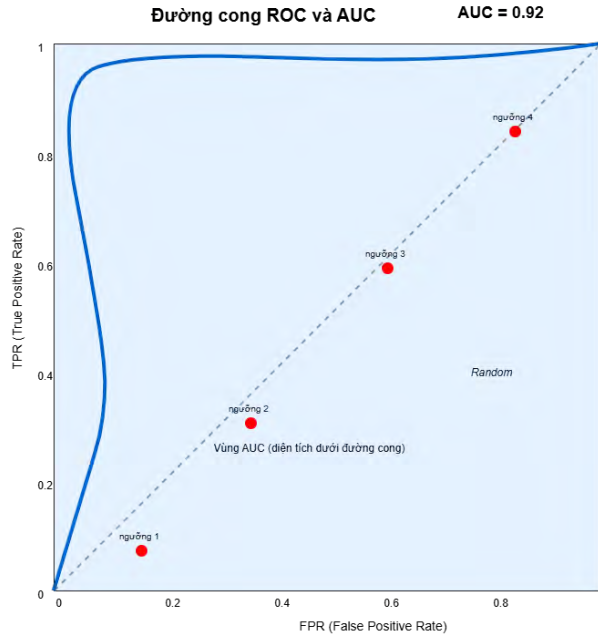
False Positive Rate được xác định bởi Công thức (1.48).

$$FPR := \frac{FP}{FP + TN}. \quad (1.48)$$

Trong đó, FPR là tỉ lệ dương tính giả; FP là số mẫu âm tính bị dự đoán sai thành dương tính; và TN là số mẫu âm tính đúng.

Hình 1.8 cho biết đường cong ROC và diện tích dưới đường cong AUC. Đường cong càng tiến gần góc trên bên trái thì mô hình càng có khả năng phân biệt tốt; tương

ứng, AUC càng lớn thì khả năng xếp hạng mẫu dương cao hơn mẫu âm càng tốt [101, 103–105].



Hình 1.8: Đường cong ROC minh họa mối quan hệ giữa True Positive Rate và False Positive Rate khi thay đổi ngưỡng quyết định. Diện tích dưới đường cong, ký hiệu là AUC, phản ánh khả năng phân biệt tổng quát của mô hình.

Trong thiết lập đa lớp hoặc khi dữ liệu mất cân bằng mạnh, các độ đo tổng hợp theo kiểu macro-average được sử dụng để tránh hiện tượng lớp lớn chi phối kết quả chung [102, 108]. Macro F1 được xác định bởi Công thức (1.49).

$$\text{MacroF1} := \frac{1}{C} \sum_{c=1}^C \text{F1}_c. \quad (1.49)$$

Trong đó, MacroF1 là F1-score trung bình theo lớp; C là số lớp; và F1_c là F1-score của lớp c .

Macro AUC được xác định bởi Công thức (1.50).

$$\text{MacroAUC} := \frac{1}{C} \sum_{c=1}^C \text{AUC}_c. \quad (1.50)$$

Trong đó, MacroAUC là AUC trung bình theo lớp; C là số lớp; và AUC_c là AUC của lớp c trong thiết lập one-vs-rest.

Macro sensitivity được xác định bởi Công thức (1.51).

$$\text{MacroSens} := \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FN_c}. \quad (1.51)$$

Trong đó, MacroSens là sensitivity trung bình theo lớp; TP_c là số mẫu dương tính đúng của lớp c ; và FN_c là số mẫu của lớp c bị dự đoán sai.

Macro specificity được xác định bởi Công thức (1.52).

$$\text{MacroSpec} := \frac{1}{C} \sum_{c=1}^C \frac{TN_c}{TN_c + FP_c}. \quad (1.52)$$

Trong đó, MacroSpec là specificity trung bình theo lớp; TN_c là số mẫu âm tính đúng của lớp c ; và FP_c là số mẫu âm tính bị dự đoán sai thành lớp c .

Ngoài các độ đo trên, trong nhiều nghiên cứu hiện đại, quá trình đánh giá còn được thực hiện thông qua cross-validation nhằm giảm ảnh hưởng của việc chia dữ liệu ngẫu nhiên và tăng tính ổn định của kết quả [106, 107]. Trong các bài toán ảnh y tế với dữ liệu hạn chế, repeated stratified k-fold cross-validation thường được sử dụng để đảm bảo phân bố lớp ổn định giữa các fold và phản ánh khách quan hơn khả năng tổng quát hóa của mô hình [103, 104].

Các độ đo tổng hợp này đặc biệt phù hợp trong các bài toán phân loại tổn thương da và dữ liệu mất cân bằng, nơi hiệu năng trên lớp hiếm có ý nghĩa thực tiễn cao và không nên bị che khuất bởi lớp chiếm ưu thế.

1.7. Cơ sở hình thành hướng tiếp cận của luận án

Trên cơ sở các nội dung tổng quan đã trình bày, mục này hệ thống hóa những khoảng trống nghiên cứu liên quan trực tiếp đến hướng tiếp cận của luận án. Các khoảng trống này tập trung vào hai nhóm vấn đề: chọn lọc và hợp nhất đặc trưng đa nguồn; học bán giám sát trên đồ thị có kiểm soát độ tin cậy.

1.7.1. Các vấn đề nghiên cứu cần tiếp tục giải quyết

Các nghiên cứu gần đây trong phân loại ảnh tổn thương da cho thấy xu hướng chuyển từ mô hình ảnh đơn nguồn sang mô hình khai thác ảnh, đặc trưng học sâu, đặc trưng thủ công và dữ liệu mô tả lâm sàng [14, 28, 49, 72, 109, 110]. Tuy nhiên, vẫn còn một số vấn đề cần giải quyết.

Thứ nhất, các mô hình học sâu tạo ra nhiều bản đồ đặc trưng ở các tầng khác nhau [59–61]. Các nghiên cứu phân loại tổn thương da đã hợp nhất đặc trưng từ nhiều tầng hoặc nhiều CNN [11, 66]; các cơ chế Squeeze-and-Excitation và CBAM đã học trọng số để tái hiệu chỉnh đáp ứng theo kênh hoặc theo cả kênh và không gian [67, 68]; các phương pháp chọn lọc dựa trên entropy hoặc lượng thông tin của bản đồ đã được dùng để đánh giá bộ lọc nhằm nén mạng [69, 70]. Tuy nhiên, các nhóm phương pháp này lần lượt tập trung vào hợp nhất véc-tơ đặc trưng, gán trọng số mềm trong mạng hoặc nén kiến trúc. Vấn đề còn lại đối với luận án là chọn lọc tương đối các bản đồ trong từng block mà không sử dụng nhân tại bước tính tiêu chí, giữ nguyên backbone, rồi hợp nhất tập biểu diễn đã chọn với đặc trưng thủ công trong bối cảnh dữ liệu tổn thương da mất cân bằng. Entropy chỉ được xem là tiêu chí hỗ trợ đánh giá về phân bố kích hoạt; hiệu quả phân loại của lựa chọn này phải được kiểm chứng thực nghiệm.

Thứ hai, hợp nhất đặc trưng thủ công và đặc trưng học sâu đã được sử dụng để khai thác tính bổ trợ giữa các nguồn biểu diễn [9, 10, 111]. Tuy nhiên, nhiều phương pháp vẫn dựa trên ghép nối vector hoặc kết hợp tuyến tính. Cách tiếp cận này đơn giản nhưng chưa điều tiết đóng góp của từng nguồn đặc trưng theo chất lượng thông tin, đồng thời chưa kiểm soát tốt thành phần dư thừa hoặc nhiễu trong không gian biểu diễn [48, 72, 112]. Một số nghiên cứu đã sử dụng attention hoặc cross-attention để học quan hệ giữa các nguồn dữ liệu [17, 29], nhưng việc lựa chọn đặc trưng trước khi hợp nhất vẫn cần

được nghiên cứu thêm.

Thứ ba, dữ liệu mô tả lâm sàng như tuổi, giới tính và vị trí tổn thương là nguồn thông tin bổ trợ cho phân loại [14, 28, 109, 110]. Nguồn này thường chỉ được đưa vào mô hình như một véc-tơ phụ; vai trò của nó trong xây dựng biểu diễn nút và quan hệ giữa các mẫu trên đồ thị vẫn chưa được khai thác đầy đủ [14, 18, 19].

Thứ tư, học bán giám sát và gán nhãn giả phù hợp với ảnh y tế do dữ liệu có nhãn thường hạn chế, trong khi dữ liệu chưa gán nhãn nhiều hơn [23, 24, 95, 113]. Tuy nhiên, chất lượng nhãn giả ảnh hưởng trực tiếp đến quá trình huấn luyện lại. Nếu mẫu gán nhãn sai được đưa vào huấn luyện, sai lệch có thể lan truyền, nhất là trong dữ liệu mất cân bằng. Các nghiên cứu gần đây đã nhấn mạnh rủi ro sai lệch nhãn giả và nhu cầu kiểm soát độ tin cậy trong học bán giám sát [25, 84, 85]. Trong học bán giám sát trên đồ thị, việc đánh giá mức độ không chắc chắn và độ tin cậy của nhãn giả vẫn còn hạn chế [92, 93].

Thứ năm, mạng nơ-ron đồ thị đã được sử dụng để khai thác quan hệ giữa các mẫu [20, 22, 91]. Tuy nhiên, việc kết hợp đặc trưng học sâu, đặc trưng thủ công và dữ liệu mô tả lâm sàng trong một biểu diễn đồ thị thống nhất vẫn cần được nghiên cứu, nhất là trong học bán giám sát và dữ liệu mất cân bằng [89, 95].

Bảng 1.3: Tổng hợp các hướng học đa nguồn, học trên đồ thị và khai thác dữ liệu chưa gán nhãn

Nhóm phương pháp	Metadata	Quan hệ liên mẫu	Dữ liệu chưa nhãn	Kiểm soát nhãn giả	Đặc điểm và giới hạn
Hợp nhất ảnh và metadata [14, 28]	Có	Không	Không	Không	Bổ sung ngữ cảnh lâm sàng cho biểu diễn ảnh, nhưng chủ yếu xử lý từng mẫu độc lập.
Attention hoặc cross-attention đa nguồn [17, 18]	Có	Không	Không	Không	Học tương tác thích nghi giữa các nguồn trong từng mẫu, nhưng chưa mô hình hóa quan hệ liên mẫu.
Gán nhãn giả có kiểm soát độ tin cậy [84, 85, 113]	Tùy phương pháp	Không	Có	Có	Giảm ảnh hưởng của nhãn giả kém tin cậy, nhưng ít khai thác đồng thời cấu trúc quan hệ và biểu diễn đa nguồn.
Học bán giám sát trên đồ thị [93, 95]	Hạn chế	Có	Có	Hạn chế	Khai thác quan hệ liên mẫu, nhưng cơ chế phối hợp metadata, đặc trưng thủ công và lọc nhãn giả chưa được xem xét đồng thời.

Bảng 1.3 cho thấy bốn hướng nghiên cứu phát triển theo các trục bổ trợ nhưng chưa hội tụ hoàn toàn. Hợp nhất ảnh và metadata làm giàu ngữ cảnh cho biểu diễn của từng mẫu; attention và cross-attention cho phép điều tiết tương tác giữa các nguồn; gán nhãn giả mở rộng phạm vi khai thác dữ liệu chưa gán nhãn; còn học bán giám sát trên đồ thị đưa quan hệ liên mẫu vào quá trình lan truyền thông tin. Sự tách biệt giữa các trục này tạo ra ba giới hạn liên quan. Thứ nhất, tương tác đa nguồn thường được mô hình hóa trong phạm vi từng mẫu nên chưa phản ánh cấu trúc lân cận của dữ liệu. Thứ hai, các phương pháp gán nhãn giả có kiểm soát độ tin cậy chưa khai thác đầy đủ biểu diễn đồng

thời từ ảnh, đặc trưng thủ công và metadata trên đồ thị. Thứ ba, học bán giám sát trên đồ thị đã tận dụng quan hệ liên mẫu nhưng việc kết hợp cơ chế hợp nhất đa nguồn với lựa chọn nhân giả theo độ tin cậy vẫn còn hạn chế. Từ đó hình thành nhu cầu nghiên cứu một khung thống nhất giữa biểu diễn đa nguồn, cấu trúc đồ thị và cơ chế khai thác dữ liệu chưa gán nhãn có kiểm soát.

Từ các phân tích trên, khoảng trống thứ nhất là thiếu một cơ chế thống nhất để chọn lọc bản đồ đặc trưng theo từng block, giữ nguyên backbone, sau đó hợp nhất biểu diễn đã chọn với đặc trưng thủ công và xử lý dữ liệu mất cân bằng. Khoảng trống này không phủ nhận các nghiên cứu về attention kênh hoặc chọn lọc bộ lọc bằng entropy nhằm nén mạng, mà xác định phần chúng chưa giải quyết trong quy trình hợp nhất đặc trưng của luận án. Khoảng trống thứ hai là thiếu một khung học bán giám sát kết hợp đặc trưng ảnh, đặc trưng thủ công và metadata để mô hình hóa quan hệ giữa các mẫu trên đồ thị, đồng thời kiểm soát độ tin cậy của nhân giả. Hai khoảng trống tương ứng với hai hướng phương pháp được phát triển trong Chương 2 và Chương 3.

1.7.2. Định hướng tiếp cận của luận án

Từ các khoảng trống nghiên cứu đã phân tích, luận án xác định hai hướng tiếp cận chính.

Hướng thứ nhất sử dụng entropy kích hoạt như một tiêu chí sàng lọc không giám sát trước khi hợp nhất. Entropy mô tả mức phân tán của histogram kích hoạt và không trực tiếp biểu thị khả năng phân biệt lớp. BlockNet xác định khoảng entropy tương đối trong từng block, xây dựng biểu diễn từ tập bản đồ được giữ lại và đánh giá lựa chọn này bằng hiệu năng phân loại trong Chương 2.

Hướng thứ hai mô hình hóa dữ liệu đa nguồn bằng học trên đồ thị và học bán giám sát. Đặc trưng ảnh, đặc trưng thủ công và dữ liệu mô tả lâm sàng được dùng để xây dựng biểu diễn nút và quan hệ giữa các mẫu; mạng nơ-ron đồ thị sau đó khai thác quan hệ liên mẫu thay vì xử lý từng mẫu độc lập [20, 22, 89, 91].

Bên cạnh đó, luận án khai thác dữ liệu chưa gán nhãn thông qua gán nhãn giả có kiểm soát độ tin cậy. Mô hình ban đầu được huấn luyện trên dữ liệu có nhãn, sau đó dự đoán nhãn giả cho các mẫu chưa gán nhãn. Thay vì sử dụng toàn bộ nhãn giả, luận án lựa chọn các mẫu có độ tin cậy phù hợp nhằm hạn chế lan truyền sai lệch nhãn trong quá trình huấn luyện lại [25, 85]. Nội dung này là cơ sở cho phương pháp được trình bày trong Chương 3.

Các định hướng trên được hình thành từ quá trình phát triển nghiên cứu của Nghiên cứu sinh. [CT1] khảo sát việc kết hợp đặc trưng thủ công và đặc trưng học sâu trong phân loại ảnh; [CT2] đánh giá các mô hình CNN và ensemble learning trên dữ liệu tổn thương da. Từ nền tảng đó, [CT3] tập trung vào lựa chọn bản đồ đặc trưng theo entropy và hợp nhất đặc trưng đa nguồn; [CT4] mở rộng sang học bán giám sát trên đồ thị với cơ chế đánh giá độ tin cậy dựa trên entropy.

Kết luận Chương 1

Chương 1 đã trình bày bài toán phân loại ảnh tổn thương da với đặc trưng học sâu, đặc trưng thủ công và dữ liệu mô tả lâm sàng; đồng thời hệ thống hóa cơ sở về hợp nhất, entropy, học trên đồ thị, học bán giám sát và đánh giá mô hình.

Từ tổng quan nghiên cứu, chương đã xác định các khoảng trống chính: thiếu cơ chế lựa chọn bản đồ đặc trưng trước khi hợp nhất; hợp nhất đa nguồn còn hạn chế trong kiểm

soát thông tin dư thừa và nhiễu; metadata chưa được khai thác đầy đủ trong xây dựng quan hệ giữa các mẫu; và gán nhãn giả trong học bán giám sát cần cơ chế kiểm soát độ tin cậy.

Các phân tích này định hướng hai nội dung nghiên cứu tiếp theo của luận án. Chương 2 đề xuất phương pháp hợp nhất đặc trưng dựa trên chọn lọc entropy, còn Chương 3 trình bày mô hình học bán giám sát đa nguồn trên đồ thị cho phân loại ảnh tổn thương da.

CHƯƠNG 2: HỢP NHẤT ĐẶC TRƯNG ĐA NGUỒN DỰA TRÊN CHỌN LỌC ENTROPY CHO PHÂN LOẠI ẢNH TỔN THƯƠNG DA

Mở đầu

Các mô hình học sâu, trong đó có mạng nơ-ron tích chập (CNN), được sử dụng rộng rãi trong phân loại ảnh y tế, bao gồm ảnh tổn thương da [1, 12, 27]. Các mạng CNN tạo ra nhiều bản đồ đặc trưng ở các block khác nhau, nhưng mức đóng góp của chúng đối với bài toán phân loại không đồng đều. Việc giữ lại các phản hồi yếu, dư thừa, ít liên quan hoặc không ổn định có thể gây nhiễu và làm giảm chất lượng biểu diễn trước quá trình hợp nhất và phân loại. Vì vậy, cần đánh giá và sàng lọc các bản đồ đặc trưng trước khi xây dựng biểu diễn hợp nhất.

Bên cạnh đó, nhiều hướng nghiên cứu đã kết hợp đặc trưng học sâu với đặc trưng thủ công để khai thác tính bổ trợ giữa biểu diễn tự động và tri thức miền [8–10]. Tuy nhiên, phần lớn các phương pháp hiện nay vẫn thực hiện hợp nhất trực tiếp các vector đặc trưng thông qua các phép nối vector hoặc các phép kết hợp tuyến tính đơn giản. Khi biểu diễn đầu vào còn chứa nhiều thành phần dư thừa hoặc nhiễu, quá trình hợp nhất có thể làm tăng độ phức tạp của mô hình mà không cải thiện tương ứng khả năng phân loại. Đồng thời, dữ liệu ảnh tổn thương da thường tồn tại hiện tượng mất cân bằng lớp nghiêm trọng, trong đó các lớp thiểu số có số lượng mẫu rất hạn chế. Điều này làm cho mô hình dễ thiên lệch về lớp chiếm ưu thế và làm giảm khả năng nhận diện các tổn thương nguy hiểm.

Chương này kế thừa một phần định hướng nghiên cứu từ [CT1] về kết hợp đặc trưng thủ công và đặc trưng học sâu trong phân loại ảnh. Tuy nhiên, [CT1] chủ yếu thực hiện hợp nhất trực tiếp các vector đặc trưng và chưa xem xét việc sàng lọc từng bản đồ đặc trưng trong các block CNN trước khi hợp nhất. Vì vậy, Chương 2 phát triển mô hình BlockNet, trong đó entropy kích hoạt được dùng để mô tả độ phân tán của các giá trị trên từng bản đồ và thiết lập một quy tắc sàng lọc không giám sát; giá trị phân loại của tập bản đồ được giữ lại được đánh giá bằng thực nghiệm ở bước sau.

Khác với các phương pháp chỉ sử dụng đặc trưng từ một tầng CNN cuối hoặc hợp nhất trực tiếp toàn bộ bản đồ đặc trưng, phương pháp đề xuất khai thác đồng thời nhiều block tích chập nhằm thu nhận các mức biểu diễn khác nhau của ảnh tổn thương da. Trên cơ sở các bản đồ đặc trưng được trích xuất từ các block của mạng CNN, entropy được sử dụng như một tiêu chí định lượng để đánh giá mức độ phân bố thông tin của từng bản đồ đặc trưng. Các bản đồ đặc trưng có entropy phù hợp được lựa chọn để giảm thành phần dư thừa và nâng cao khả năng biểu diễn của đặc trưng học sâu trước khi thực hiện hợp nhất với các đặc trưng thủ công.

Bên cạnh đó, chương cũng đề xuất cơ chế cân bằng trong không gian đặc trưng nhằm cải thiện khả năng học đối với các lớp thiểu số trong dữ liệu mất cân bằng. Thay vì tăng cường trực tiếp trên ảnh đầu vào, quá trình cân bằng được thực hiện trên các biểu diễn đặc trưng đã được chọn lọc theo entropy và hợp nhất đa nguồn. Cách tiếp cận này hướng tới cải thiện chất lượng biểu diễn đầu vào trước khi thực hiện quá trình hợp nhất

và phân loại.

Nội dung chương bao gồm mô hình biểu diễn đặc trưng đa nguồn, cơ chế lựa chọn bản đồ đặc trưng dựa trên entropy, phương pháp hợp nhất đặc trưng, chiến lược cân bằng trong không gian đặc trưng và phân tích kết quả thực nghiệm của mô hình đề xuất.

Nội dung của chương liên quan đến công bố của Nghiên cứu sinh ở [CT3].

2.1. Phát biểu bài toán

Kế thừa phát biểu bài toán phân loại ảnh và các ký hiệu nền tảng đã trình bày ở Chương 1, chương này tập trung vào bài toán xây dựng biểu diễn đặc trưng đa nguồn cho phân loại ảnh tổn thương da dựa trên chọn lọc bản đồ đặc trưng theo entropy.

Trong phương pháp đề xuất, mô hình phân loại không được xây dựng trực tiếp trên ảnh đầu vào, mà trên không gian biểu diễn đặc trưng sau quá trình lựa chọn bản đồ đặc trưng và hợp nhất đa nguồn. Tập dữ liệu có nhãn được ký hiệu theo Công thức (2.1).

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N, \quad (2.1)$$

trong đó x_i là mẫu dữ liệu đầu vào thứ i , $y_i \in \mathcal{Y} = \{1, 2, \dots, C\}$ là nhãn lớp tương ứng, C là số lớp và N là số lượng mẫu.

Bộ phân loại được xây dựng trên không gian đặc trưng hợp nhất, được xác định theo Công thức (2.2):

$$g_{\omega}^{(b)} : Z \rightarrow \mathcal{Y}, \quad (2.2)$$

trong đó Z là không gian biểu diễn đặc trưng sau hợp nhất và ω là tập tham số của mô hình.

Trong chương này, mỗi mẫu dữ liệu được biểu diễn dưới dạng đa nguồn, bao gồm thông tin ảnh và các đặc trưng bổ trợ:

$$x_i = \left(x_i^{(img)}, x_i^{(h)} \right), \quad (2.3)$$

trong đó $x_i^{(img)}$ là ảnh đầu vào và $x_i^{(h)}$ là các đặc trưng thủ công như màu sắc, kết cấu và hình dạng.

Khác với các phương pháp chỉ sử dụng đặc trưng từ một tầng CNN cuối, BlockNet khai thác đồng thời nhiều block tích chập nhằm thu nhận các mức biểu diễn khác nhau của ảnh tổn thương da. Các block của mạng CNN được xác định theo Công thức (2.4).

$$\mathcal{B} = \{\text{block1_conv1}, \text{block1_conv2}, \dots, \text{block5_conv3}\}, \quad (2.4)$$

Với mỗi block $b \in \mathcal{B}$, tập bản đồ đặc trưng được xác định theo Công thức (2.5).

$$\mathcal{F}^{(b)}(x) = \{F_m^{(b)}(x)\}_{m=1}^{M_b}, \quad (2.5)$$

trong đó $F_m^{(b)}(x)$ là bản đồ đặc trưng thứ m của block b , và M_b là số lượng bản đồ đặc trưng của block đó.

Các bản đồ đặc trưng không chỉ được sử dụng như biểu diễn trung gian của CNN, mà còn được xem như các thành phần cần đánh giá và lựa chọn trước khi hợp nhất. Từ các bản đồ được trích xuất tại từng block, đặc trưng học sâu được xây dựng và kết hợp với đặc trưng thủ công để tạo biểu diễn cuối cùng của mẫu. Với mẫu x_i , trong đó $i = 1, 2, \dots, N$ và N là số lượng mẫu, biểu diễn đặc trưng hợp nhất tại block b được xác

định theo Công thức (2.6):

$$z_i^{(b)} = \psi \left(z_i^{(d,b)}, z_i^{(h)} \right), \quad (2.6)$$

trong đó $b \in \mathcal{B}$ là chỉ số block CNN được xét; $z_i^{(d,b)}$ là vector đặc trưng học sâu của mẫu x_i tại block b sau bước chọn lọc bản đồ đặc trưng; $z_i^{(h)}$ là vector đặc trưng thủ công của mẫu x_i ; và $\psi(\cdot)$ là phép hợp nhất đặc trưng.

Các bản đồ đặc trưng có phân bố kích hoạt khác nhau và không nhất thiết đóng góp như nhau cho biểu diễn hợp nhất. Dựa trên khái niệm entropy đã trình bày ở Chương 1, chương này sử dụng entropy kích hoạt như một tiêu chí không giám sát để định lượng độ phân tán của các giá trị trên bản đồ đặc trưng. Đại lượng này không sử dụng nhãn lớp và không trực tiếp đo khả năng phân biệt:

$$H_{\text{act}}(F_m^{(b)}(x)) = - \sum_{r=1}^{K_m} p_r \log_2 p_r, \quad (2.7)$$

trong đó $F_m^{(b)}(x)$ là bản đồ đặc trưng thứ m của block b , K_m là số mức giá trị kích hoạt thực sự xuất hiện và p_r là tần suất tương đối của mức thứ r . Các bản đồ có entropy nằm trong khoảng được xác định cho từng block được giữ lại để xây dựng biểu diễn học sâu trước khi hợp nhất đa nguồn.

ĐỊNH NGHĨA 2.1 [BlockNet [CT3]] Cho ảnh đầu vào x_i , với $i = 1, 2, \dots, N$, trong đó N là số lượng mẫu trong tập dữ liệu đang xét. Gọi $\mathcal{B}$ là tập các block CNN được sử dụng để trích xuất đặc trưng, và $b \in \mathcal{B}$ là chỉ số của một block cụ thể. Tại block b , tập các bản đồ đặc trưng được trích xuất từ ảnh x_i được xác định theo Công thức (2.8):

$$\mathcal{F}^{(b)}(x_i) = \left\{ F_m^{(b)}(x_i) \right\}_{m=1}^{M_b}, \quad (2.8)$$

trong đó $\mathcal{F}^{(b)}(x_i)$ là tập toàn bộ bản đồ đặc trưng của mẫu x_i tại block b ; $F_m^{(b)}(x_i)$ là bản đồ đặc trưng thứ m của mẫu x_i tại block b ; $m = 1, 2, \dots, M_b$ là chỉ số bản đồ đặc trưng; và M_b là số lượng bản đồ đặc trưng được sinh ra tại block b .

BlockNet được định nghĩa là một mô hình hợp nhất đặc trưng đa nguồn gồm bốn thành phần chính: (i) trích xuất bản đồ đặc trưng từ các block tích chập của CNN; (ii) đánh giá và lựa chọn bản đồ đặc trưng dựa trên entropy; (iii) xây dựng đặc trưng học sâu từ các bản đồ đặc trưng được lựa chọn và hợp nhất với đặc trưng thủ công; và (iv) cân bằng dữ liệu trong không gian đặc trưng để hỗ trợ quá trình phân loại.

Với mỗi block $b \in \mathcal{B}$, sau bước lựa chọn bản đồ đặc trưng dựa trên entropy, BlockNet tạo ra một biểu diễn đặc trưng hợp nhất cho mẫu x_i theo Công thức (2.9):

$$z_i^{(b)} = \psi \left(z_i^{(d,b)}, z_i^{(h)} \right), \quad (2.9)$$

trong đó $z_i^{(b)}$ là vector đặc trưng hợp nhất của mẫu x_i tại block b ; $z_i^{(d,b)}$ là vector đặc trưng học sâu của mẫu x_i tại block b , được xây dựng từ các bản đồ đặc trưng được lựa chọn sau bước đánh giá entropy, với chỉ số d biểu thị nhóm đặc trưng học sâu; $z_i^{(h)}$ là vector đặc trưng thủ công của mẫu x_i , với chỉ số h biểu thị nhóm đặc trưng thủ công; và $\psi(\cdot)$ là hàm hợp nhất đặc trưng. Trong thực nghiệm của luận án, $\psi(\cdot)$ được cài đặt bằng phép ghép nối vector nhằm kết hợp thông tin từ đặc trưng học sâu và đặc trưng thủ công.

Sau bước chọn lọc entropy, mỗi block của CNN hình thành một tập biểu diễn đặc trưng riêng phục vụ cho quá trình huấn luyện và đánh giá. Cách tiếp cận này cho phép mô hình khai thác đồng thời các mức biểu diễn khác nhau của ảnh tổn thương da, từ các đặc trưng mức thấp như màu sắc và kết cấu đến các đặc trưng ngữ nghĩa phức tạp hơn ở các block sâu.

Bên cạnh đó, dữ liệu thực tế thường có phân bố nhãn không cân bằng, trong đó một số lớp có số lượng mẫu ít hơn so với các lớp còn lại. Để giảm ảnh hưởng của hiện tượng này, chương này xem xét cơ chế cân bằng trong không gian đặc trưng thông qua việc tăng cường biểu diễn cho các lớp thiểu số sau bước trích xuất đặc trưng, thay vì tăng cường trực tiếp ở mức ảnh đầu vào.

Bài toán của chương này là xây dựng một phương pháp trích xuất, lựa chọn và hợp nhất đặc trưng đa nguồn dựa trên entropy để tạo ra các biểu diễn đặc trưng theo từng block CNN, giảm thành phần dư thừa trong biểu diễn và hỗ trợ phân loại trong điều kiện dữ liệu mất cân bằng.

2.2. Ý tưởng phương pháp đề xuất

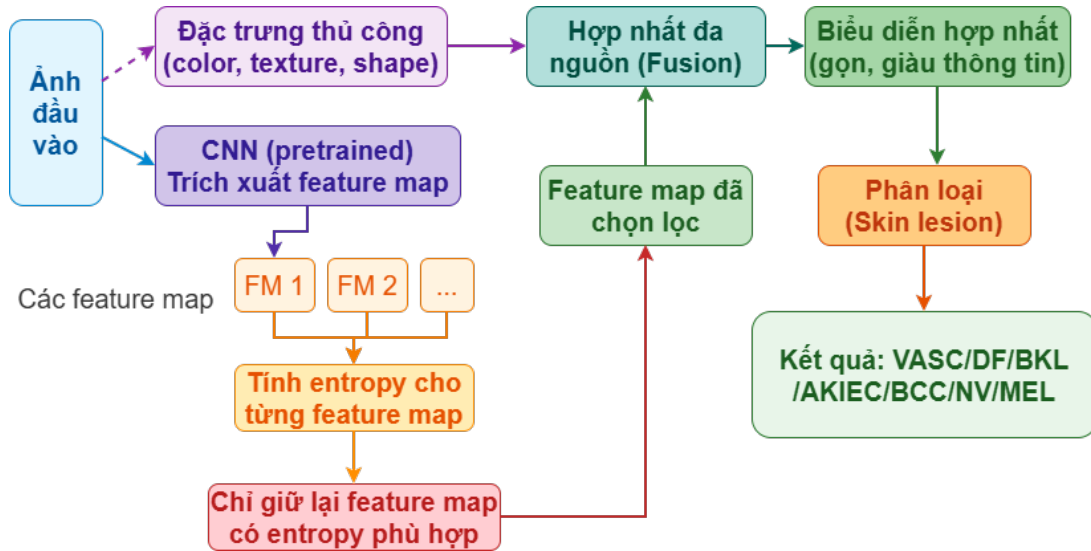
Từ phần phát biểu bài toán, hiệu năng phân loại phụ thuộc vào chất lượng biểu diễn đầu vào. Mạng CNN tạo ra nhiều bản đồ đặc trưng ở các tầng khác nhau, nhưng mức đóng góp của chúng không đồng đều. Một số bản đồ phản ánh cấu trúc tổn thương, màu sắc hoặc kết cấu, trong khi một số khác chứa phản hồi yếu, dư thừa, ít liên quan hoặc không ổn định. Việc giữ lại đồng thời các thành phần này có thể gây nhiễu và làm giảm chất lượng biểu diễn trước quá trình hợp nhất và phân loại.

Bên cạnh đó, nhiều nghiên cứu hiện nay chỉ khai thác đặc trưng từ một tầng CNN cuối hoặc hợp nhất trực tiếp toàn bộ bản đồ đặc trưng bằng các phép kết hợp đơn giản [9–11]. Cách tiếp cận này chưa tận dụng đầy đủ khả năng biểu diễn đa mức của mạng CNN. Trong thực tế, các block khác nhau của CNN thường phản ánh các mức đặc trưng khác nhau của ảnh tổn thương da: các block nông có xu hướng biểu diễn màu sắc, biên và kết cấu cục bộ, trong khi các block sâu hơn phản ánh các cấu trúc ngữ nghĩa phức tạp hơn của tổn thương [50–52]. Vì vậy, việc tổ chức và khai thác đặc trưng theo từng block CNN là cần thiết nhằm đánh giá và lựa chọn các biểu diễn phù hợp cho bài toán phân loại.

Ý tưởng chính của chương là đánh giá và chọn lọc bản đồ đặc trưng trước khi hợp nhất. Entropy được dùng để mô tả mức phân tán của các giá trị kích hoạt trong từng bản đồ, không trực tiếp đo khả năng phân biệt lớp. Entropy thấp có thể phản ánh phân bố tập trung, còn entropy cao phản ánh phân bố rộng hơn nhưng cũng có thể liên quan đến nhiễu. Vì vậy, BlockNet sử dụng khoảng entropy tương đối trong từng block như một giả thuyết sàng lọc; giá trị của tập bản đồ được giữ lại được xác nhận bằng hiệu năng phân loại trong phần thực nghiệm.

Để giải quyết các vấn đề trên, chương này đề xuất BlockNet – một phương pháp hợp nhất đặc trưng đa nguồn có chọn lọc dựa trên entropy theo từng block của mạng CNN. Ý tưởng tổng thể của phương pháp được minh họa trong Hình 2.1. Từ ảnh đầu vào, đặc trưng được xây dựng đồng thời theo hai hướng: (i) đặc trưng học sâu được trích xuất từ các bản đồ đặc trưng của các block CNN đã được huấn luyện trước và (ii) đặc trưng thủ công phản ánh màu sắc, kết cấu và hình dạng của tổn thương.

Trong BlockNet, mỗi block $b \in \mathcal{B}$ của mạng CNN tạo ra một tập bản đồ đặc trưng riêng $\mathcal{F}^{(b)}(x)$. Thay vì gộp trực tiếp toàn bộ bản đồ đặc trưng từ các tầng CNN, phương



Hình 2.1: Ý tưởng phương pháp đề xuất: chọn lọc bản đồ đặc trưng dựa trên entropy và hợp nhất đặc trưng đa nguồn.

pháp xử lý độc lập từng block để xây dựng các biểu diễn đặc trưng tương ứng cho từng mức biểu diễn của mạng CNN. Cách tổ chức này cho phép đánh giá khả năng biểu diễn của từng block thay vì chỉ sử dụng một bản đồ đặc trưng đơn lẻ hoặc một tầng CNN cuối.

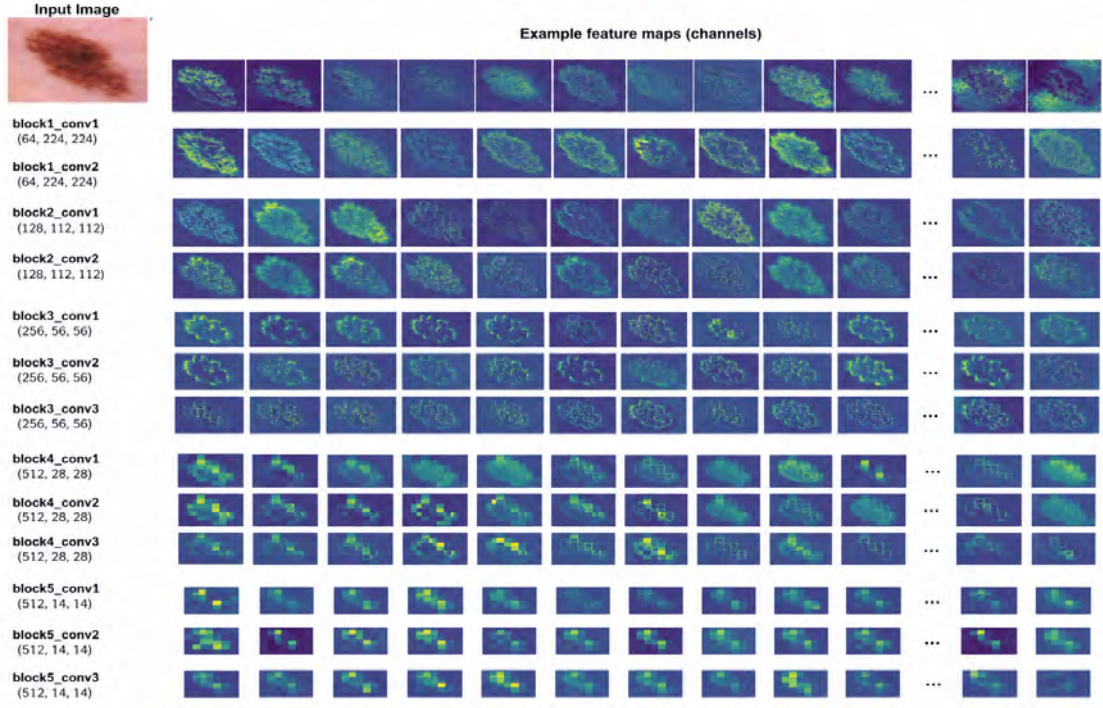
BlockNet sử dụng entropy để so sánh mức phân tán kích hoạt giữa các bản đồ trong từng block. Phương pháp không đồng nhất entropy cao với thông tin hữu ích hoặc entropy thấp với thông tin kém giá trị. Thay vào đó, các bản đồ nằm trong khoảng entropy được xác định tương đối cho từng block được giữ lại để xây dựng biểu diễn, sau đó tính hữu ích của lựa chọn này được đánh giá bằng kết quả phân loại.

Hình 2.2 minh họa các bản đồ đặc trưng được trích xuất từ các block khác nhau của mạng CNN. Mỗi ảnh đầu vào tạo ra số lượng lớn bản đồ đặc trưng với mức độ thông tin khác nhau. Một số bản đồ đặc trưng phản ánh rõ vùng tổn thương hoặc cấu trúc đặc trưng của ảnh, trong khi một số khác có phản hồi yếu hoặc chứa ít biến thiên. Điều này cho thấy sự cần thiết của việc đánh giá và lựa chọn bản đồ đặc trưng trước khi thực hiện hợp nhất đặc trưng.

Sau bước chọn lọc entropy, các bản đồ đặc trưng hợp lệ của từng block được chuyển đổi thành vector đặc trưng học sâu. Các vector này sau đó được kết hợp với các đặc trưng thủ công để tạo thành biểu diễn đa nguồn được xác định theo Công thức (2.9).

Việc hợp nhất này cho phép khai thác đồng thời khả năng học biểu diễn của mạng CNN và tri thức miền từ các đặc trưng thủ công, từ đó nâng cao khả năng phân biệt giữa các lớp tổn thương da. Đồng thời, mỗi block CNN tạo ra một tập biểu diễn đặc trưng riêng, cho phép đánh giá ảnh hưởng của các mức biểu diễn khác nhau đến kết quả phân loại.

Ngoài ra, dữ liệu ảnh tổn thương da thường tồn tại hiện tượng mất cân bằng nhãn, trong đó các lớp ác tính hoặc các lớp hiếm có số lượng mẫu thấp hơn so với các lớp phổ biến. Để giảm ảnh hưởng của hiện tượng này, phương pháp thực hiện tăng cường trong không gian đặc trưng bằng cách bổ sung các biểu diễn đặc trưng cho các lớp thiếu số dựa trên các bản đồ đặc trưng đã được lựa chọn. Quá trình này được thực hiện riêng cho từng block CNN, với $b \in \mathcal{B}$ là chỉ số block đang xét, từ đó tạo ra tập dữ liệu đặc trưng



Hình 2.2: Ví dụ các bản đồ đặc trưng được trích xuất từ các block khác nhau của mạng CNN.

cân bằng $D_{\text{balanced}}^{(b)}$.

Tập dữ liệu đặc trưng cân bằng tại block b được xác định theo Công thức (2.10):

$$D_{\text{balanced}}^{(b)} = \left\{ \left(z_i^{(b)}, y_i \right) \right\}_{i=1}^{N_b}, \quad (2.10)$$

trong đó $D_{\text{balanced}}^{(b)}$ là tập dữ liệu đặc trưng sau cân bằng tại block b ; $z_i^{(b)}$ là vector đặc trưng hợp nhất của mẫu thứ i tại block b ; y_i là nhãn lớp tương ứng của mẫu x_i ; và N_b là số lượng mẫu trong tập dữ liệu đặc trưng sau cân bằng tại block b . Tập dữ liệu $D_{\text{balanced}}^{(b)}$ được sử dụng để huấn luyện bộ phân loại tương ứng với block b . Việc cân bằng dữ liệu chỉ được thực hiện trên tập huấn luyện, không áp dụng cho tập kiểm tra, nhằm bảo đảm tính độc lập của quá trình đánh giá.

Cách tiếp cận này giúp tập trung vào các biểu diễn đặc trưng đã được đánh giá bằng entropy thay vì tạo thêm dữ liệu trực tiếp ở mức ảnh đầu vào, giảm nguy cơ khuếch đại nhiễu hoặc biến đổi không phù hợp trong dữ liệu ảnh.

Tổng thể, phương pháp đề xuất bao gồm các bước chính: (i) trích xuất bản đồ đặc trưng từ các block của mạng CNN, (ii) tính toán entropy và lựa chọn bản đồ đặc trưng hợp lệ, (iii) trích xuất đặc trưng thủ công, (iv) hợp nhất đặc trưng đa nguồn theo từng block, (v) cân bằng dữ liệu trong không gian đặc trưng, và (vi) huấn luyện bộ phân loại trên các tập biểu diễn đặc trưng tương ứng.

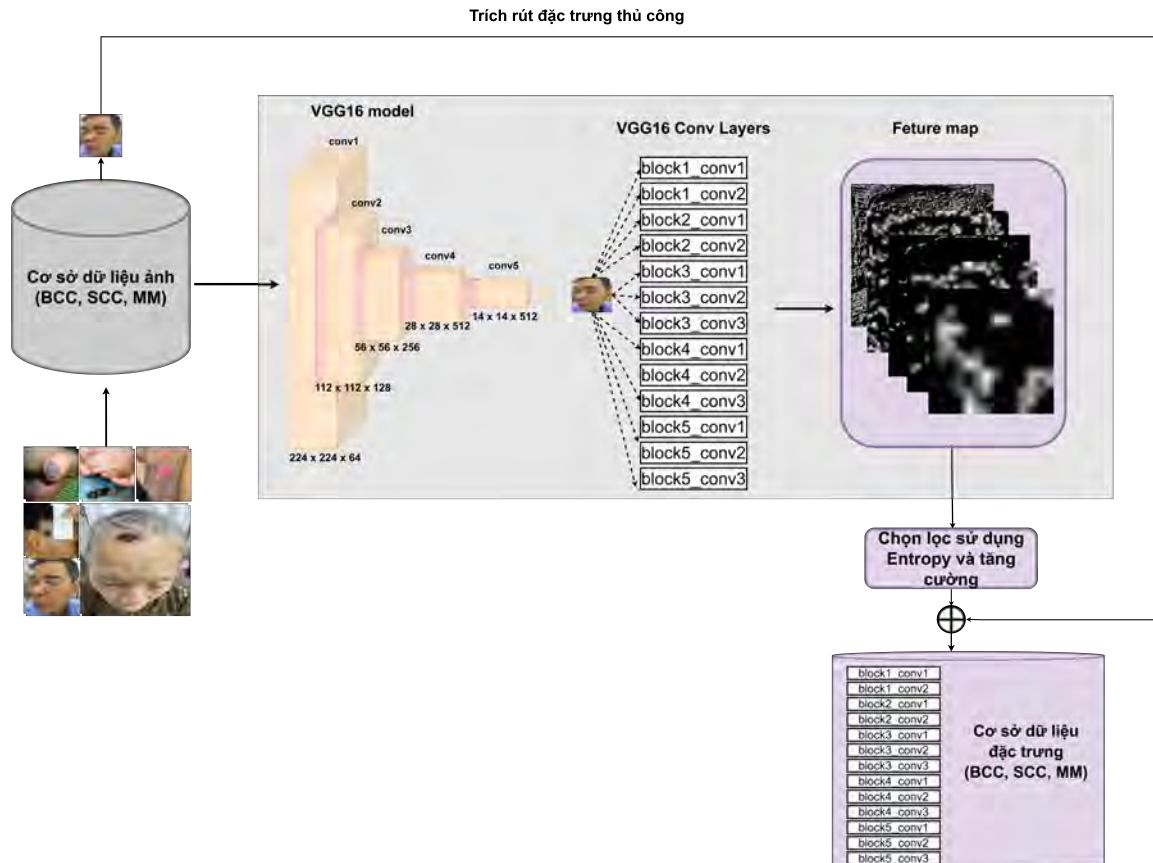
Sự khác biệt giữa BlockNet và các hướng hợp nhất trực tiếp, tái hiệu chỉnh kênh hoặc không gian, cùng các phương pháp chọn lọc bộ lọc bằng entropy nhằm nén mạng đã được tổng quan và đối chiếu trong Chương 1. Trên cơ sở khoảng trống đã xác định, phần tiếp theo tập trung trình bày thiết kế, thuật toán và kiểm chứng thực nghiệm của

BlockNet.

2.3. Phương pháp BlockNet đề xuất

2.3.1. Kiến trúc tổng quan của BlockNet

Kiến trúc tổng quan của mô hình BlockNet được minh họa trong Hình 2.3. Hình này mô tả quy trình xử lý của phương pháp, bắt đầu từ cơ sở dữ liệu ảnh đầu vào, qua mạng CNN tiền huấn luyện, trích xuất bản đồ đặc trưng tại các block tích chập, lựa chọn bản đồ đặc trưng dựa trên entropy, hợp nhất đặc trưng đa nguồn và tăng cường trong không gian đặc trưng trước khi xây dựng các cơ sở dữ liệu đặc trưng phục vụ huấn luyện và đánh giá bộ phân loại.



Hình 2.3: Kiến trúc tổng quan của mô hình BlockNet.

Với một ảnh đầu vào x , BlockNet sử dụng mạng VGG16 đã được huấn luyện trước để trích xuất đặc trưng từ các block tích chập trung gian. Khác với các phương pháp chỉ sử dụng đặc trưng từ tầng CNN cuối, BlockNet khai thác đồng thời các bản đồ đặc trưng tại nhiều block khác nhau của VGG16, từ *block1_conv1* đến *block5_conv3*. Cách tiếp cận này cho phép mô hình thu nhận đồng thời các đặc trưng mức thấp như biên, màu sắc và kết cấu, cũng như các đặc trưng mức cao hơn phản ánh cấu trúc hoặc hình dạng tổng thể.

Tập các block tích chập được sử dụng trong BlockNet đã được xác định theo Công thức (2.4). Với mỗi block $b \in \mathcal{B}$, tập bản đồ đặc trưng được trích xuất từ ảnh đầu vào x được xác định theo Công thức (2.5). Điều này cho thấy mỗi block của mạng CNN tạo ra một tập bản đồ đặc trưng riêng với khả năng biểu diễn khác nhau.

Sau bước trích xuất, các bản đồ đặc trưng tại mỗi block được đánh giá theo entropy.

Những bản đồ có entropy nằm trong khoảng phù hợp của block được giữ lại để xây dựng biểu diễn học sâu, còn các bản đồ nằm ngoài khoảng này được sàng lọc. Điều kiện lựa chọn và tập bản đồ đặc trưng thu được được xác định chính thức trong Định nghĩa 2.4. Cách trình bày này phân biệt phần mô tả kiến trúc tổng quan với phần định nghĩa toán học của bước chọn lọc.

Trong BlockNet, bước lựa chọn bản đồ đặc trưng được thực hiện trước khi xây dựng cơ sở dữ liệu đặc trưng. Nhờ đó, các đặc trưng được lưu trữ và sử dụng cho huấn luyện không phải là toàn bộ bản đồ đặc trưng thô, mà là các biểu diễn đã được chọn lọc theo tiêu chí entropy. Điều này giúp giảm thành phần dư thừa trong không gian biểu diễn, hạn chế ảnh hưởng của nhiễu và làm cho đặc trưng đầu vào gọn hơn nhưng vẫn giữ được các thành phần thông tin.

Như đã đối chiếu trong Chương 1, BlockNet khác các cơ chế attention ở chỗ không gán trọng số mềm cho toàn bộ kênh, và khác các phương pháp chọn lọc bộ lọc nhằm nén mạng ở chỗ không thay đổi backbone. Việc chọn lọc được thực hiện trên biểu diễn của từng block trước khi hợp nhất với đặc trưng thủ công.

Sau bước chọn lọc entropy, các bản đồ đặc trưng hợp lệ của từng block được chuyển đổi thành vector đặc trưng học sâu $z_i^{(d,b)}$. Bên cạnh đó, các đặc trưng thủ công như màu sắc, kết cấu và hình dạng được trích xuất để tạo thành vector $z_i^{(h)}$. Hai nguồn đặc trưng này được hợp nhất theo từng block để tạo thành biểu diễn đa nguồn theo Công thức (2.9).

Cách tổ chức này cho phép mỗi block CNN hình thành một tập biểu diễn đặc trưng riêng. Thay vì chỉ sử dụng một bản đồ đặc trưng đơn lẻ trong mỗi block, BlockNet khai thác tập các bản đồ đặc trưng hợp lệ để giữ lại nhiều thành phần thông tin bổ sung trong cùng một block. Điều này cho phép đánh giá khả năng biểu diễn của từng block CNN thông qua các tập dữ liệu đặc trưng tương ứng.

Ngoài ra, dữ liệu ảnh tổn thương da thường tồn tại hiện tượng mất cân bằng lớp, trong đó các lớp hiếm hoặc lớp ác tính có số lượng mẫu thấp hơn so với các lớp phổ biến. Để giảm ảnh hưởng của hiện tượng này, BlockNet thực hiện tăng cường trong không gian đặc trưng sau bước chọn lọc entropy và hợp nhất đặc trưng.

Với mỗi lớp k , giả sử n_k là số lượng mẫu của lớp đó và $n_{\max}$ là số lượng mẫu lớn nhất trong các lớp. Hệ số tăng cường được xác định theo Công thức (2.26).

Khi đó, số lượng biểu diễn đặc trưng sau tăng cường của lớp k được xác định theo Công thức (2.27). Do cơ chế tăng cường được thực hiện theo hệ số nhân nguyên m_k , số lượng biểu diễn sau tăng cường có thể xấp xỉ hoặc lớn hơn $n_{\max}$, thay vì bảo đảm bằng tuyệt đối giữa tất cả các lớp. Vì vậy, bước này được xem là cơ chế giảm mất cân bằng trong không gian đặc trưng, thay vì cân bằng tuyệt đối số lượng mẫu theo từng lớp. Sau bước tăng cường, mỗi block b tạo thành một cơ sở dữ liệu đặc trưng cân bằng $D_{\text{balanced}}^{(b)}$.

Các cơ sở dữ liệu đặc trưng này được sử dụng để huấn luyện và đánh giá bộ phân loại tương ứng với từng block CNN. BlockNet có khả năng đánh giá ảnh hưởng của các mức biểu diễn khác nhau đối với kết quả phân loại ảnh tổn thương da.

Tổng thể, kiến trúc BlockNet gồm các thành phần chính: (i) cơ sở dữ liệu ảnh đầu vào, (ii) mạng VGG16 tiền huấn luyện, (iii) trích xuất bản đồ đặc trưng từ các block tích chập, (iv) lựa chọn bản đồ đặc trưng dựa trên entropy, (v) hợp nhất đặc trưng đa nguồn theo từng block, (vi) tăng cường trong không gian đặc trưng, (vii) xây dựng cơ sở dữ liệu đặc trưng cân bằng, và (viii) huấn luyện bộ phân loại.

Các bước xử lý này hình thành quy trình xây dựng biểu diễn đặc trưng theo từng

block CNN và giảm thành phần dư thừa trong không gian biểu diễn.

2.3.2. Trích xuất và biểu diễn bản đồ đặc trưng từ CNN

Trong BlockNet, mạng CNN được sử dụng như một bộ trích xuất đặc trưng học sâu để khai thác các biểu diễn không gian của ảnh đầu vào. Khác với các phương pháp chỉ sử dụng đặc trưng ở tầng CNN cuối, phương pháp đề xuất khai thác bản đồ đặc trưng từ nhiều block tích chập trung gian nhằm thu nhận đồng thời các đặc trưng mức thấp như biên, màu sắc, kết cấu và các đặc trưng mức cao hơn liên quan đến cấu trúc tổng thể.

VGG16 được sử dụng làm backbone để trích xuất bản đồ đặc trưng. Việc lựa chọn VGG16 xuất phát từ một số lý do. Thứ nhất, VGG16 có kiến trúc tuần tự rõ ràng với các block tích chập được tổ chức theo cấu trúc phân cấp, thuận lợi cho việc trích xuất và phân tích bản đồ đặc trưng ở nhiều mức biểu diễn khác nhau. Thứ hai, các tầng tích chập của VGG16 sử dụng bộ lọc kích thước nhỏ 3×3 , giúp tạo ra các bản đồ đặc trưng ổn định và dễ phân tích theo từng block. Thứ ba, VGG16 đã được huấn luyện trước trên tập dữ liệu lớn nên có khả năng cung cấp các biểu diễn thị giác tổng quát, phù hợp trong điều kiện dữ liệu ảnh y tế có số lượng mẫu hạn chế. Ngoài ra, mục tiêu của chương này không phải là đề xuất một backbone CNN mới, mà là đề xuất cơ chế lựa chọn và hợp nhất bản đồ đặc trưng dựa trên entropy; do đó một kiến trúc CNN chuẩn, ổn định và dễ diễn giải như VGG16 là phù hợp để kiểm chứng ý tưởng đề xuất.

Trong BlockNet, đặc trưng học sâu không được xây dựng từ một lớp tích chập duy nhất mà được tổ chức theo từng block của VGG16. Tập các lớp tích chập được sử dụng trong nghiên cứu đã được xác định theo Công thức (2.4).

Với mỗi block $b \in \mathcal{B}$, tập bản đồ đặc trưng được xác định theo Công thức (2.5). Mỗi block CNN tạo ra một tập bản đồ đặc trưng riêng với khả năng biểu diễn khác nhau.

Sau bước lựa chọn entropy và cân bằng dữ liệu trong không gian đặc trưng, với mỗi block $b \in \mathcal{B}$, BlockNet hình thành một tập dữ liệu đặc trưng cân bằng $D_{\text{balanced}}^{(b)}$. Tập này gồm các cặp $(z_i^{(b)}, y_i)$, trong đó $z_i^{(b)}$ là vector đặc trưng hợp nhất của mẫu x_i tại block b , và y_i là nhãn lớp tương ứng.

Cách tổ chức này cho phép BlockNet đánh giá khả năng biểu diễn của từng block CNN thông qua các tập dữ liệu đặc trưng tương ứng, thay vì chỉ sử dụng một bản đồ đặc trưng đơn lẻ hoặc chỉ khai thác đặc trưng ở tầng CNN cuối.

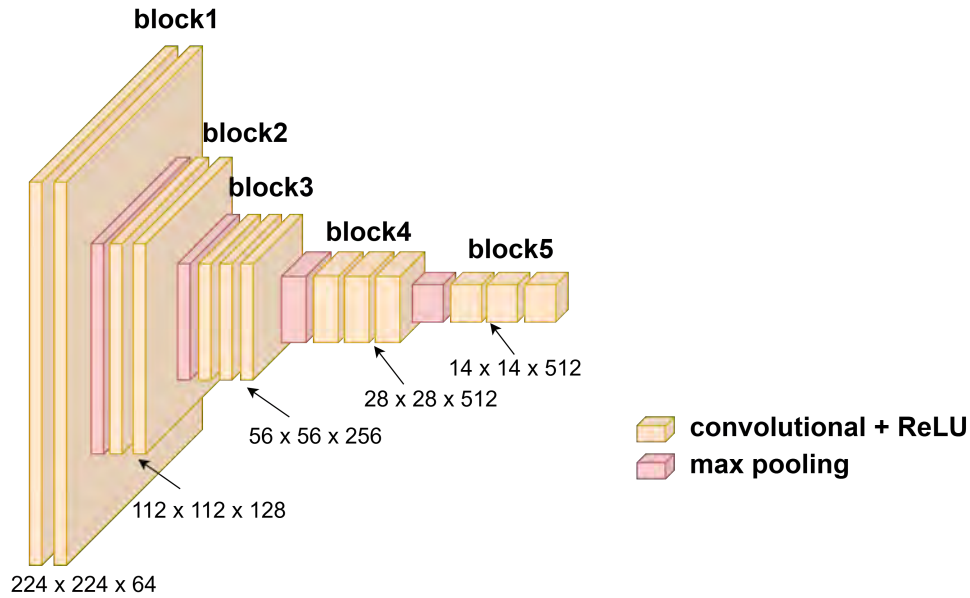
Kiến trúc phân cấp của các block tích chập trong VGG16 được minh họa trong Hình 2.4. Hình 2.4 cho thấy kích thước không gian của bản đồ đặc trưng giảm dần theo độ sâu của mạng, trong khi số lượng kênh đặc trưng tăng lên. Điều này phản ánh quá trình chuyển đổi từ các biểu diễn mức thấp ở các block đầu sang các biểu diễn có tính trừu tượng cao hơn ở các block sâu.

Bảng 2.1 tóm tắt các block của VGG16 được sử dụng trong BlockNet, bao gồm số lớp tích chập, kích thước đầu ra và số lượng bản đồ đặc trưng tương ứng.

Với một ảnh đầu vào x , mỗi lớp tích chập trong VGG16 tạo ra một tập các bản đồ đặc trưng. Về mặt hình thức, bản đồ đặc trưng thứ m của block b được xác định theo Công thức (2.11):

$$F_m^{(b)}(x) = \sigma(W_m^{(b)} * x + b_m^{(b)}), \quad (2.11)$$

trong đó $W_m^{(b)}$ là bộ lọc tích chập thứ m của block b , $b_m^{(b)}$ là hệ số bias, ký hiệu $*$ biểu diễn phép tích chập và $\sigma(\cdot)$ là hàm kích hoạt phi tuyến.



Hình 2.4: Cấu trúc phân cấp các block tích chập của VGG16 được sử dụng để trích xuất bản đồ đặc trưng.

Bảng 2.1: Các block tích chập của VGG16 được sử dụng để trích xuất bản đồ đặc trưng.

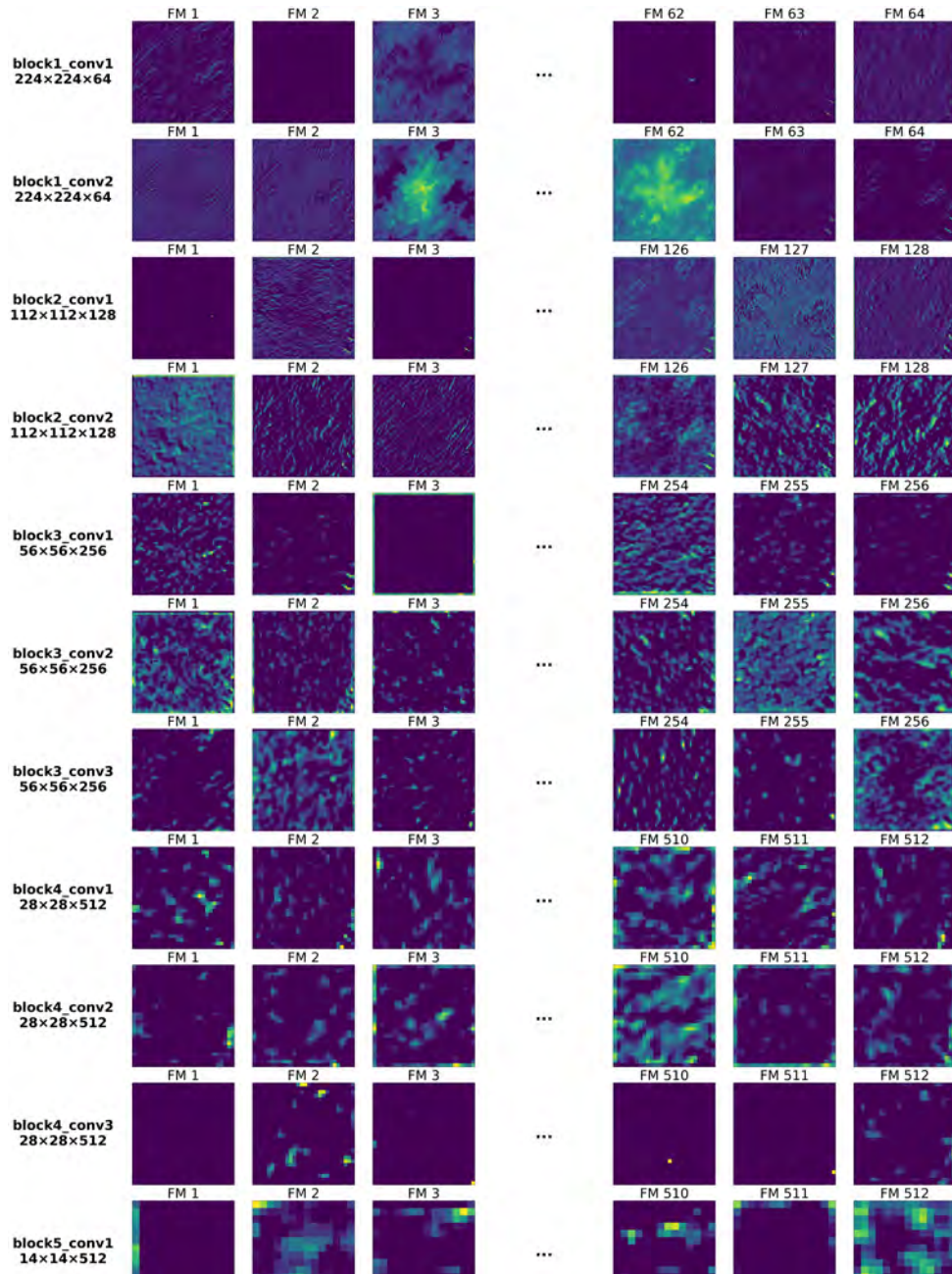
Block	Các lớp sử dụng	Số lớp	Kích thước đầu ra	Số bản đồ đặc trưng
block1	block1_conv1, block1_conv2	2	224×224	64
block2	block2_conv1, block2_conv2	2	112×112	128
block3	block3_conv1, block3_conv2, block3_conv3	3	56×56	256
block4	block4_conv1, block4_conv2, block4_conv3	3	28×28	512
block5	block5_conv1, block5_conv2, block5_conv3	3	14×14	512

Tập bản đồ đặc trưng được trích xuất từ block b đã được xác định theo Công thức (2.5). Điều này cho phép biểu diễn đặc trưng được tổ chức theo từng block CNN thay vì gom toàn bộ bản đồ đặc trưng vào một tập duy nhất.

Hình 2.5 minh họa các bản đồ đặc trưng thực tế được trích xuất từ toàn bộ các lớp tích chập được sử dụng trong VGG16. Mỗi dòng tương ứng với một lớp tích chập, trong đó một số bản đồ đặc trưng đầu và cuối được hiển thị; dấu “...” biểu diễn các bản đồ đặc trưng trung gian không được trình bày để tránh làm hình quá dày.

Từ Hình 2.5, có thể quan sát thấy rằng các lớp đầu như *block1_conv1* và *block1_conv2* thường tạo ra các bản đồ đặc trưng có độ phân giải không gian cao, phản ánh các đặc trưng mức thấp như biên, màu sắc và kết cấu cục bộ. Các lớp ở giữa như *block3_conv1*, *block3_conv2* và *block3_conv3* bắt đầu biểu diễn các mẫu hình cục bộ phức tạp hơn. Trong khi đó, các lớp sâu hơn như *block4_conv** và *block5_conv** có kích thước không gian nhỏ hơn nhưng biểu diễn các đặc trưng trừu tượng hơn, liên quan đến cấu trúc tổng thể hoặc vùng phản hồi của tổn thương.

Hình minh họa cho thấy các bản đồ đặc trưng có kiểu phản hồi khác nhau. Một số bản đồ thể hiện vùng có cấu trúc, trong khi một số khác gần như phẳng, có phản hồi yếu



Hình 2.5: Ví dụ các bản đồ đặc trưng được trích xuất từ các lớp tích chập của VGG16. Mỗi dòng biểu diễn một lớp tích chập, gồm một số bản đồ đặc trưng đầu và cuối; dấu “...” biểu diễn các bản đồ đặc trưng trung gian không hiển thị.

hoặc ít biến thiên. Quan sát thị giác này chỉ mô tả sự khác biệt của đáp ứng, không đủ để kết luận trực tiếp khả năng phân biệt lớp. Tuy nhiên, việc giữ các phản hồi yếu, dư thừa hoặc không ổn định có thể gây nhiễu biểu diễn, vì vậy cần đánh giá chúng trước bước hợp nhất.

Sau khi được trích xuất, các bản đồ đặc trưng của từng block được chuyển đổi thành vector đặc trưng học sâu để phục vụ cho các bước lựa chọn, hợp nhất và phân loại. Biểu diễn đặc trưng học sâu của block b đối với ảnh x được xác định theo Công thức (2.12):

$$z^{(d,b)} = \phi \left(\mathcal{F}^{(b)}(x) \right), \quad (2.12)$$

trong đó $\phi(\cdot)$ là hàm chuyển đổi tập bản đồ đặc trưng của block b sang vector đặc trưng học sâu.

Trong BlockNet, quá trình này không dừng lại ở việc sử dụng toàn bộ bản đồ đặc trưng, mà tiếp tục đo độ phân tán kích hoạt của từng bản đồ bằng entropy trước khi xây dựng biểu diễn hợp nhất cuối cùng. Sau bước sàng lọc theo khoảng entropy, mỗi block CNN hình thành một tập biểu diễn riêng để huấn luyện và đánh giá bộ phân loại; chính kết quả phân loại được dùng để kiểm chứng quy tắc sàng lọc.

Như vậy, quá trình trích xuất và biểu diễn bản đồ đặc trưng từ CNN đóng vai trò nền tảng trong BlockNet. Nó tạo ra các không gian biểu diễn theo từng block CNN với mức độ trừu tượng khác nhau. Tuy nhiên, do số lượng bản đồ đặc trưng lớn và chất lượng thông tin không đồng đều, cần có cơ chế xác định khoảng entropy phù hợp để lựa chọn các bản đồ đặc trưng hữu ích trước khi thực hiện hợp nhất đặc trưng. Nội dung này được trình bày trong mục tiếp theo.

2.3.3. Xác định khoảng entropy phù hợp

Sau khi trích xuất bản đồ đặc trưng từ các block tích chập của CNN, cần xem xét sự khác nhau giữa các kiểu phản hồi. Một số bản đồ gần như đồng nhất và ít biến thiên; một số khác có kích hoạt phân tán hoặc không ổn định. Các hiện tượng này không trực tiếp xác định khả năng phân biệt lớp, nhưng cung cấp cơ sở để xây dựng giả thuyết sàng lọc. Thay vì chỉ chọn các bản đồ có entropy lớn nhất, BlockNet xác định một khoảng entropy tương đối cho từng block và kiểm chứng tập bản đồ được giữ lại bằng hiệu năng phân loại.

Trong BlockNet, entropy được sử dụng như một độ đo phản ánh mức độ đa dạng trong phân bố giá trị của bản đồ đặc trưng. Entropy được tính riêng cho từng bản đồ đặc trưng của từng block CNN. Định nghĩa 2.2 trình bày cách xác định entropy của một bản đồ đặc trưng.

ĐỊNH NGHĨA 2.2 [Entropy của bản đồ đặc trưng [CT3]] Từ Định nghĩa 1.2 ở Chương 1, xét bản đồ đặc trưng $F_m^{(b)}(x)$ gồm N_m phần tử. Gọi K_m là số mức giá trị kích hoạt khác nhau thực sự xuất hiện, n_r là số lần xuất hiện của mức thứ r và $p_r = n_r/N_m$. Trong quá trình thực nghiệm, entropy được tính bằng hàm `shannon_entropy(feature_map)` với cơ số mặc định bằng 2. Do đó, entropy kích hoạt được xác định:

$$H_{\text{act}}(F_m^{(b)}(x)) = - \sum_{r=1}^{K_m} p_r \log_2 p_r, \quad (2.13)$$

Entropy có đơn vị bit và thỏa mãn:

$$0 \leq H_{\text{act}}(F_m^{(b)}(x)) \leq \log_2 K_m \leq \log_2 N_m. \quad (2.14)$$

Entropy tại đây phản ánh mức độ phân tán của các giá trị kích hoạt trong một bản đồ đặc trưng. Entropy thấp cho biết các giá trị tập trung vào ít bin hơn, còn entropy cao cho biết phân bố rộng hơn. Vì histogram này không chứa nhãn lớp, entropy kích hoạt

không phải độ đo độ chính xác, độ ổn định hay khả năng phân biệt lớp của bản đồ; bản đồ entropy cao có thể chứa nhiều và bản đồ entropy thấp vẫn có thể hữu ích. Do đó, khoảng entropy của BlockNet chỉ là giả thuyết sàng lọc, được kiểm chứng ở mức tập hợp bằng hiệu năng phân loại.

VÍ DỤ 2.1 [So sánh entropy giữa hai bản đồ đặc trưng đơn giản] Xét hai bản đồ đặc trưng đơn giản F_A và F_B được minh họa trong Hình 2.6.

120	120	120	120
120	120	120	120
120	120	120	120
120	120	120	120

(a) Bản đồ đặc trưng F_A

0	40	80	120
160	200	240	255
30	90	150	210
15	75	135	195

(b) Bản đồ đặc trưng F_B

Hình 2.6: Hai bản đồ đặc trưng minh họa cho quá trình tính entropy.

Bản đồ đặc trưng F_A có toàn bộ giá trị đều bằng 120. Khi đó, xác suất xuất hiện của mức giá trị 120 là:

$$p(120) = 1. \quad (2.15)$$

Entropy cơ sở 2 của bản đồ đặc trưng F_A được tính:

$$\begin{aligned} H(F_A) &= -(1 \times \log_2(1)) \\ &= -(1 \times 0) \\ &= 0. \end{aligned} \quad (2.16)$$

Theo Công thức (2.16), bản đồ đặc trưng F_A có entropy bằng 0 bit do toàn bộ giá trị đều giống nhau.

Đối với bản đồ đặc trưng F_B , tất cả 16 giá trị đều khác nhau. Do đó, xác suất xuất hiện của mỗi giá trị là:

$$p_j = \frac{1}{16}. \quad (2.17)$$

Mỗi thành phần entropy cơ sở 2 được tính:

$$\begin{aligned} p_j \log_2(p_j) &= \frac{1}{16} \log_2\left(\frac{1}{16}\right) \\ &= -\frac{\log_2(16)}{16} \\ &\approx -0.25. \end{aligned} \quad (2.18)$$

Do có 16 thành phần giống nhau nên entropy của bản đồ đặc trưng F_B là:

$$\begin{aligned} H(F_B) &= - \left(16 \times \left(-\frac{\log_2(16)}{16} \right) \right) \\ &= \log_2(16) = 4 \text{ bit.} \end{aligned} \quad (2.19)$$

Kết quả trong Công thức (2.19) cho thấy bản đồ đặc trưng F_B có entropy cao hơn F_A , nghĩa là các giá trị của F_B phân tán trên nhiều mức hơn. Kết quả này không cho phép kết luận F_B chứa nhiều thông tin phân biệt lớp hơn F_A .

Ví dụ 2.1 chỉ minh họa cách entropy phản ánh độ phân tán của các giá trị kích hoạt. Trong BlockNet, đại lượng này được dùng để thiết lập tập bản đồ ứng viên; mức độ phù hợp của biểu diễn sau sàng lọc được xác định bằng kết quả phân loại.

Trong BlockNet, khoảng entropy được xác định riêng cho từng block CNN vì các block tạo ra các phân bố kích hoạt khác nhau. Các block đầu thường duy trì độ phân giải không gian cao, trong khi các block sâu hơn tạo phản hồi trừu tượng hơn. Vì vậy, một khoảng entropy chung cho toàn bộ các block sẽ không phản ánh đúng sự khác nhau về thang và độ phân tán kích hoạt giữa các nhóm bản đồ.

ĐỊNH NGHĨA 2.3 [Khoảng entropy phù hợp [CT3]]

Cho block $b \in \mathcal{B}$, trong đó $\mathcal{B}$ là tập các block CNN được sử dụng để trích xuất đặc trưng. Với bản đồ đặc trưng $F_m^{(b)}(x_i)$ của mẫu x_i tại block b , giá trị entropy tương ứng được ký hiệu là $H(F_m^{(b)}(x_i))$. Gọi $\theta^{(b)}$ là ngưỡng entropy tối thiểu của block b . Trước hết, các giá trị entropy nhỏ hơn hoặc bằng $\theta^{(b)}$ được loại bỏ để thu được tập entropy hợp lệ, được xác định theo Công thức (2.20):

$$E^{(b)} = \left\{ H(F_m^{(b)}(x_i)) \mid H(F_m^{(b)}(x_i)) > \theta^{(b)}, i = 1, 2, \dots, N, m = 1, 2, \dots, M_b \right\}, \quad (2.20)$$

trong đó $E^{(b)}$ là tập các giá trị entropy hợp lệ của block b ; N là số lượng mẫu trong tập dữ liệu đang xét; M_b là số lượng bản đồ đặc trưng tại block b ; m là chỉ số bản đồ đặc trưng; và $\theta^{(b)}$ là ngưỡng entropy tối thiểu dùng để loại bỏ các bản đồ đặc trưng có mức phản hồi quá thấp.

Gọi $\mu_e^{(b)}$ và $\sigma_e^{(b)}$ lần lượt là giá trị trung bình và độ lệch chuẩn của các giá trị entropy trong $E^{(b)}$. Khi đó, khoảng entropy phù hợp của block b được xác định bởi $[e_{\min}^{(b)}, e_{\max}^{(b)}]$, trong đó $e_{\min}^{(b)}$ và $e_{\max}^{(b)}$ lần lượt là cận dưới và cận trên của khoảng entropy phù hợp:

$$e_{\min}^{(b)} = \max \left(\min(E^{(b)}), \mu_e^{(b)} - \sigma_e^{(b)} \right), \quad (2.21)$$

$$e_{\max}^{(b)} = \min \left(\max(E^{(b)}), \mu_e^{(b)} + \sigma_e^{(b)} \right). \quad (2.22)$$

Trong đó, $\mu_e^{(b)}$ phản ánh mức entropy trung bình của các bản đồ đặc trưng hợp lệ tại block b , còn $\sigma_e^{(b)}$ phản ánh mức độ phân tán của các giá trị entropy này. Khoảng $[e_{\min}^{(b)}, e_{\max}^{(b)}]$ giữ lại các bản đồ có entropy nằm trong vùng được xác định tương đối từ phân bố entropy tại block b , đồng thời sàng lọc các giá trị ở hai phía ngoài khoảng này.

Việc nằm ngoài khoảng entropy không tự nó chứng minh một bản đồ không có khả năng phân biệt lớp. Entropy thấp cho biết phân bố kích hoạt tập trung hơn, còn entropy

cao cho biết phân bố rộng hơn và trong một số trường hợp có thể đi kèm nhiều. Vì vậy, khoảng trên là tiêu chí sàng lọc của BlockNet; tính hữu ích của các bản đồ được giữ lại phải được xác nhận bằng thực nghiệm phân loại.

Cần lưu ý rằng khoảng entropy phù hợp trong Định nghĩa 2.3 không được hiểu là một miền tối ưu tuyệt đối cho mọi tập dữ liệu hoặc mọi kiến trúc mạng. Khoảng này được xác định theo phân bố entropy của từng block CNN trong thiết lập thực nghiệm, nhằm sàng lọc các bản đồ có phân bố kích hoạt quá tập trung hoặc quá phân tán trước khi xây dựng biểu diễn học sâu.

Trong Định nghĩa 2.3, ngưỡng $\theta^{(b)}$ có vai trò sàng lọc các bản đồ có phân bố kích hoạt gần như đồng nhất trong block b . Giá trị $\mu_e^{(b)}$ phản ánh trung tâm của phân bố entropy hợp lệ, trong khi $\sigma_e^{(b)}$ phản ánh độ phân tán của các giá trị entropy. Việc sử dụng khoảng:

$$[\mu_e^{(b)} - \sigma_e^{(b)}, \mu_e^{(b)} + \sigma_e^{(b)}]$$

giúp giữ lại các bản đồ đặc trưng có mức thông tin ổn định, đồng thời loại bỏ các phản hồi quá thấp hoặc quá bất thường.

VÍ DỤ 2.2 [Xác định khoảng entropy phù hợp từ phân bố thực nghiệm] Xét các bản đồ đặc trưng được trích xuất từ lớp *block1_conv1* của mạng VGG16 đối với ảnh đầu vào *2_Basal Cell Carcinoma_Picture1.jpg*. Lớp *block1_conv1* tạo ra 64 bản đồ đặc trưng có kích thước 224×224 .

Hình 2.7 minh họa ba bản đồ đặc trưng đầu và ba bản đồ đặc trưng cuối; dấu “.....” biểu diễn các bản đồ đặc trưng trung gian không hiển thị.



Hình 2.7: Minh họa một số bản đồ đặc trưng đầu và cuối được trích xuất từ lớp *block1_conv1* của VGG16 đối với một ảnh đầu vào.

Sau khi tính entropy cho toàn bộ 64 bản đồ đặc trưng của lớp *block1_conv1*, ta thu được phân bố entropy như minh họa trong Hình 2.8.

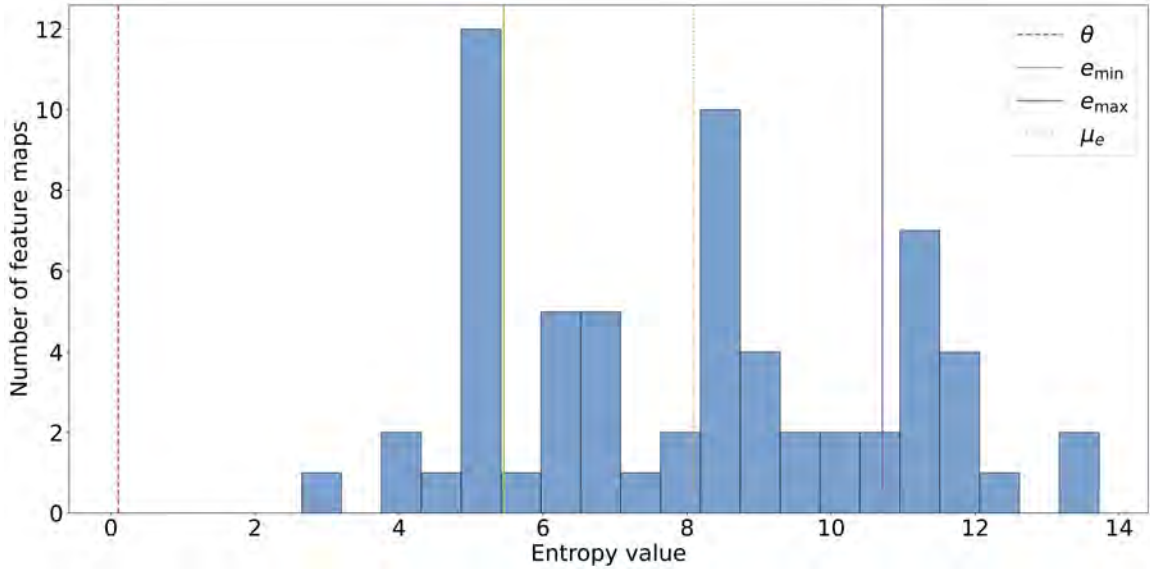
Trong Hình 2.8, đường nét đứt biểu diễn ngưỡng tối thiểu $\theta^{(b)}$, dùng để loại bỏ các bản đồ đặc trưng có entropy quá thấp. Đường chấm biểu diễn giá trị trung bình $\mu_e^{(b)}$ của các entropy hợp lệ. Hai đường còn lại biểu diễn $e_{\min}^{(b)}$ và $e_{\max}^{(b)}$, là cận dưới và cận trên của khoảng entropy phù hợp đối với block đang xét.

Từ phân bố này, các bản đồ đặc trưng được lựa chọn nếu entropy của chúng nằm trong khoảng:

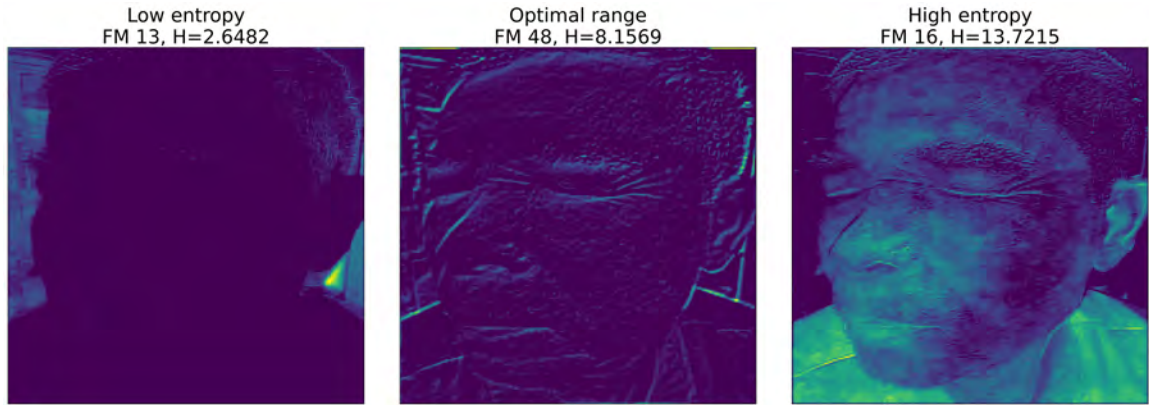
$$e_{\min}^{(b)} \leq H(F_m^{(b)}(x)) \leq e_{\max}^{(b)}. \quad (2.23)$$

Hình 2.9 minh họa ba bản đồ đặc trưng đại diện: một bản đồ đặc trưng có entropy thấp, một bản đồ đặc trưng nằm trong khoảng entropy phù hợp và một bản đồ đặc trưng có entropy cao.

Mỗi bản đồ của *block1_conv1* có $N_m = 224 \times 224 = 50\,176$ phần tử, nên entropy cơ



Hình 2.8: Phân bố entropy của các bản đồ đặc trưng tại lớp *block1_conv1* và các ngưỡng dùng để xác định khoảng entropy phù hợp.



Hình 2.9: Ba bản đồ đặc trưng đại diện theo các mức entropy khác nhau: thấp, nằm trong khoảng phù hợp và cao.

số 2 thỏa mãn

$$H_{\text{act}} \leq \log_2(50176) \approx 15.6147 \text{ bit.}$$

Do đó, các giá trị $H = 2.6482$, $H = 8.1569$ và $H = 13.7215$ đều thuộc miền hợp lệ. Các giá trị này lần lượt biểu thị phân bố kích hoạt tập trung hơn, trung gian và phân tán hơn; chúng không trực tiếp xác định khả năng phân biệt lớp.

CHÚ Ý 2.1 Hình 2.8 biểu diễn phân bố entropy của các bản đồ đặc trưng được trích xuất từ lớp *block1_conv1*. Trục hoành biểu diễn giá trị entropy, trong khi trục tung biểu diễn số lượng bản đồ đặc trưng tương ứng với từng khoảng entropy.

Phân bố này cho thấy độ phân tán kích hoạt của các bản đồ đặc trưng không đồng đều: một số bản đồ có entropy rất thấp, trong khi một số khác có entropy rất cao. Sự khác nhau này chưa tự nó xác định giá trị phân biệt lớp của từng bản đồ.

Ngưỡng $\theta^{(b)}$ được sử dụng để sàng lọc các bản đồ có entropy rất thấp. Các bản đồ dưới ngưỡng thường có phân bố kích hoạt tập trung, gần như đồng nhất hoặc chỉ phản hồi ở một số vùng nhỏ. Đây là dấu hiệu về kiểu phân bố kích hoạt, không phải kết luận

trực tiếp về khả năng phân biệt lớp; tác động của việc sàng lọc được đánh giá bằng thực nghiệm.

Các bản đồ đặc trưng có entropy nằm trong khoảng:

$$\theta^{(b)} \leq H(F_m^{(b)}(x)) < e_{\min}^{(b)}$$

được xem là các bản đồ đặc trưng có mức thông tin thấp nhưng bắt đầu xuất hiện các phản hồi cục bộ. Tuy nhiên, các phản hồi này vẫn còn hạn chế và chưa đủ ổn định để biểu diễn rõ cấu trúc của đối tượng.

Khoảng entropy phù hợp được xác định trong miền:

$$e_{\min}^{(b)} \leq H(F_m^{(b)}(x)) \leq e_{\max}^{(b)}.$$

Các bản đồ đặc trưng trong khoảng này thường thể hiện sự cân bằng giữa mức độ phản hồi và khả năng biểu diễn cấu trúc. Chúng giữ lại được các đặc điểm như biên, kết cấu hoặc vùng tổn thương mà không bị phân tán quá mức. Đây là nhóm bản đồ đặc trưng mang tính biểu diễn kết quả nhất cho bài toán phân loại.

Ngược lại, các bản đồ có entropy lớn hơn $e_{\max}^{(b)}$ có phân bố kích hoạt rộng hơn. Entropy cao không đồng nghĩa với hữu ích cho phân loại và trong một số trường hợp có thể đi kèm nhiều hoặc phản hồi không đặc hiệu. Vì vậy, các bản đồ này được sàng lọc theo giả thuyết của BlockNet và tác động của quyết định đó được kiểm chứng trong phần thực nghiệm.

Do đó, khoảng entropy phù hợp không phải là vùng có entropy lớn nhất, mà là vùng cân bằng giữa khả năng biểu diễn thông tin và mức độ ổn định của bản đồ đặc trưng.

Do đó, BlockNet không lựa chọn bản đồ đặc trưng theo nguyên tắc entropy càng lớn càng tốt. Phương pháp xác định một khoảng entropy tương đối cho từng block để tạo tập bản đồ ứng viên, thay vì khẳng định trước khả năng phân biệt của từng bản đồ.

Từ khoảng entropy phù hợp của từng block CNN, một bản đồ đặc trưng được giữ lại nếu thỏa mãn điều kiện trong Công thức (2.23). Điều kiện này sàng lọc các bản đồ có phân bố kích hoạt nằm ngoài khoảng đã xác định. Chất lượng của tập bản đồ được lựa chọn đối với hợp nhất và phân loại được đánh giá bằng thực nghiệm, không được suy ra trực tiếp từ entropy.

Khoảng entropy phù hợp đóng vai trò là cơ chế lọc thông tin trong BlockNet. Cơ chế này không chỉ giúp giảm số lượng bản đồ đặc trưng dư thừa mà còn nâng cao chất lượng biểu diễn đầu vào bằng cách giữ lại các bản đồ đặc trưng có mức thông tin phù hợp trong từng block CNN. Đây là bước trung gian trước khi thực hiện lựa chọn bản đồ đặc trưng và hợp nhất với các đặc trưng thủ công.

2.3.4. Chọn lọc bản đồ đặc trưng dựa trên entropy

Sau khi xác định khoảng entropy phù hợp cho từng block CNN, bước tiếp theo của BlockNet là sử dụng khoảng này để lựa chọn các bản đồ đặc trưng phục vụ quá trình xây dựng biểu diễn đặc trưng. Mục tiêu của bước chọn lọc không phải là giữ lại các bản đồ đặc trưng có entropy lớn nhất, mà là loại bỏ các bản đồ đặc trưng có mức phản hồi quá yếu hoặc quá phân tán, từ đó giảm thành phần dư thừa và hạn chế các phản hồi không ổn định trước khi hợp nhất đặc trưng.

Với mỗi block $b \in \mathcal{B}$, tập bản đồ đặc trưng được trích xuất từ mẫu đầu vào x_i được ký hiệu là $\mathcal{F}^{(b)}(x_i) = \{F_m^{(b)}(x_i)\}_{m=1}^{M_b}$. Entropy của từng bản đồ đặc trưng $F_m^{(b)}(x_i)$ được

tính theo Công thức (2.13). Dựa trên khoảng entropy phù hợp $[e_{\min}^{(b)}, e_{\max}^{(b)}]$ đã xác định cho block b , các bản đồ đặc trưng được giữ lại được xác định theo Định nghĩa 2.4.

ĐỊNH NGHĨA 2.4 [Bản đồ đặc trưng hợp lệ [CT3]]

Cho tập bản đồ đặc trưng $\mathcal{F}^{(b)}(x_i)$ của mẫu x_i tại block b . Tập các bản đồ đặc trưng hợp lệ sau bước chọn lọc entropy được xác định theo Công thức (2.24):

$$\mathcal{F}^{*(b)}(x_i) = \left\{ F_m^{(b)}(x_i) \in \mathcal{F}^{(b)}(x_i) \mid e_{\min}^{(b)} \leq H(F_m^{(b)}(x_i)) \leq e_{\max}^{(b)} \right\}. \quad (2.24)$$

Trong đó, $\mathcal{F}^{*(b)}(x_i)$ là tập các bản đồ đặc trưng được giữ lại của mẫu x_i tại block b sau bước chọn lọc entropy; $F_m^{(b)}(x_i)$ là bản đồ đặc trưng thứ m tại block b ; $H(F_m^{(b)}(x_i))$ là entropy của bản đồ đặc trưng đó; $e_{\min}^{(b)}$ và $e_{\max}^{(b)}$ lần lượt là cận dưới và cận trên của khoảng entropy phù hợp đã xác định cho block b .

Định nghĩa 2.4 xác định bước chọn lọc entropy ở mức bản đồ đặc trưng của từng block CNN. Các bản đồ đặc trưng có entropy nhỏ hơn $e_{\min}^{(b)}$ thường có phản hồi yếu, ít biến thiên hoặc chỉ biểu diễn một số vùng nhỏ chưa đủ ổn định. Ngược lại, các bản đồ đặc trưng có entropy lớn hơn $e_{\max}^{(b)}$ có thể có phản hồi quá phân tán trên ảnh, chứa nhiễu, vùng nền hoặc các phản hồi không đặc hiệu. Do đó, BlockNet lựa chọn các bản đồ đặc trưng nằm trong khoảng entropy phù hợp, thay vì lựa chọn theo nguyên tắc entropy càng lớn càng cao.

VÍ DỤ 2.3 [Minh họa xác định khoảng entropy phù hợp và chọn lọc bản đồ đặc trưng] Xét ảnh đầu vào *2_Basal Cell Carcinoma_Picture1.jpg* và lớp *block1_conv1* của mạng VGG16. Lớp này tạo ra 64 bản đồ đặc trưng, ký hiệu từ FM1 đến FM64. Ví dụ này thực hiện phép tính entropy, xác định khoảng entropy phù hợp và lựa chọn các bản đồ đặc trưng được giữ lại; tập bản đồ sau lựa chọn được trình bày tiếp trong Ví dụ 2.4.

Hình 2.10 minh họa quá trình chọn lọc bản đồ đặc trưng. Entropy của từng bản đồ đặc trưng được tính theo Công thức (2.13). Các giá trị entropy hiển thị trong Hình 2.10 được sử dụng để minh họa quá trình xác định khoảng entropy phù hợp.

Với ngưỡng entropy tối thiểu $\theta^{(block1_conv1)} = 0.10$, các bản đồ đặc trưng có entropy nhỏ hơn hoặc bằng ngưỡng này sẽ bị loại khỏi tập entropy hợp lệ theo Công thức (2.20). Trong ví dụ này, tất cả 64 giá trị entropy đều lớn hơn 0.10, do đó tập entropy hợp lệ $E^{(block1_conv1)}$ gồm toàn bộ 64 giá trị sau: Vì các giá trị được tính theo logarit cơ số 2, giới hạn trên của mỗi entropy là $\log_2(50176) \approx 15.6147$ bit. Giá trị lớn nhất trong danh sách là 13.72 bit, nên toàn bộ 64 giá trị đều thỏa mãn giới hạn toán học của định nghĩa đã sửa.

$$E^{(block1_conv1)} = \{8.75, 5.10, 9.22, 8.76, 10.35, 5.41, 7.05, 11.20, 13.40, 5.24, 8.72, 4.76, 2.65, 5.09, 6.92, 13.72, 11.12, 6.15, 6.37, 12.29, 5.10, 9.64, 5.41, 6.93, 4.88, 4.13, 8.21, 11.68, 8.39, 11.50, 6.46, 10.70, 11.03, 11.72, 8.46, 8.52, 5.41, 9.57, 5.04, 6.12, 3.95, 7.49, 8.27, 7.78, 4.90, 8.50, 11.72, 8.16, 11.18, 4.94, 5.55, 8.36, 11.87, 11.43, 6.70, 4.96, 10.59, 6.71, 8.89, 11.19, 8.24, 6.09, 8.45, 10.07\}.$$

Từ các giá trị entropy trong $E^{(block1_conv1)}$, giá trị trung bình được tính bởi:

$$\mu_e^{(block1_conv1)} = \frac{1}{64} \sum_{m=1}^{64} H\left(F_m^{(block1_conv1)}(x_i)\right) \approx 8.0806.$$

Độ lệch chuẩn của các giá trị entropy được tính bởi:

$$\sigma_e^{(block1_conv1)} = \sqrt{\frac{1}{64} \sum_{m=1}^{64} \left(H\left(F_m^{(block1_conv1)}(x_i)\right) - \mu_e^{(block1_conv1)}\right)^2} \approx 2.6297.$$

Theo Công thức (2.21), cận dưới của khoảng entropy phù hợp được xác định bởi:

$$\begin{aligned} e_{\min}^{(block1_conv1)} &= \max\left(\min(E^{(block1_conv1)}), \mu_e^{(block1_conv1)} - \sigma_e^{(block1_conv1)}\right) \\ &= \max(2.65, 8.0806 - 2.6297) \\ &= \max(2.65, 5.4509) \\ &\approx 5.45. \end{aligned}$$

Tương tự, theo Công thức (2.22), cận trên của khoảng entropy phù hợp được xác định bởi:

$$\begin{aligned} e_{\max}^{(block1_conv1)} &= \min\left(\max(E^{(block1_conv1)}), \mu_e^{(block1_conv1)} + \sigma_e^{(block1_conv1)}\right) \\ &= \min(13.72, 8.0806 + 2.6297) \\ &= \min(13.72, 10.7103) \\ &\approx 10.71. \end{aligned}$$

Như vậy, khoảng entropy phù hợp của lớp $block1_conv1$ là:

$$\left[e_{\min}^{(block1_conv1)}, e_{\max}^{(block1_conv1)}\right] \approx [5.45, 10.71].$$

Một bản đồ đặc trưng $F_m^{(block1_conv1)}(x_i)$ được giữ lại nếu entropy của nó thỏa mãn điều kiện trong Công thức (2.24), tức là:

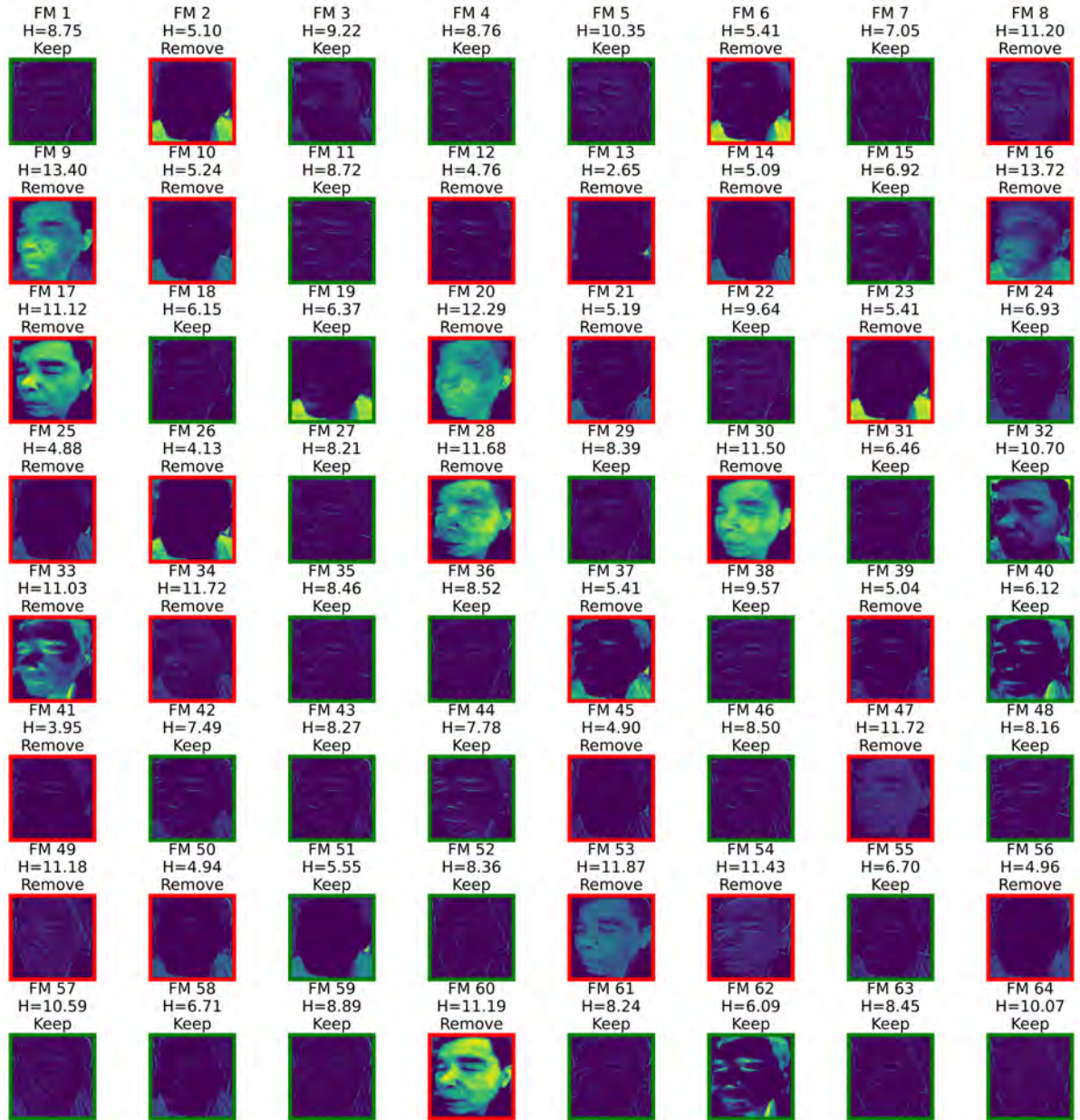
$$5.45 \leq H\left(F_m^{(block1_conv1)}(x_i)\right) \leq 10.71.$$

Chẳng hạn, trong Hình 2.10, FM1 có $H = 8.75$ nên được giữ lại vì $5.45 \leq 8.75 \leq 10.71$; FM2 có $H = 5.10$ nên bị loại vì $5.10 < 5.45$; FM5 có $H = 10.35$ nên được giữ lại vì $5.45 \leq 10.35 \leq 10.71$; FM8 có $H = 11.20$ nên bị loại vì $11.20 > 10.71$; và FM9 có $H = 13.40$ cũng bị loại vì $13.40 > 10.71$.

Với điều kiện lựa chọn trên, trong tổng số 64 bản đồ đặc trưng của lớp $block1_conv1$, có 34 bản đồ đặc trưng được giữ lại và 30 bản đồ đặc trưng bị loại bỏ, đúng như minh họa trong Hình 2.10. Kết quả này minh họa nguyên tắc lựa chọn của BlockNet: giữ lại các bản đồ đặc trưng có entropy nằm trong khoảng phù hợp của từng block CNN, thay vì chọn theo giá trị entropy lớn nhất.

CHÚ Ý 2.2 Hình 2.10 minh họa rằng các bản đồ bị sàng lọc gồm cả trường hợp entropy thấp và entropy cao. Entropy thấp thường đi cùng phân bố kích hoạt tập trung, còn entropy cao phản ánh phân bố rộng hơn và trong một số trường hợp có thể đi kèm phản hồi không ổn định. BlockNet giữ các bản đồ nằm trong khoảng entropy của từng block

Entropy-based feature map selection at block1_conv1
Selected: 34, Removed: 30, theta=0.10, e_min=5.45, e_max=10.71



Hình 2.10: Minh họa chọn lọc bản đồ đặc trưng dựa trên entropy tại lớp block1_conv1. Các bản đồ đặc trưng có viền xanh được giữ lại do nằm trong khoảng entropy phù hợp; các bản đồ đặc trưng có viền đỏ bị loại bỏ.

như một quy tắc sàng lọc, sau đó đánh giá quy tắc này bằng kết quả phân loại.

Sau bước chọn lọc, tập $\mathcal{F}^{(b)}(x_i)$ được sử dụng để xây dựng vector đặc trưng học sâu $z_i^{(d,b)}$ của mẫu x_i tại block b . Số lượng bản đồ đặc trưng được giữ lại không được ấn định trước theo một giá trị cố định như top- k , mà phụ thuộc vào số bản đồ đặc trưng có entropy nằm trong khoảng phù hợp $[e_{\min}^{(b)}, e_{\max}^{(b)}]$. Nói cách khác, BlockNet không sử dụng toàn bộ bản đồ đặc trưng ban đầu, đồng thời cũng không chọn một số lượng cố định bản đồ đặc trưng cho mọi block, mà lựa chọn các bản đồ đặc trưng thỏa mãn tiêu chí entropy phù hợp của từng block CNN.

Trong phạm vi bước chọn lọc entropy, số lượng bản đồ đặc trưng được giữ lại được kiểm soát bởi khoảng entropy phù hợp của từng block. Sau đó, khi các bản đồ đặc trưng hợp lệ được chuyển đổi thành vector đặc trưng học sâu và tổ chức thành tập dữ liệu đặc trưng phục vụ huấn luyện, số lượng mẫu trong từng lớp tiếp tục được điều chỉnh thông qua cơ chế cân bằng dữ liệu trong không gian đặc trưng, được trình bày ở mục tiếp theo.

Vector đặc trưng học sâu $z_i^{(d,b)}$ sau đó được hợp nhất với vector đặc trưng thủ công $z_i^{(h)}$ để tạo thành biểu diễn đặc trưng hợp nhất $z_i^{(b)}$ phục vụ cho quá trình phân loại. Như vậy, bước chọn lọc bản đồ đặc trưng dựa trên entropy đóng vai trò trung gian giữa quá trình trích xuất bản đồ đặc trưng từ CNN và quá trình hợp nhất đặc trưng đa nguồn trong BlockNet.

Cách tổ chức này giảm số lượng bản đồ đặc trưng cần xử lý, hạn chế các phản hồi dư thừa hoặc không ổn định và giữ lại các bản đồ đặc trưng có mức thông tin phù hợp cho từng block CNN.

2.3.5. Hợp nhất đặc trưng đa nguồn

Sau bước chọn lọc bản đồ đặc trưng dựa trên entropy, BlockNet thu được các tập bản đồ đặc trưng hợp lệ để xây dựng vector đặc trưng học sâu. Tuy nhiên, các đặc trưng học sâu được trích xuất từ CNN chủ yếu phản ánh các biểu diễn được học tự động từ dữ liệu ảnh, trong khi một số đặc điểm trực quan như màu sắc, kết cấu hoặc hình dạng tổn thương vẫn có thể được mô tả kết quả bằng các đặc trưng thủ công. Vì vậy, BlockNet thực hiện hợp nhất đặc trưng đa nguồn để khai thác tính bổ trợ giữa đặc trưng học sâu và đặc trưng thủ công.

ĐỊNH NGHĨA 2.5 [Trích rút đặc trưng từ BlockNet [CT3]]

Cho ảnh đầu vào x và block $b \in \mathcal{B}$, BlockNet thực hiện quá trình trích xuất đặc trưng theo bốn bước chính: (i) trích xuất bản đồ đặc trưng từ block tích chập của CNN; (ii) đánh giá entropy và chọn lọc các bản đồ đặc trưng hợp lệ; (iii) chuyển đổi các bản đồ đặc trưng được giữ lại thành vector đặc trưng học sâu; và (iv) hợp nhất vector đặc trưng học sâu với vector đặc trưng thủ công để tạo biểu diễn đặc trưng cuối cùng.

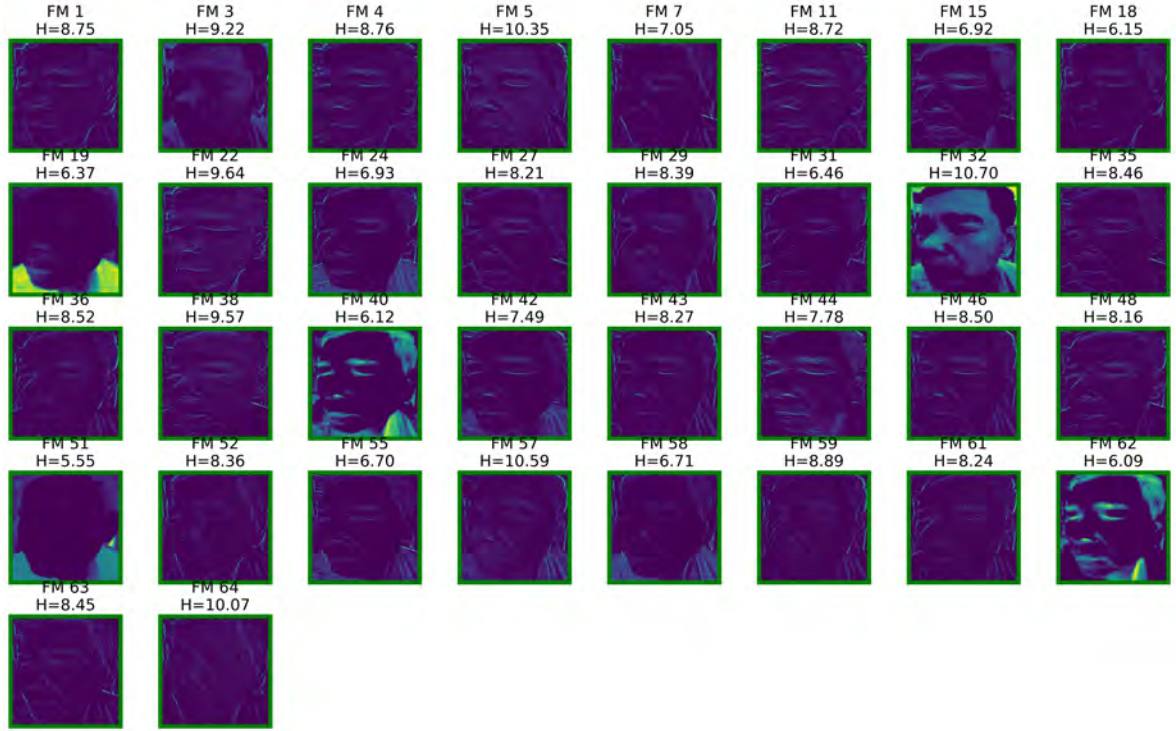
Với mỗi block b , biểu diễn đặc trưng hợp nhất của BlockNet được xác định theo Công thức (2.6).

Định nghĩa 2.5 cho thấy BlockNet không trực tiếp sử dụng toàn bộ bản đồ đặc trưng được tạo ra từ CNN, mà xây dựng biểu diễn đặc trưng theo từng block thông qua cơ chế chọn lọc entropy trước khi thực hiện hợp nhất đa nguồn. Cách tiếp cận này giúp giảm dư thừa trong không gian biểu diễn và giữ lại các thành phần thông tin phù hợp hơn cho phân loại.

VÍ DỤ 2.4 [Minh họa tập bản đồ đặc trưng được giữ lại] Ví dụ này tiếp tục sử dụng ảnh *2_Basal Cell Carcinoma_Picture1.jpg* và lớp *block1_conv1* đã xét trong Ví dụ 2.3. Theo kết quả tính toán từ 64 giá trị entropy trong ví dụ trước, khoảng entropy phù hợp là [5.45, 10.71]. Có 34 bản đồ đặc trưng thỏa mãn điều kiện lựa chọn và 30 bản đồ đặc trưng bị loại. Hình 2.11 trình bày riêng tập bản đồ đặc trưng được giữ lại để xây dựng vector đặc trưng học sâu.

Trong BlockNet, tập bản đồ đặc trưng hợp lệ của mẫu x_i tại block b , ký hiệu là $\mathcal{F}^{*(b)}(x_i)$, đóng vai trò là nguồn biểu diễn học sâu cho bước hợp nhất đa nguồn. Các bản đồ đặc trưng này được chuyển đổi thành vector đặc trưng học sâu thông qua các phép biến đổi như chuẩn hóa, pooling hoặc flatten. Vector đặc trưng học sâu thu được ký hiệu

Kept feature maps at block1_conv1
Kept: 34/64, theta=0.10, e_min=5.45, e_max=10.71



Hình 2.11: Các bản đồ đặc trưng hợp lệ được giữ lại từ lớp block1_conv1.

là $z_i^{(d,b)}$. Do đầu vào của bước này đã được chọn lọc theo entropy, vector $z_i^{(d,b)}$ có xu hướng gọn hơn và ít chứa thành phần dư thừa hơn so với trường hợp sử dụng toàn bộ bản đồ đặc trưng ban đầu.

Song song với đặc trưng học sâu, BlockNet sử dụng ba bộ mô tả thủ công được trích xuất trực tiếp từ ảnh đầu vào, gồm histogram HSV mô tả phân bố màu, GLCM mô tả kết cấu và fractal mô tả độ phức tạp hình dạng. Véc-tơ đặc trưng thủ công được ký hiệu là $z_i^{(h)}$. Sau đó, véc-tơ đặc trưng học sâu $z_i^{(d,b)}$ và véc-tơ đặc trưng thủ công $z_i^{(h)}$ được hợp nhất để tạo biểu diễn đặc trưng cuối cùng của block b , được xác định theo Công thức (2.25).

$$z_i^{(b)} = \psi \left(z_i^{(d,b)}, z_i^{(h)} \right), \quad (2.25)$$

trong đó $\psi(\cdot)$ là phép hợp nhất đặc trưng.

Phép hợp nhất được thực hiện theo cơ chế ghép nối vector đặc trưng. Cách tiếp cận này giữ lại thông tin từ từng nguồn biểu diễn, đồng thời cho phép bộ phân loại khai thác quan hệ hỗ trợ giữa đặc trưng học sâu và đặc trưng thủ công.

Trong BlockNet, quá trình hợp nhất đặc trưng được thực hiện riêng cho từng block CNN. Điều này có nghĩa là mỗi block b sẽ tạo ra một biểu diễn đặc trưng đa nguồn riêng:

$$z_i^{(b)}.$$

Từ đó, mỗi block hình thành một tập dữ liệu đặc trưng tương ứng phục vụ cho quá trình huấn luyện và đánh giá bộ phân loại. Cách tổ chức này cho phép đánh giá khả năng

biểu diễn của từng block CNN thông qua các biểu diễn đa nguồn đã được chọn lọc bằng entropy. Việc hợp nhất đặc trưng đa nguồn có các vai trò sau. Thứ nhất, đặc trưng học sâu có khả năng học các mẫu biểu diễn phức tạp từ dữ liệu ảnh, là các cấu trúc không dễ mô tả bằng công thức thủ công. Thứ hai, đặc trưng thủ công giúp bổ sung các mô tả trực quan ổn định như màu sắc, kết cấu và hình dạng, vốn có ý nghĩa trong phân tích ảnh tổn thương da. Thứ ba, việc kết hợp hai nhóm đặc trưng giúp tăng tính phong phú của biểu diễn đầu vào, qua đó hỗ trợ bộ phân loại trong điều kiện dữ liệu có độ biến thiên cao.

Ngoài ra, việc thực hiện hợp nhất sau bước chọn lọc entropy giảm ảnh hưởng của các bản đồ đặc trưng dư thừa hoặc có phản hồi không ổn định. Điều này góp phần làm cho biểu diễn cuối cùng gọn hơn nhưng vẫn duy trì được các thành phần thông tin đối với phân loại.

Bước hợp nhất đặc trưng đa nguồn đóng vai trò kết nối giữa cơ chế chọn lọc bản đồ đặc trưng dựa trên entropy và giai đoạn phân loại. Với mỗi block CNN, BlockNet xây dựng một biểu diễn đặc trưng đa nguồn riêng, tạo cơ sở cho các bước cân bằng dữ liệu trong không gian đặc trưng và huấn luyện bộ phân loại ở các mục tiếp theo.

2.3.6. Cân bằng dữ liệu trong không gian đặc trưng

Trong các bộ dữ liệu ảnh y tế, số lượng mẫu giữa các lớp thường không đồng đều. Một số lớp có nhiều mẫu hơn so với các lớp còn lại, trong khi các lớp thiếu số lại có ít dữ liệu để mô hình học được đặc trưng phân biệt. Hiện tượng này có thể làm cho bộ phân loại thiên lệch về các lớp có số lượng mẫu lớn và làm giảm khả năng nhận dạng đối với các lớp ít mẫu.

Trong BlockNet, cân bằng dữ liệu được thực hiện trong không gian đặc trưng để phù hợp với quy trình trích xuất và chọn lọc đặc trưng đã trình bày ở các mục trước. Sau khi các bản đồ đặc trưng hợp lệ của từng block CNN được lựa chọn theo Định nghĩa 2.4, các biểu diễn đặc trưng đa nguồn thu được sẽ được sử dụng làm cơ sở cho bước tăng cường dữ liệu.

Như vậy, quá trình cân bằng không được thực hiện trực tiếp trên ảnh đầu vào, mà được thực hiện trên các vector đặc trưng đã được xây dựng từ các bản đồ đặc trưng hợp lệ và đặc trưng thủ công. Cách tiếp cận này giúp duy trì tính nhất quán giữa bước chọn lọc entropy và bước tăng cường dữ liệu trong BlockNet.

Với mỗi block $b \in \mathcal{B}$, tập biểu diễn đặc trưng đa nguồn được xác định:

$$\mathcal{Z}^{(b)} = \left\{ z_i^{(b)} \right\}_{i=1}^N,$$

trong đó $z_i^{(b)}$ là vector đặc trưng hợp nhất của mẫu thứ i đối với block b .

Giả sử $n_k^{(b)}$ là số lượng vector đặc trưng thuộc lớp k trong tập biểu diễn của block b , và:

$$n_{\max}^{(b)} = \max_k \left(n_k^{(b)} \right)$$

là số lượng mẫu lớn nhất giữa các lớp của block đó.

Để giảm chênh lệch số lượng mẫu giữa các lớp, BlockNet sử dụng hệ số tăng cường đã được xác định theo Công thức (2.26):

$$m_k^{(b)} = \left\lceil \frac{n_{\max}^{(b)}}{n_k^{(b)}} \right\rceil. \quad (2.26)$$

Khi đó, số lượng biểu diễn đặc trưng sau tăng cường của lớp k trong block b được xác định:

$$n_k'^{(b)} = n_k^{(b)} \cdot m_k^{(b)}. \quad (2.27)$$

Trong cách biểu diễn này, $m_k^{(b)}$ cho biết số lần các vector đặc trưng hợp lệ của lớp k trong block b được sử dụng lại trong quá trình tăng cường. Do $m_k^{(b)}$ được xác định bằng phép làm tròn lên, số lượng biểu diễn sau tăng cường $n_k'^{(b)}$ có thể bằng hoặc lớn hơn $n_{\max}^{(b)}$. Điều này cho thấy cơ chế này hướng tới việc làm giảm mất cân bằng lớp trong không gian đặc trưng, thay vì bắt buộc số lượng mẫu sau tăng cường của mọi lớp phải bằng nhau tuyệt đối.

Để trình bày rõ hơn quá trình thực hiện trong thuật toán, số lượng vector đặc trưng cần bổ sung cho lớp k của block b được ký hiệu là:

$$\Delta_k^{(b)} = n_k'^{(b)} - n_k^{(b)}. \quad (2.28)$$

Như vậy, $m_k^{(b)}$, $n_k'^{(b)}$ và $\Delta_k^{(b)}$ không biểu diễn ba cơ chế cân bằng khác nhau, mà là ba đại lượng mô tả cùng một quá trình tăng cường trong không gian đặc trưng. Cụ thể, $m_k^{(b)}$ là hệ số tăng cường, $n_k'^{(b)}$ là số lượng biểu diễn đặc trưng sau tăng cường, còn $\Delta_k^{(b)}$ là số lượng vector đặc trưng cần bổ sung trong quá trình thực hiện thuật toán.

Nếu $\Delta_k^{(b)} > 0$, các vector đặc trưng hợp lệ của lớp k trong block b được chọn và sử dụng lại để bổ sung vào tập dữ liệu huấn luyện. Nếu $\Delta_k^{(b)} = 0$, lớp đó không cần bổ sung thêm biểu diễn đặc trưng.

Sau bước tăng cường, mỗi block b hình thành tập dữ liệu đặc trưng cân bằng $D_{\text{balanced}}^{(b)}$, trong đó các biểu diễn đặc trưng đã được chọn lọc theo entropy, hợp nhất đa nguồn và tăng cường trong không gian đặc trưng.

Trong BlockNet, quá trình cân bằng dữ liệu được thực hiện riêng cho từng block CNN. Điều này có nghĩa là mỗi block sau bước chọn lọc entropy và hợp nhất đặc trưng sẽ tạo thành một tập biểu diễn đặc trưng riêng phục vụ cho quá trình huấn luyện và đánh giá bộ phân loại.

Cách tổ chức này cho phép đánh giá khả năng biểu diễn của từng block thông qua các tập dữ liệu đặc trưng tương ứng. Thay vì sử dụng một bản đồ đặc trưng đơn lẻ của mỗi block, nghiên cứu khai thác tập các bản đồ đặc trưng hợp lệ sau bước chọn lọc entropy để giữ lại nhiều thành phần thông tin bổ sung trong cùng một block. Do đó, mỗi block được xem như một nguồn biểu diễn đặc trưng riêng biệt trong BlockNet.

Việc xây dựng các tập dữ liệu đặc trưng riêng cho từng block cũng giúp đánh giá ảnh hưởng của mức độ trừu tượng của đặc trưng CNN đối với bài toán phân loại. Các block đầu thường phản ánh các đặc trưng mức thấp như biên hoặc kết cấu cục bộ, trong khi các block sâu hơn có xu hướng biểu diễn các cấu trúc ngữ nghĩa phức tạp hơn. Thông qua các tập dữ liệu đặc trưng $D_{\text{balanced}}^{(b)}$, BlockNet có thể đánh giá kết quả biểu diễn của từng block trước khi xác định cấu hình phù hợp cho quá trình phân loại.

CHÚ Ý 2.3 cơ chế cân bằng dữ liệu được thực hiện trong không gian đặc trưng để duy trì tính nhất quán với thiết kế của BlockNet. Trọng tâm của chương không phải là so sánh toàn diện các chiến lược tăng cường dữ liệu khác nhau, mà là trình bày một quy trình thống nhất gồm trích xuất bản đồ đặc trưng, chọn lọc bằng entropy, hợp nhất đặc trưng

đa nguồn và cân bằng trên biểu diễn đặc trưng đã chọn.

Về bản chất, cơ chế trên là lấy mẫu tăng cường ngẫu nhiên có hoàn lại trong không gian đặc trưng (random oversampling in feature space). Phương pháp sử dụng lại các vector đặc trưng hợp nhất hiện có của lớp thiểu số; không sinh thêm ảnh và không nội suy để tạo vector tổng hợp mới. Toàn bộ thao tác chỉ được thực hiện trên phần huấn luyện của từng fold sau khi dữ liệu đã được chia; tập xác thực và tập kiểm tra được giữ nguyên để tránh rò rỉ thông tin.

Cơ chế này can thiệp ở mức phân bố mẫu nên khác với hàm mất mát có trọng số theo lớp (class-weighted loss), trong đó dữ liệu được giữ nguyên nhưng sai số của lớp thiểu số được gán trọng số lớn hơn; và khác với focal loss, trong đó đóng góp của các mẫu dễ được giảm để mô hình tập trung hơn vào các mẫu khó [114, 115]. So với SMOTE, phương pháp của BlockNet không tạo vector mới bằng nội suy giữa các mẫu thiểu số, do đó tránh đưa thêm biểu diễn tổng hợp vào không gian đặc trưng [116]. Mỗi hướng xử lý có ưu điểm và hạn chế riêng; phân đối chiếu này nhằm làm rõ vị trí can thiệp của cơ chế được sử dụng, không được xem là một so sánh hiệu năng thực nghiệm giữa các phương pháp xử lý mất cân bằng [117, 118].

VÍ DỤ 2.5 [Minh họa cân bằng dữ liệu trong không gian đặc trưng] Giả sử đối với một block CNN, lớp có số lượng mẫu lớn nhất có:

$$n_{\max}^{(b)} = 100.$$

Xét một lớp thiểu số có:

$$n_k^{(b)} = 60.$$

Theo Công thức (2.26), hệ số tăng cường của lớp này được xác định:

$$m_k^{(b)} = \left\lceil \frac{100}{60} \right\rceil = 2.$$

Khi đó, số lượng biểu diễn đặc trưng sau tăng cường được xác định theo Công thức (2.27):

$$n_k'^{(b)} = 60 \times 2 = 120.$$

Số lượng vector đặc trưng cần bổ sung được xác định theo Công thức (2.28):

$$\Delta_k^{(b)} = 120 - 60 = 60.$$

Ví dụ này cho thấy số lượng mẫu sau tăng cường có thể lớn hơn $n_{\max}^{(b)}$ do sử dụng phép làm tròn lên trong hệ số tăng cường. Vì vậy, kết quả sau tăng cường không nhất thiết tạo ra số lượng mẫu tuyệt đối bằng nhau giữa các lớp, mà nhằm làm giảm mức độ mất cân bằng trong không gian đặc trưng.

Sau khi hoàn tất bước cân bằng, tập dữ liệu đặc trưng $D_{\text{balanced}}^{(b)}$ của từng block CNN sẽ được sử dụng làm đầu vào cho bộ phân loại. Nội dung về bộ phân loại và hàm mục tiêu được trình bày trong mục tiếp theo.

2.3.7. Bộ phân loại và hàm mục tiêu

Sau khi hoàn tất các bước chọn lọc bản đồ đặc trưng, hợp nhất đặc trưng đa nguồn và cân bằng dữ liệu trong không gian đặc trưng, mỗi block CNN tạo ra tập dữ liệu đặc trưng cân bằng $D_{\text{balanced}}^{(b)}$. Mỗi mẫu đầu vào trong tập dữ liệu này được biểu diễn bởi vector đặc trưng hợp nhất $z_i^{(b)}$, và vector này được sử dụng làm đầu vào cho bộ phân loại nhằm dự đoán nhãn lớp tương ứng.

ĐỊNH NGHĨA 2.6 [Hàm phân loại] Hàm phân loại tổng quát đã được xác định trong Công thức (2.2). Trong giai đoạn huấn luyện của BlockNet, đầu vào cụ thể của hàm phân loại là vector đặc trưng hợp nhất $z_i^{(b)}$ của mẫu x_i đối với block b .

Khi đó, bộ phân loại học ánh xạ từ biểu diễn đặc trưng $z_i^{(b)}$ sang nhãn lớp $y_i \in \mathcal{Y}$.

Trong thực tế, hàm phân loại trả về vector xác suất dự đoán:

$$\hat{\mathbf{p}}_i^{(b)} = g_{\omega}^{(b)} \left(z_i^{(b)} \right), \quad (2.29)$$

trong đó $\hat{\mathbf{p}}_i^{(b)}$ là vector xác suất dự đoán của block b , còn $\omega^{(b)}$ là tập tham số của bộ phân loại tương ứng với block đó.

bộ phân loại Multi-Layer Perceptron (MLP) được sử dụng để đánh giá kết quả của các biểu diễn đặc trưng do BlockNet tạo ra. Việc sử dụng MLP với kiến trúc tương đối đơn giản giúp tập trung đánh giá chất lượng của các biểu diễn đặc trưng sau các bước chọn lọc entropy, hợp nhất đa nguồn và cân bằng trong không gian đặc trưng, thay vì để kết quả bị chi phối bởi một kiến trúc phân loại quá phức tạp.

Điều này cho thấy trọng tâm của BlockNet không nằm ở việc đề xuất một kiến trúc học sâu mới cho bộ phân loại, mà nằm ở cơ chế xây dựng biểu diễn đặc trưng kết quả từ bản đồ đặc trưng của CNN và các đặc trưng thủ công. Do đó, MLP được lựa chọn nhằm phản ánh rõ hơn vai trò của các bước xử lý đặc trưng trong quy trình đề xuất.

Trong BlockNet, quá trình huấn luyện và đánh giá bộ phân loại được thực hiện riêng cho từng block CNN. Cụ thể, mỗi block $b \in \mathcal{B}$ tạo ra một tập dữ liệu đặc trưng cân bằng $D_{\text{balanced}}^{(b)}$ và một bộ phân loại tương ứng $g_{\omega}^{(b)}$. Cách tổ chức này cho phép đánh giá khả năng biểu diễn của từng block CNN thông qua kết quả phân loại trên các tập đặc trưng tương ứng.

Với vector đặc trưng hợp nhất $z_i^{(b)}$, bộ phân loại MLP gồm các lớp fully-connected liên tiếp. Mỗi lớp ẩn thực hiện phép biến đổi tuyến tính kết hợp với hàm kích hoạt phi tuyến nhằm học các quan hệ giữa các thành phần đặc trưng.

Ở lớp đầu ra, hàm softmax được sử dụng để chuyển các giá trị đầu ra thành phân bố xác suất trên C lớp:

$$\hat{p}_{ic}^{(b)} = \frac{\exp(o_{ic}^{(b)})}{\sum_{r=1}^C \exp(o_{ir}^{(b)})}, \quad (2.30)$$

trong đó $o_{ic}^{(b)}$ là giá trị đầu ra trước softmax tương ứng với lớp c của mẫu i đối với block b , còn $\hat{p}_{ic}^{(b)}$ là xác suất dự đoán mẫu i thuộc lớp c .

Nhãn dự đoán cuối cùng của mẫu x_i đối với block b được xác định:

$$\hat{y}_i^{(b)} = \arg \max_{c \in \{1, \dots, C\}} \hat{p}_{ic}^{(b)}. \quad (2.31)$$

Để huấn luyện bộ phân loại, hàm mất mát cross-entropy được sử dụng. Với nhãn thật ở dạng one-hot:

$$y_{ic},$$

hàm mất mát trên tập huấn luyện của block b được xác định:

$$\mathcal{L}^{(b)} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log \left(\hat{p}_{ic}^{(b)} \right). \quad (2.32)$$

Trong Công thức (2.32), nếu mẫu i thuộc lớp c thì:

$$y_{ic} = 1,$$

ngược lại:

$$y_{ic} = 0.$$

Hàm mất mát này khuyến khích mô hình gán xác suất cao cho lớp đúng và xác suất thấp cho các lớp còn lại.

Quá trình tối ưu hóa bộ phân loại được thực hiện bằng lan truyền ngược và thuật toán tối ưu Adam. Các tham số $\omega^{(b)}$ của bộ phân loại được cập nhật nhằm cực tiểu hóa hàm mất mát trong Công thức (2.32).

Sau khi huấn luyện, bộ phân loại của từng block CNN nhận vector đặc trưng hợp nhất $z_i^{(b)}$ của ảnh đầu vào và trả về nhãn dự đoán tương ứng.

Bộ phân loại trong BlockNet đóng vai trò là thành phần ra quyết định cuối cùng trên các biểu diễn đặc trưng đã được xây dựng qua các bước trích xuất bản đồ đặc trưng, chọn lọc entropy, hợp nhất đa nguồn và cân bằng dữ liệu trong không gian đặc trưng.

Cách thiết kế này giúp nhấn rằng kết quả của mô hình đề xuất chủ yếu đến từ chất lượng biểu diễn đặc trưng được xây dựng theo từng block CNN, thay vì từ độ phức tạp của bộ phân loại.

2.4. Thuật toán đề xuất và quy trình thực hiện

Từ các thành phần đã trình bày ở các mục trước, quy trình BlockNet được tổng hợp thành ba thuật toán chính. Thuật toán 2.1 thực hiện chọn lọc bản đồ đặc trưng dựa trên entropy. Thuật toán 2.2 thực hiện cân bằng dữ liệu trong không gian đặc trưng. Thuật toán 2.3 mô tả quy trình tổng thể của BlockNet theo từng block CNN, từ trích xuất bản đồ đặc trưng, chọn lọc entropy, hợp nhất đặc trưng, cân bằng dữ liệu trong không gian đặc trưng đến huấn luyện và đánh giá bộ phân loại.

Phân tích thuật toán. Thuật toán 2.1 thực hiện chọn lọc bản đồ đặc trưng ở mức từng block CNN. Đầu vào của thuật toán là tập bản đồ đặc trưng $\mathcal{F}^{(b)}(x_i)$ đã được trích xuất từ mẫu x_i tại block b . Trước hết, entropy của từng bản đồ đặc trưng được tính nhằm đánh giá mức độ phân bố thông tin của bản đồ đặc trưng đó. Các bản đồ đặc trưng có entropy nhỏ hơn hoặc bằng ngưỡng $\theta^{(b)}$ bị loại khỏi tập entropy hợp lệ. Khi đó, $E^{(b)}$ là tập các giá trị entropy hợp lệ tại block b , được sử dụng để xác định khoảng entropy phù hợp $[e_{\min}^{(b)}, e_{\max}^{(b)}]$.

Sau khi xác định khoảng entropy phù hợp, thuật toán giữ lại các bản đồ có entropy nằm trong khoảng $[e_{\min}^{(b)}, e_{\max}^{(b)}]$ để tạo thành tập $\mathcal{F}^{*(b)}(x_i)$. Số lượng bản đồ được giữ lại không ấn định trước theo giá trị top- k , mà phụ thuộc vào phân bố entropy tại từng block.

Thuật toán 2.1 Chọn lọc bản đồ đặc trưng dựa trên entropy

Đầu vào: Tập bản đồ đặc trưng $\mathcal{F}^{(b)}(x_i)$ của mẫu x_i tại block b ; ngưỡng entropy tối thiểu $\theta^{(b)}$.

Đầu ra: Khoảng entropy phù hợp $[e_{\min}^{(b)}, e_{\max}^{(b)}]$; Tập bản đồ đặc trưng hợp lệ $\mathcal{F}^{*(b)}(x_i)$.

- 1: Khởi tạo tập entropy hợp lệ $E^{(b)} \leftarrow \emptyset$.
 - 2: **for** mỗi bản đồ đặc trưng $F_m^{(b)}(x_i) \in \mathcal{F}^{(b)}(x_i)$ **do**
 - 3: Tính entropy $H(F_m^{(b)}(x_i))$ theo Công thức (2.13).
 - 4: **if** $H(F_m^{(b)}(x_i)) > \theta^{(b)}$ **then**
 - 5: Thêm $H(F_m^{(b)}(x_i))$ vào $E^{(b)}$.
 - 6: **end if**
 - 7: **end for**
 - 8: Xác định $e_{\min}^{(b)}$ theo Công thức (2.21).
 - 9: Xác định $e_{\max}^{(b)}$ theo Công thức (2.22).
 - 10: Xác định tập bản đồ đặc trưng hợp lệ $\mathcal{F}^{*(b)}(x_i)$ theo Công thức (2.24).
 - 11: **return** $\mathcal{F}^{*(b)}(x_i), [e_{\min}^{(b)}, e_{\max}^{(b)}]$.
-

Cách làm này sàng lọc hai phía của phân bố trước khi xây dựng véc-tơ đặc trưng học sâu.

Độ phức tạp. Giả sử block b có M_b bản đồ đặc trưng, mỗi bản đồ đặc trưng có kích thước $H_b \times W_b$. Việc tính entropy cho toàn bộ M_b bản đồ đặc trưng có độ phức tạp $O(M_b H_b W_b)$. Sau khi các giá trị entropy đã được tính, việc xác định khoảng entropy phù hợp và chọn lọc bản đồ đặc trưng có độ phức tạp $O(M_b)$. Do đó, độ phức tạp tổng quát của Thuật toán 2.1 là:

$$O(M_b H_b W_b).$$

Sau Thuật toán 2.1, mỗi mẫu x_i tại block b thu được tập bản đồ đặc trưng hợp lệ $\mathcal{F}^{*(b)}(x_i)$. Tập này được chuyển đổi thành vector đặc trưng học sâu $z_i^{(d,b)}$ và hợp nhất với vector đặc trưng thủ công $z_i^{(h)}$ để tạo thành vector đặc trưng hợp nhất $z_i^{(b)}$. Khi toàn bộ các mẫu huấn luyện đã được biểu diễn dưới dạng các cặp $(z_i^{(b)}, y_i)$, bước tiếp theo là cân bằng tập biểu diễn này trong không gian đặc trưng để giảm ảnh hưởng của mất cân bằng lớp trước khi huấn luyện bộ phân loại. Nội dung này được trình bày trong Thuật toán 2.2.

Phân tích thuật toán. Thuật toán 2.2 thực hiện cân bằng dữ liệu trong không gian đặc trưng, tức là thao tác trên các vector đặc trưng hợp nhất $z_i^{(b)}$ đã được xây dựng tại từng block CNN, thay vì tạo thêm ảnh mới ở mức dữ liệu đầu vào. Với mỗi block b , thuật toán xác định số lượng mẫu của từng lớp trong tập $\mathcal{Z}^{(b)}$, sau đó tính số lượng vector đặc trưng cần bổ sung cho các lớp thiếu số.

Việc bổ sung dữ liệu được thực hiện bằng cách chọn ngẫu nhiên có hoàn lại các vector đặc trưng hiện có của lớp cần cân bằng. Cơ chế này là một dạng lấy mẫu lặp trong không gian đặc trưng; thuật toán không sinh ảnh mới và cũng không tạo vector tổng hợp mới. Do đó, $D_{\text{balanced}}^{(b)}$ là tập dữ liệu đặc trưng sau cân bằng tại block b , gồm các cặp $(z_i^{(b)}, y_i)$ và được sử dụng để huấn luyện bộ phân loại tương ứng với block đó. Cách tổ chức này giúp giảm ảnh hưởng của mất cân bằng lớp trong quá trình huấn luyện, đồng thời tránh làm thay đổi dữ liệu ảnh gốc.

Thuật toán 2.2 Cân bằng dữ liệu trong không gian đặc trưng

Đầu vào: Tập biểu diễn đặc trưng hợp nhất tại block b , $\mathcal{Z}^{(b)} = \{(z_i^{(b)}, y_i)\}_{i=1}^N$; tập nhãn $\mathcal{Y}$.

Đầu ra: Tập dữ liệu đặc trưng sau cân bằng $D_{\text{balanced}}^{(b)}$.

- 1: Khởi tạo $D_{\text{balanced}}^{(b)} \leftarrow \mathcal{Z}^{(b)}$.
 - 2: Xác định số lượng mẫu $n_k^{(b)}$ của từng lớp $k \in \mathcal{Y}$.
 - 3: Xác định $n_{\max}^{(b)}$ là số lượng mẫu lớn nhất giữa các lớp.
 - 4: **for** mỗi lớp $k \in \mathcal{Y}$ **do**
 - 5: Tính hệ số cân bằng $m_k^{(b)}$ theo Công thức (2.26).
 - 6: Tính số lượng biểu diễn sau cân bằng $n_k'^{(b)}$ theo Công thức (2.27).
 - 7: Tính số lượng vector cần bổ sung $\Delta_k^{(b)}$ theo Công thức (2.28).
 - 8: **if** $\Delta_k^{(b)} > 0$ **then**
 - 9: Chọn ngẫu nhiên có hoàn lại $\Delta_k^{(b)}$ vector đặc trưng từ tập $\{z_i^{(b)} \mid y_i = k\}$.
 - 10: Bổ sung các vector được chọn cùng nhãn lớp k vào $D_{\text{balanced}}^{(b)}$.
 - 11: **end if**
 - 12: **end for**
 - 13: **return** $D_{\text{balanced}}^{(b)}$.
-

Độ phức tạp. Giả sử $\mathcal{Z}^{(b)}$ có N mẫu đặc trưng và tập nhãn có C lớp. Việc đếm số lượng mẫu của từng lớp cần duyệt qua N phần tử nên có độ phức tạp $O(N)$. Việc tính hệ số cân bằng, số lượng sau cân bằng và số lượng vector cần bổ sung cho từng lớp có độ phức tạp $O(C)$. Gọi $R_b = \sum_{k=1}^C \Delta_k^{(b)}$ là tổng số vector đặc trưng được bổ sung tại block b . Khi đó, chi phí bổ sung vector là $O(R_b)$, và độ phức tạp tổng quát của Thuật toán 2.2 là:

$$O(N + C + R_b).$$

Thuật toán 2.2 tạo ra tập dữ liệu đặc trưng sau cân bằng $D_{\text{balanced}}^{(b)}$ cho một block b cụ thể. Tuy nhiên, BlockNet không chỉ xử lý một block đơn lẻ mà lặp lại toàn bộ chuỗi thao tác cho từng block CNN được lựa chọn. Vì vậy, cần mô tả một quy trình tổng thể để liên kết các bước trích xuất bản đồ đặc trưng, chọn lọc entropy, hợp nhất đặc trưng, cân bằng dữ liệu trong không gian đặc trưng và huấn luyện bộ phân loại theo từng block. Quy trình đầy đủ này được trình bày trong Thuật toán 2.3.

Phân tích thuật toán. Thuật toán 2.3 mô tả quy trình tổng thể của BlockNet theo từng block CNN. Với mỗi block b , thuật toán trích xuất tập bản đồ đặc trưng từ ảnh đầu vào, áp dụng Thuật toán 2.1 để giữ lại các bản đồ đặc trưng hợp lệ, sau đó chuyển đổi chúng thành vector đặc trưng học sâu. Vector này được hợp nhất với đặc trưng thủ công để tạo biểu diễn đặc trưng hợp nhất $z_i^{(b)}$.

Tập biểu diễn đặc trưng hợp nhất của block b được cân bằng bằng Thuật toán 2.2 trước khi huấn luyện bộ phân loại. Nhờ thực hiện quy trình riêng cho từng block CNN, BlockNet tạo ra các tập biểu diễn đặc trưng tương ứng với từng mức biểu diễn của mạng CNN. Cách tổ chức này cho phép đánh giá vai trò của từng block trong bài toán phân loại, đồng thời liên kết thống nhất các bước chọn lọc entropy, hợp nhất đặc trưng, cân bằng dữ liệu trong không gian đặc trưng và huấn luyện bộ phân loại.

Thuật toán 2.3 Quy trình tổng thể của BlockNet

Đầu vào: Tập dữ liệu huấn luyện $D = \{(x_i, y_i)\}_{i=1}^N$; tập block CNN $\mathcal{B}$; đặc trưng thủ công $z_i^{(h)}$; ngưỡng entropy $\theta^{(b)}$ với $b \in \mathcal{B}$.

Đầu ra: Các tập dữ liệu đặc trưng sau cân bằng $D_{\text{balanced}}^{(b)}$; các bộ phân loại $g_{\omega}^{(b)}$; nhãn dự đoán $\hat{y}_i^{(b)}$.

- 1: **for** mỗi block $b \in \mathcal{B}$ **do**
 - 2: Khởi tạo tập biểu diễn đặc trưng hợp nhất $\mathcal{Z}^{(b)} \leftarrow \emptyset$.
 - 3: **for** mỗi mẫu $(x_i, y_i) \in D$ **do**
 - 4: Trích xuất tập bản đồ đặc trưng $\mathcal{F}^{(b)}(x_i)$ theo Công thức (2.5).
 - 5: Chọn tập bản đồ đặc trưng hợp lệ $\mathcal{F}^{*(b)}(x_i)$ bằng Thuật toán 2.1.
 - 6: Chuyển đổi $\mathcal{F}^{*(b)}(x_i)$ thành vector đặc trưng học sâu $z_i^{(d,b)}$.
 - 7: Trích xuất hoặc nạp vector đặc trưng thủ công $z_i^{(h)}$.
 - 8: Hợp nhất $z_i^{(d,b)}$ và $z_i^{(h)}$ để tạo $z_i^{(b)}$ theo Công thức (2.25).
 - 9: Thêm $(z_i^{(b)}, y_i)$ vào $\mathcal{Z}^{(b)}$.
 - 10: **end for**
 - 11: Tạo tập dữ liệu đặc trưng sau cân bằng $D_{\text{balanced}}^{(b)}$ bằng Thuật toán 2.2.
 - 12: Huấn luyện bộ phân loại $g_{\omega}^{(b)}$ trên $D_{\text{balanced}}^{(b)}$.
 - 13: Tính xác suất dự đoán theo Công thức (2.30).
 - 14: Xác định nhãn dự đoán theo Công thức (2.31).
 - 15: Tối ưu bộ phân loại theo hàm mất mát trong Công thức (2.32).
 - 16: **end for**
 - 17: **return** $D_{\text{balanced}}^{(b)}, g_{\omega}^{(b)}, \hat{y}_i^{(b)}$ với mọi $b \in \mathcal{B}$.
-

Độ phức tạp. Giả sử số block CNN được sử dụng là $T = |\mathcal{B}|$. Với mỗi block b , chi phí chọn lọc entropy trên một mẫu là $O(M_b H_b W_b)$ như đã phân tích trong Thuật toán 2.1. Với N mẫu, chi phí trích xuất và chọn lọc bản đồ đặc trưng cho block b là $O(NM_b H_b W_b)$. Gọi d_b là chiều của vector đặc trưng hợp nhất $z_i^{(b)}$, chi phí xây dựng và tổ chức tập đặc trưng tại block b là $O(Nd_b)$. Gọi $T_{\text{balance}}^{(b)} = O(N + C + R_b)$ là chi phí cân bằng dữ liệu trong không gian đặc trưng và $T_{\text{clf}}^{(b)}$ là chi phí huấn luyện bộ phân loại tại block b . Khi đó, độ phức tạp tổng quát của Thuật toán 2.3 là:

$$O\left(\sum_{b \in \mathcal{B}} \left(NM_b H_b W_b + Nd_b + T_{\text{balance}}^{(b)} + T_{\text{clf}}^{(b)}\right)\right).$$

Trong đó, R_b là tổng số vector đặc trưng được bổ sung trong bước cân bằng dữ liệu của block b . Biểu thức trên cho thấy chi phí chính của BlockNet đến từ quá trình trích xuất, tính entropy và chọn lọc bản đồ đặc trưng theo từng block, cùng với chi phí cân bằng dữ liệu trong không gian đặc trưng và huấn luyện bộ phân loại trên các tập đặc trưng tương ứng.

Sau khi đã mô tả đầy đủ ba thuật toán của BlockNet, phần tiếp theo trình bày cách thiết lập thực nghiệm để kiểm chứng kết quả của quy trình đề xuất. Cụ thể, các nội dung tiếp theo bao gồm bộ dữ liệu sử dụng, các mô hình so sánh, cấu hình thực nghiệm và các độ đo đánh giá. Các thiết lập này là cơ sở để phân tích ảnh hưởng của chọn lọc bản

đồ đặc trưng dựa trên entropy, hợp nhất đặc trưng đa nguồn và cân bằng dữ liệu trong không gian đặc trưng ở các mục kết quả thực nghiệm.

2.5. Thiết lập thực nghiệm

2.5.1. Bộ dữ liệu

Các thực nghiệm trong chương này được thực hiện trên hai bộ dữ liệu ảnh tổn thương da gồm bộ dữ liệu VNCancer [119] và bộ dữ liệu HAM10000 [3]. Hai bộ dữ liệu này được lựa chọn nhằm đánh giá kết quả của phương pháp BlockNet trong các điều kiện dữ liệu khác nhau về chất lượng ảnh, thiết bị thu nhận ảnh, mức độ đa dạng tổn thương và tính mất cân bằng lớp.

Bộ dữ liệu VNCancer được xây dựng bởi các chuyên gia da liễu tại Bệnh viện Da liễu Trung ương, Việt Nam, từ dữ liệu ảnh thu nhận trong quá trình thăm khám lâm sàng [119]. Đây là ảnh lâm sàng chụp bằng điện thoại thông thường, không phải ảnh soi da (dermoscopy) thu nhận bằng thiết bị chuyên dụng. Ảnh được ghi nhận trong các điều kiện ánh sáng, góc chụp, khoảng cách và nền khác nhau; vì vậy dữ liệu có độ sáng, độ tương phản, tỷ lệ vùng tổn thương và chất lượng ảnh không đồng nhất. Những đặc điểm này phản ánh bối cảnh thu nhận ảnh thực tế tại Việt Nam và tạo ra thách thức khác với các bộ dữ liệu dermoscopy đã được chuẩn hóa.

Ngược lại, HAM10000 là bộ dữ liệu dermoscopy chuẩn quốc tế, được thu nhận bằng các thiết bị chuyên dụng dùng trong da liễu. Các ảnh dermoscopy màu trong HAM10000 có chất lượng cao hơn, vùng tổn thương rõ ràng hơn và ít chịu ảnh hưởng của nhiễu môi trường. Vì vậy, HAM10000 thường được sử dụng như một bộ dữ liệu chuẩn để đánh giá các phương pháp phân loại tổn thương da dựa trên học sâu.

Việc sử dụng đồng thời VNCancer và HAM10000 cho phép đánh giá phương pháp đề xuất trên cả dữ liệu lâm sàng thực tế và dữ liệu dermoscopy chuẩn hóa. Điều này giúp kiểm tra khả năng thích nghi và tính ổn định của BlockNet trong các điều kiện dữ liệu khác nhau.

Bộ dữ liệu VNCancer

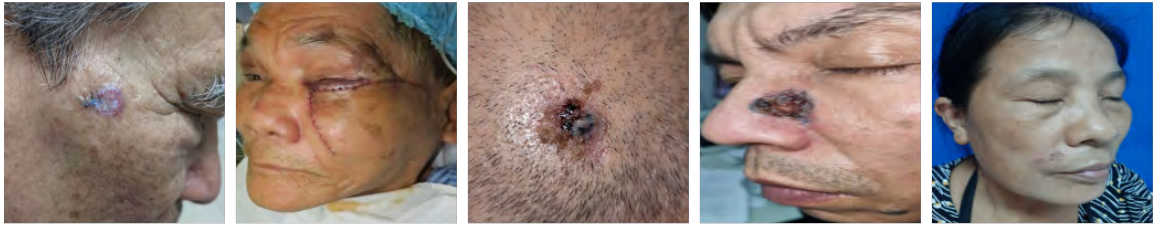
Bảng 2.2 trình bày chi tiết các lớp tổn thương da cùng với số lượng ảnh tương ứng trong bộ dữ liệu VNCancer [119].

Bảng 2.2: Các lớp tổn thương da và phân bố số lượng trong bộ dữ liệu VNCancer

Ký hiệu	Tên tiếng Anh	Tên tiếng Việt	Số lượng
BCC	Basal Cell Carcinoma	Ung thư biểu mô tế bào đáy	2719
MM	Malignant Melanoma	U hắc tố ác tính	314
SCC	Squamous Cell Carcinoma	Ung thư biểu mô tế bào vảy	833
Tổng số ảnh			3866

Các ảnh trong bộ dữ liệu phản ánh nhiều điều kiện chụp khác nhau, sự đa dạng về vị trí tổn thương, kích thước, màu sắc và hình thái bề mặt. Một số ví dụ ảnh trong bộ dữ liệu VNCancer được minh họa trong Hình 2.12.

Tổng cộng bộ dữ liệu VNCancer gồm 3866 ảnh thuộc ba lớp ung thư da, được ký hiệu lần lượt là BCC, MM và SCC. Lớp BCC có 2719 ảnh, chiếm khoảng 70,33%; lớp MM có 314 ảnh, chiếm khoảng 8,12%; và lớp SCC có 833 ảnh, chiếm khoảng 21,55%.

BCC**MM****SCC**

Hình 2.12: Một số ảnh mẫu trong bộ dữ liệu VNCancer với các lớp tổn thương da khác nhau.

Tỷ lệ giữa lớp lớn nhất và lớp nhỏ nhất xấp xỉ 8,66 : 1, cho thấy mức mất cân bằng đáng kể. Hệ ký hiệu BCC, MM và SCC được sử dụng thống nhất trong mô tả dữ liệu, thiết lập thực nghiệm và phân tích kết quả trên VNCancer; ký hiệu MEL chỉ được sử dụng cho lớp melanoma của bộ dữ liệu HAM10000.

VNCancer được sử dụng để đánh giá BlockNet trong điều kiện ảnh lâm sàng thực tế, có chất lượng và bối cảnh thu nhận không đồng nhất. Vai trò này bổ sung cho HAM10000, là bộ dữ liệu ảnh dermoscopy được thu nhận bằng thiết bị chuyên dụng. Trước khi trích xuất đặc trưng, ảnh VNCancer được chuẩn hóa về kích thước 224×224 để phù hợp với đầu vào VGG16. Việc chia dữ liệu được thực hiện theo chiến lược phân tầng; các tham số chuẩn hóa đặc trưng, khoảng entropy và bước lấy mẫu tăng cường trong không gian đặc trưng chỉ được xác định từ phần huấn luyện của từng fold, không sử dụng dữ liệu xác thực để xác định các tham số này.

Các thông tin được cung cấp cho nghiên cứu gồm ảnh và nhãn lớp tương ứng. Các thuộc tính như số lượng bệnh nhân, thời gian thu nhận, model điện thoại, độ phân giải ảnh gốc, thông tin nhân khẩu học, tiêu chuẩn xác nhận chẩn đoán và mã phê duyệt đạo đức không được nêu đầy đủ trong tài liệu mô tả bộ dữ liệu hiện có và không được sử dụng trong thực nghiệm của chương này. Vì vậy, luận án không suy đoán các thông tin chưa được kiểm chứng và giới hạn mô tả VNCancer ở các thuộc tính có căn cứ từ nguồn dữ liệu và quy trình thực nghiệm.

Bộ dữ liệu HAM10000

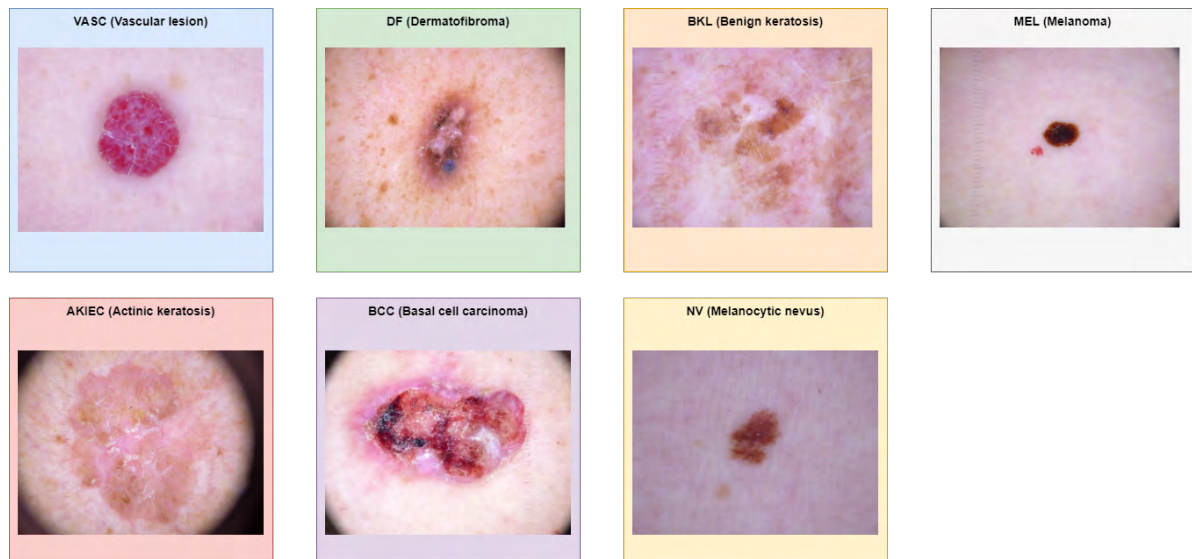
HAM10000 [3] là bộ dữ liệu dermoscopy chuẩn gồm 10015 ảnh màu của bảy lớp tổn thương da. Bảng 2.3 liệt kê các lớp cùng với số lượng ảnh tương ứng.

Bảng 2.3: Các lớp tổn thương da và phân bố số lượng trong bộ dữ liệu HAM10000

Ký hiệu	Tên tiếng Anh	Tên tiếng Việt	Số lượng
MEL	Melanoma	U hắc tố	1113
NV	Melanocytic nevus	Nốt ruồi hắc tố	6705
BCC	Basal cell carcinoma	Ung thư biểu mô tế bào đáy	514
AKIEC	Actinic keratoses and intraepithelial carcinoma	Dày sừng quang hóa và ung thư biểu mô nội biểu bì	327
BKL	Benign keratosis-like lesions	Tổn thương dày sừng lành tính	1099
DF	Dermatofibroma	U xơ da	115
VASC	Vascular lesions	Tổn thương mạch máu	142
Tổng số ảnh			10015

Ghi chú: Dữ liệu dermoscopy, các lớp được sắp xếp theo số lượng giảm dần.

Một số ảnh đại diện của các lớp trong bộ dữ liệu HAM10000 được trình bày trong Hình 2.13.



Hình 2.13: Một số ảnh mẫu trong bộ dữ liệu HAM10000 thuộc các lớp tổn thương da khác nhau.

Hai bộ dữ liệu đều có hiện tượng mất cân bằng lớp, tuy nhiên mức độ và đặc điểm mất cân bằng khác nhau. Bộ dữ liệu VNCancer có số lớp ít hơn (3 lớp) nhưng phản ánh dữ liệu lâm sàng thực tế với điều kiện chụp đa dạng; trong khi đó HAM10000 có số lớp nhiều hơn (7 lớp) và mức độ mất cân bằng rất nghiêm trọng, lớp NV chiếm tới 6705 ảnh (gần 67%) trong khi các lớp DF chỉ có 115 ảnh, VASC có 142 ảnh. Sự khác biệt này cho phép đánh giá toàn diện phương pháp BlockNet trong nhiều kịch bản dữ liệu khác nhau, đồng thời kiểm tra kết quả của cơ chế chọn lọc entropy và cân bằng dữ liệu trong không

gian đặc trưng.

Tiền xử lý dữ liệu

Trước khi đưa vào mô hình, tất cả ảnh được chuẩn hóa kích thước và xử lý phù hợp với đầu vào của mạng CNN dùng trong bước trích xuất bản đồ đặc trưng. Sau bước tiền xử lý, dữ liệu được sử dụng để:

- trích xuất bản đồ đặc trưng từ các block của VGG16,
- tính entropy và chọn lọc bản đồ đặc trưng,
- trích xuất đặc trưng thủ công,
- hợp nhất đặc trưng đa nguồn,
- huấn luyện bộ phân loại BlockNet.

Việc sử dụng đồng thời hai bộ dữ liệu với đặc điểm khác nhau giúp đánh giá toàn diện tính kết quả, tính ổn định và khả năng tổng quát hóa của phương pháp đề xuất.

Các đặc trưng sử dụng trong thực nghiệm

các đặc trưng sử dụng trong thực nghiệm được xây dựng để khai thác đồng thời khả năng biểu diễn của đặc trưng học sâu và đặc trưng thủ công từ ảnh tổn thương da. Hệ đặc trưng sử dụng trong thực nghiệm gồm hai nhóm chính: đặc trưng học sâu và đặc trưng thủ công.

Đặc trưng học sâu: Đặc trưng học sâu được trích xuất từ các bản đồ đặc trưng trung gian của mạng VGG16 huấn luyện trước. Cụ thể, BlockNet khai thác toàn bộ 13 lớp tích chập của VGG16, bao gồm:

- block1_conv1, block1_conv2,
- block2_conv1, block2_conv2,
- block3_conv1, block3_conv2, block3_conv3,
- block4_conv1, block4_conv2, block4_conv3,
- block5_conv1, block5_conv2, block5_conv3.

Đối với mỗi lớp tích chập, entropy kích hoạt được tính theo Công thức (2.13). Các bản đồ có entropy nằm trong khoảng xác định bởi Công thức (2.21) và Công thức (2.22) được giữ lại để xây dựng vector đặc trưng học sâu. Đây là quy tắc sàng lọc theo dạng phân bố kích hoạt, không phải phép xác định trước giá trị phân biệt của từng bản đồ; hiệu quả của tập được chọn được xác nhận bằng so sánh phân loại có và không có bước sàng lọc.

Sau bước chọn lọc entropy, các bản đồ đặc trưng hợp lệ được thực hiện pooling và flatten để tạo véc-tơ đặc trưng học sâu $z_i^{(d,b)}$ cho từng block. Véc-tơ đặc trưng học sâu có chiều dài 4096.

Đặc trưng thủ công: Các đặc trưng thủ công được xây dựng theo cơ sở trình bày tại Mục 1.2.2. Trong thực nghiệm, BlockNet sử dụng histogram HSV 32 chiều theo Công

thức (1.10), GLCM 4 chiều theo Công thức (1.15) và véc-tơ fractal 128 chiều theo Công thức (1.21). Do đó,

$$d_{\text{handcrafted}} = d_c + d_t + d_s = 32 + 4 + 128 = 164. \quad (2.33)$$

Mức liên quan của từng nhóm đặc trưng với nhãn được đánh giá bổ trợ bằng Chi-square theo Mục 1.2.1. Toàn bộ thao tác tính điểm chỉ sử dụng phần huấn luyện của từng fold.

Véc-tơ đặc trưng thủ công 164 chiều được chuẩn hóa rồi ghép với véc-tơ đặc trưng học sâu 4096 chiều theo Công thức (2.25). Tổng số chiều đầu vào bộ phân loại là

$$d_{\text{BlockNet}} = d_{\text{deep}} + d_{\text{handcrafted}} = 4096 + 164 = 4260. \quad (2.34)$$

Như vậy, véc-tơ đặc trưng thủ công có 164 chiều và véc-tơ hợp nhất đưa vào bộ phân loại có tổng cộng 4260 chiều.

2.5.2. Các mô hình so sánh

Để đánh giá kết quả của phương pháp đề xuất, nghiên cứu so sánh BlockNet với các phương pháp trích rút đặc trưng sâu dựa trên nhiều backbone học sâu phổ biến. Cần nhấn rằng các mô hình này không được sử dụng như các mô hình phân loại huấn luyện end-to-end, mà được sử dụng như các bộ trích rút đặc trưng. Sau khi đặc trưng được trích xuất, cùng một bộ phân loại MLP được sử dụng để huấn luyện và đánh giá trên các tập đặc trưng tương ứng. Cách thiết lập này giúp so sánh trực tiếp chất lượng biểu diễn đặc trưng giữa BlockNet và các mô hình cơ sở, đồng thời hạn chế ảnh hưởng của sự khác biệt do kiến trúc bộ phân loại gây ra.

Đối với BlockNet, đặc trưng học sâu được trích xuất từ 13 lớp tích chập của VGG16, từ *block1_conv1* đến *block5_conv3*. Với mỗi lớp, các bản đồ đặc trưng được đánh giá bằng entropy, sau đó chỉ các bản đồ nằm trong khoảng entropy phù hợp mới được giữ lại. Các bản đồ hợp lệ được chuyển đổi thành véc-tơ đặc trưng học sâu 4096 chiều. Véc-tơ này được ghép với histogram HSV 32 chiều, GLCM 4 chiều và fractal 128 chiều. Vì vậy, véc-tơ cuối cùng có $4096 + 32 + 4 + 128 = 4260$ chiều theo Công thức (2.34).

Đối với các mô hình cơ sở, đặc trưng sâu được trích xuất trực tiếp từ các backbone đã huấn luyện trước. Cụ thể, các mô hình được sử dụng gồm VGG16 và VGG19 [50], ResNet50, ResNet101 và ResNet152 [51], InceptionV3 [120], MobileNetV2 [121], EfficientNetB0 và EfficientNetB7 [54], DenseNet121, DenseNet169 và DenseNet201 [52], và Vision Transformer ViT-B16 [56]. Các backbone này được sử dụng để tạo vector đặc trưng ảnh, sau đó vector đặc trưng được đưa vào cùng một kiến trúc MLP như trong phương pháp đề xuất để bảo đảm tính công bằng khi so sánh.

Bảng 2.4 tóm tắt các mô hình trích rút đặc trưng được sử dụng trong thực nghiệm, cách lấy đặc trưng và chiều dài vector đặc trưng tương ứng.

Trong bảng trên, các mô hình cơ sở chỉ sử dụng đặc trưng sâu đơn nguồn từ backbone tương ứng. Ngược lại, BlockNet sử dụng đặc trưng học sâu đã được chọn lọc theo entropy và kết hợp với các đặc trưng thủ công miền chuyên biệt. Vì vậy, phép so sánh trong thực nghiệm không nhằm chứng minh một backbone mới cao hơn các backbone hiện có, mà nhằm đánh giá xem cơ chế chọn lọc bản đồ đặc trưng dựa trên entropy và hợp nhất đặc trưng đa nguồn có tạo ra biểu diễn kết quả hơn so với các đặc trưng sâu trích trực tiếp từ các kiến trúc phổ biến hay không.

Tất cả các vector đặc trưng sau khi trích rút đều được chuẩn hóa trước khi đưa vào

Bảng 2.4: Các mô hình trích rút đặc trưng được sử dụng trong thực nghiệm

Mô hình / phương pháp	Cách trích rút đặc trưng	Chiều dài vector
BlockNet	Bản đồ đặc trưng từ 13 lớp tích chập của VGG16, chọn lọc theo entropy và hợp nhất với đặc trưng thủ công	4260
VGG16 [50]	Đặc trưng từ lớp $fc2$ của VGG16	4096
VGG19 [50]	Đặc trưng từ lớp $fc2$ của VGG19	4096
ResNet50 [51]	Đặc trưng từ tầng pooling cuối sau khối tích chập	2048
ResNet101 [51]	Đặc trưng từ tầng pooling cuối sau khối tích chập	2048
ResNet152 [51]	Đặc trưng từ tầng pooling cuối sau khối tích chập	2048
InceptionV3 [120]	Đặc trưng từ tầng pooling toàn cục cuối	2048
MobileNetV2 [121]	Đặc trưng từ tầng pooling toàn cục cuối	1280
EfficientNetB0 [54]	Đặc trưng từ tầng pooling toàn cục cuối	1280
EfficientNetB7 [54]	Đặc trưng từ tầng pooling toàn cục cuối	2560
DenseNet121 [52]	Đặc trưng từ tầng pooling toàn cục cuối	1024
DenseNet169 [52]	Đặc trưng từ tầng pooling toàn cục cuối	1664
DenseNet201 [52]	Đặc trưng từ tầng pooling toàn cục cuối	1920
ViT-B16 [56]	Đặc trưng từ token biểu diễn cuối của Vision Transformer	768

bộ phân loại MLP. Việc sử dụng cùng một bộ phân loại và cùng giao thức huấn luyện cho tất cả các tập đặc trưng giúp bảo đảm rằng sự khác biệt về kết quả chủ yếu đến từ chất lượng của biểu diễn đặc trưng, thay vì từ sự khác biệt của kiến trúc phân loại.

2.5.3. Cấu hình thực nghiệm

Các thực nghiệm được thiết kế nhằm đánh giá công bằng kết quả của các biểu diễn đặc trưng khác nhau trong bài toán phân loại ảnh tổn thương da. Đối với tất cả các phương pháp, đặc trưng ảnh được trích rút trước, sau đó được sử dụng làm đầu vào cho cùng một bộ phân loại MLP. Cách thiết lập này giúp hạn chế ảnh hưởng của sự khác biệt về kiến trúc phân loại, từ đó tập trung đánh giá chất lượng của từng loại đặc trưng.

Đối với phương pháp BlockNet, ảnh đầu vào được chuẩn hóa kích thước về 224×224 để phù hợp với đầu vào của mạng VGG16. Bản đồ đặc trưng được trích xuất từ 13 lớp tích chập của VGG16, từ *block1_conv1* đến *block5_conv3*. Với mỗi lớp, entropy của các bản đồ đặc trưng được tính toán theo Công thức (2.13). Sau đó, khoảng entropy phù hợp được xác định theo Công thức (2.21) và Công thức (2.22). Các bản đồ đặc trưng hợp lệ được lựa chọn theo Công thức (2.24). Các bản đồ đặc trưng sau chọn lọc được chuyển đổi thành vector đặc trưng học sâu và hợp nhất với các đặc trưng thủ công để tạo thành vector đặc trưng đa nguồn.

Trong các thực nghiệm theo từng block, mỗi lớp tích chập của VGG16 tạo ra một tập đặc trưng riêng. Do đó, với BlockNet, mỗi block b tương ứng với một tập dữ liệu đặc trưng $D_{\text{balanced}}^{(b)}$. Các tập đặc trưng này được sử dụng độc lập để huấn luyện và đánh giá bộ phân loại MLP, qua đó phân tích ảnh hưởng của từng mức biểu diễn CNN đối với kết quả phân loại.

Đối với các mô hình so sánh, các backbone học sâu như VGG, ResNet, InceptionV3, MobileNetV2, EfficientNet, DenseNet và ViT-B16 được sử dụng như các bộ trích rút đặc trưng. Vector đặc trưng thu được từ từng backbone được chuẩn hóa và đưa vào cùng một

kiến trúc MLP như trong phương pháp đề xuất. Như vậy, các kết quả so sánh phản ánh sự khác biệt về chất lượng biểu diễn đặc trưng, thay vì sự khác biệt về bộ phân loại.

Bộ phân loại sử dụng trong tất cả các thực nghiệm là mô hình Multi-Layer Perceptron (MLP). Kiến trúc MLP được thiết kế gồm nhiều lớp fully-connected kết hợp với Batch Normalization, hàm kích hoạt ReLU và Dropout nhằm tăng khả năng học biểu diễn và giảm hiện tượng overfitting. Ngoài ra, regularization L2 được áp dụng cho các lớp Dense để tăng tính ổn định trong quá trình huấn luyện.

Kiến trúc chi tiết của MLP được sử dụng trong nghiên cứu được mô tả như sau:

Input $\rightarrow$ Dense(512) + L2(10^{-4})
 $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Dropout(0.4)
 $\rightarrow$ Dense(256) + L2(10^{-4})
 $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Dropout(0.3)
 $\rightarrow$ Dense(128) + L2(10^{-4})
 $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Dropout(0.3)
 $\rightarrow$ Dense(64) + L2(10^{-4})
 $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Dropout(0.2)
 $\rightarrow$ Dense(C) $\rightarrow$ Softmax.

Trong đó, C là số lớp của bài toán phân loại tương ứng với từng bộ dữ liệu. Hàm mất mát categorical cross-entropy được sử dụng trong quá trình huấn luyện, kết hợp với thuật toán tối ưu Adam có tốc độ học bằng 5×10^{-4} . Các chỉ số được sử dụng trong quá trình huấn luyện gồm Accuracy, Precision và Recall.

Trong quá trình đánh giá, dữ liệu được chia theo chiến lược stratified five-fold cross-validation nhằm bảo đảm phân bố lớp tương đối ổn định giữa các fold. Kết quả cuối cùng được báo cáo dưới dạng giá trị trung bình và độ lệch chuẩn trên các fold để giảm ảnh hưởng của việc chia dữ liệu ngẫu nhiên.

Để hạn chế rò rỉ thông tin trong quá trình đánh giá, việc chia dữ liệu được thực hiện trước các bước phụ thuộc vào dữ liệu huấn luyện của từng fold. Các tham số chuẩn hóa đặc trưng, khoảng entropy phù hợp và cơ chế cân bằng trong không gian đặc trưng được xác định trên phần huấn luyện của từng fold, sau đó mới áp dụng cho phần xác thực tương ứng. Tập xác thực không được sử dụng để xác định ngưỡng entropy, cân bằng dữ liệu hoặc lựa chọn siêu tham số của mô hình.

Đối với các lớp mất cân bằng, cơ chế cân bằng dữ liệu trong không gian đặc trưng được áp dụng sau bước trích rút và hợp nhất đặc trưng. Việc cân bằng được thực hiện trên vector đặc trưng thay vì trên ảnh đầu vào nhằm duy trì tính nhất quán với thiết kế của BlockNet.

Bảng 2.5 trình bày các cấu hình chính được sử dụng trong thực nghiệm.

Tất cả các thực nghiệm được thực hiện trên cùng môi trường huấn luyện nhằm bảo đảm tính nhất quán trong quá trình đánh giá. Việc sử dụng cùng một bộ phân loại MLP và cùng giao thức huấn luyện cho tất cả các tập đặc trưng giúp bảo đảm rằng sự khác biệt về kết quả chủ yếu đến từ chất lượng của biểu diễn đặc trưng, thay vì từ sự khác biệt của kiến trúc phân loại.

Bảng 2.5: Cấu hình thực nghiệm chính

Thành phần	Cấu hình
Kích thước ảnh đầu vào	224×224
Backbone của BlockNet	VGG16 tiền huấn luyện trên ImageNet
Các lớp VGG16 sử dụng	13 lớp tích chập từ <i>block1_conv1</i> đến <i>block5_conv3</i>
Chiến lược chọn lọc	Chọn lọc bản đồ đặc trưng dựa trên entropy
Ngưỡng entropy tối thiểu	$\theta = 0.1$
Đặc trưng thủ công	Histogram HSV (32 chiều), GLCM (4 chiều), fractal (128 chiều)
Chiều dài vector đặc trưng BlockNet	4260
Bộ phân loại	Multi-Layer Perceptron (MLP) gồm 4 lớp ẩn với Batch-Norm, ReLU và Dropout
Hàm mất mát	Categorical cross-entropy
Optimizer	Adam với tốc độ học 5×10^{-4}
Hệ số regularization	L2 với $\lambda = 10^{-4}$
Metric trong huấn luyện	Accuracy, Precision, Recall
Batch size	64
Số epoch tối đa	200
Early stopping patience	20 epoch
Giảm tốc độ học	ReduceLROnPlateau với hệ số 0.5
Chiến lược đánh giá	Stratified five-fold cross-validation

2.5.4. Các độ đo đánh giá

Để đánh giá kết quả của phương pháp đề xuất và các mô hình so sánh, nghiên cứu sử dụng các độ đo gồm Accuracy, Precision, Recall, F1-score và AUC. Các định nghĩa và công thức chi tiết của các độ đo này đã được trình bày trong mục 1.6.1 của Chương 1.

Do các bộ dữ liệu sử dụng trong nghiên cứu có hiện tượng mất cân bằng lớp, các độ đo Precision, Recall, F1-score và AUC được tính theo chiến lược macro-average nhằm phản ánh công bằng hơn kết quả của mô hình trên các lớp thiểu số.

Trong các thực nghiệm, kết quả được báo cáo dưới dạng giá trị trung bình và độ lệch chuẩn trên các fold của chiến lược stratified five-fold cross-validation. Ngoài các độ đo định lượng, ma trận nhầm lẫn cũng được sử dụng để phân tích khả năng phân loại của mô hình trên từng lớp tổn thương da.

2.6. Kết quả thực nghiệm

2.6.1. Đánh giá ảnh hưởng của chọn lọc bản đồ đặc trưng dựa trên entropy

Mục này đánh giá ảnh hưởng của cơ chế chọn lọc bản đồ đặc trưng dựa trên entropy đối với chất lượng biểu diễn đặc trưng học sâu trong BlockNet. Hai thiết lập được xem xét gồm: (i) không sử dụng chọn lọc entropy và (ii) sử dụng chọn lọc entropy.

Trong thiết lập không sử dụng entropy, toàn bộ bản đồ đặc trưng của từng block được sử dụng để xây dựng vector đặc trưng. Ngược lại, trong thiết lập sử dụng entropy, chỉ các bản đồ đặc trưng thỏa mãn khoảng entropy phù hợp mới được giữ lại trước bước xây dựng vector đặc trưng.

Entropy không được sử dụng như một độ đo không chắc chắn đầu ra của mô hình, mà được áp dụng trên bản đồ đặc trưng để mô tả phân bố kích hoạt của biểu diễn trung

gian. Cơ chế chọn lọc sàng lọc các bản đồ nằm ngoài khoảng entropy được xác định cho từng block; khả năng phân biệt của tập được giữ lại được đánh giá qua kết quả phân loại trước và sau chọn lọc.

Các kết quả trong mục này được trình bày trên hai bộ dữ liệu VNCancer và HAM10000, sử dụng các chỉ số Accuracy, Precision, Recall, F1-score, Sensitivity và Specificity nhằm bảo đảm tính thống nhất với các thực nghiệm ở các mục tiếp theo.

Cần lưu ý rằng các kết quả trong mục này thuộc nhóm thực nghiệm chẩn đoán nhằm khảo sát riêng sự thay đổi của biểu diễn đặc trưng học sâu khi bật hoặc tắt cơ chế chọn lọc entropy theo từng lớp tích chập. Vì vậy, các giá trị trong hai thiết lập chỉ được dùng để đối chiếu trong phạm vi thí nghiệm này và không đại diện cho kết quả cuối cùng của toàn bộ quy trình BlockNet.

2.6.1.1. Kết quả trên bộ dữ liệu VNCancer

1) Kết quả khi không sử dụng chọn lọc entropy

Bảng 2.6 trình bày kết quả phân loại trên bộ dữ liệu VNCancer khi không sử dụng chọn lọc entropy. Trong thiết lập này, toàn bộ bản đồ đặc trưng của từng block được sử dụng để tạo vector đặc trưng.

Kết quả cho thấy các lớp tích chập nông như block1_conv1, block1_conv2 và block3_conv1 đạt hiệu năng cao hơn so với nhiều lớp sâu hơn. Cấu hình cao nhất đạt được tại block1_conv2, với Accuracy bằng 80.97% và F1-score bằng 82.11%.

Đáng chú ý, các lớp tích chập sâu như block4_conv3 và block5_conv3 cho kết quả thấp hơn. Điều này cho thấy việc sử dụng toàn bộ bản đồ đặc trưng ở các tầng sâu có thể tạo ra biểu diễn chứa nhiều thông tin dư thừa hoặc phản hồi không ổn định.

Bảng 2.6: Kết quả trên bộ dữ liệu VNCancer khi không sử dụng chọn lọc entropy

Block	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
block1_conv1	80.62	81.94	80.62	82.08	80.62	87.51
block1_conv2	80.97	82.05	80.97	82.11	80.97	87.85
block2_conv1	73.57	72.80	73.57	72.14	73.57	82.33
block2_conv2	79.41	79.62	79.41	79.85	79.41	86.72
block3_conv1	80.80	81.58	80.80	81.84	80.80	87.43
block3_conv2	78.37	78.84	78.37	79.03	78.37	85.14
block3_conv3	75.43	76.01	75.43	76.90	75.43	83.02
block4_conv1	65.34	66.12	65.34	68.32	65.34	77.54
block4_conv2	69.90	70.31	69.90	71.61	69.90	80.45
block4_conv3	49.73	53.45	49.73	57.63	49.73	69.11
block5_conv1	65.22	66.44	65.22	68.67	65.22	77.85
block5_conv2	66.26	66.88	66.26	68.75	66.26	78.42
block5_conv3	58.43	59.37	58.43	60.48	58.43	73.18

2) Kết quả khi sử dụng chọn lọc entropy

Bảng 2.7 trình bày kết quả trên cùng bộ dữ liệu khi sử dụng chọn lọc entropy.

Kết quả cho thấy hiệu năng được cải thiện ở hầu hết các block. Cấu hình cao nhất đạt được tại block2_conv2, với Accuracy bằng 88.75%, F1-score bằng 88.16% và Specificity bằng 93.42%.

So với thiết lập không sử dụng entropy, Accuracy cao nhất tăng từ 80.97% lên

88.75%, tương ứng mức tăng 7.78 điểm phần trăm. Đồng thời, F1-score tăng từ 82.11% lên 88.16%, tương ứng mức tăng 6.05 điểm phần trăm.

Đặc biệt, các lớp tích chập sâu có mức cải thiện lớn. Ví dụ, tại block4_conv3, Accuracy tăng từ 49.73% lên 86.85%, trong khi F1-score tăng từ 57.63% lên 86.04%. Kết quả này cung cấp bằng chứng thực nghiệm rằng tập biểu diễn thu được sau quy tắc sàng lọc entropy phù hợp hơn với bộ phân loại trong cấu hình đang xét; không suy ra rằng entropy riêng lẻ đo trực tiếp tính hữu ích của từng bản đồ.

Bảng 2.7: Kết quả trên bộ dữ liệu VNCancer khi sử dụng chọn lọc entropy

Block	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
block1_conv1	86.85	86.74	86.85	86.54	86.85	92.34
block1_conv2	88.41	88.02	88.41	87.58	88.41	93.12
block2_conv1	86.51	86.12	86.51	85.60	86.51	92.18
block2_conv2	88.75	88.40	88.75	88.16	88.75	93.42
block3_conv1	88.24	87.95	88.24	87.57	88.24	93.05
block3_conv2	86.33	85.92	86.33	85.44	86.33	92.10
block3_conv3	83.56	83.22	83.56	82.86	83.56	90.84
block4_conv1	86.33	86.01	86.33	85.60	86.33	92.21
block4_conv2	83.74	83.31	83.74	83.01	83.74	90.95
block4_conv3	86.85	86.44	86.85	86.04	86.85	92.63
block5_conv1	86.33	86.05	86.33	85.69	86.33	92.18
block5_conv2	83.91	83.44	83.91	83.10	83.91	91.04
block5_conv3	82.18	81.97	82.18	81.97	82.18	90.12

3) Phân tích ROC trước và sau khi sử dụng entropy

Hình 2.14 minh họa đường cong ROC trên bộ dữ liệu VNCancer trước và sau khi sử dụng chọn lọc entropy.

Trong trường hợp không sử dụng entropy, mô hình tại block1_conv2 đạt AUC lần lượt là 0.8857, 0.8711 và 0.8828 đối với các lớp BCC, MM và SCC. Sau khi sử dụng entropy, mô hình tại block2_conv2 đạt AUC tương ứng là 0.9415, 0.8457 và 0.9462.

Kết quả này ghi nhận sự thay đổi khả năng phân biệt đối với các lớp BCC và SCC sau khi áp dụng chọn lọc entropy. Trong khi đó, lớp MM vẫn là lớp khó hơn do số lượng mẫu ít hơn và độ biến thiên hình thái lớn hơn.

2.6.1.2. Kết quả trên bộ dữ liệu HAM10000

1) Kết quả khi không sử dụng chọn lọc entropy

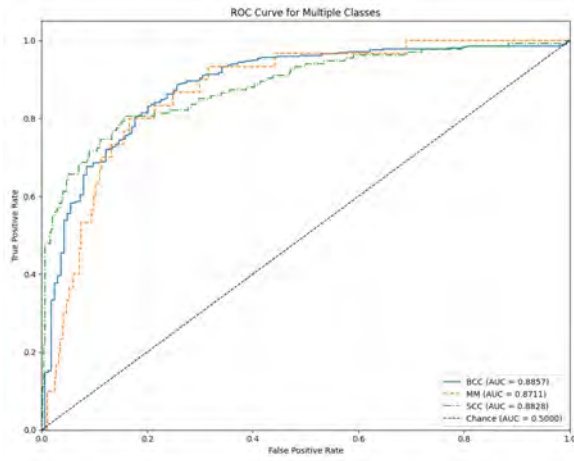
Bảng 2.8 trình bày kết quả trên HAM10000 khi không sử dụng chọn lọc entropy.

Kết quả cho thấy hiệu năng giữa các block có sự khác biệt, trong đó các lớp tích chập đầu và lớp tích chập trung gian đạt kết quả cao hơn so với các lớp sâu hơn. Cấu hình cao nhất đạt được tại block2_conv2, với Accuracy bằng 69.07% và F1-score bằng 63.78%.

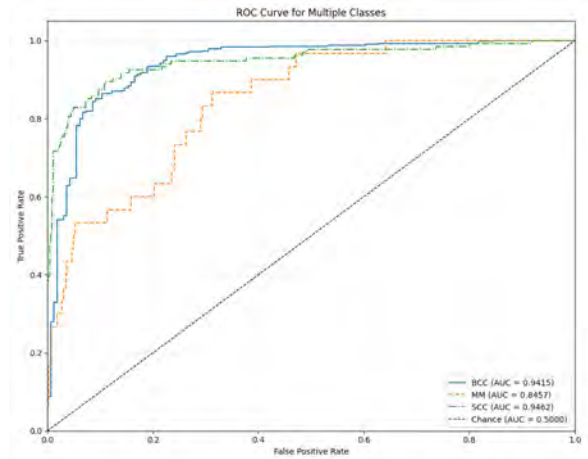
2) Kết quả khi sử dụng chọn lọc entropy

Bảng 2.9 trình bày kết quả phân loại trên HAM10000 khi sử dụng chọn lọc entropy.

Kết quả cao nhất đạt được tại block2_conv1, với Accuracy bằng 70.00% và F1-score bằng 65.00%. So với thiết lập không sử dụng entropy, F1-score tăng từ 63.78% lên 65.00%.



(a) Không sử dụng entropy, *block1_conv2*



(b) Sử dụng entropy, *block2_conv2*

Hình 2.14: So sánh ROC trên bộ dữ liệu VNCancer trước và sau khi sử dụng chọn lọc entropy.

Bảng 2.8: Kết quả trên bộ dữ liệu HAM10000 khi không sử dụng chọn lọc entropy

Block	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
block1_conv1	68.38	66.84	68.38	62.92	68.38	84.21
block1_conv2	69.04	67.52	69.04	61.97	69.04	84.45
block2_conv1	69.08	67.61	69.08	62.06	69.08	84.48
block2_conv2	69.07	68.22	69.07	63.78	69.07	84.72
block3_conv1	67.43	64.84	67.43	55.81	67.43	82.10
block3_conv2	68.69	66.20	68.69	60.78	68.69	83.52
block3_conv3	67.17	65.03	67.17	57.60	67.17	82.61
block4_conv1	67.03	63.74	67.03	54.35	67.03	81.74
block4_conv2	67.12	64.81	67.12	56.31	67.12	82.03
block4_conv3	65.79	62.30	65.79	52.29	65.79	80.44
block5_conv1	66.80	63.72	66.80	55.28	66.80	81.56
block5_conv2	66.25	62.85	66.25	52.99	66.25	80.95
block5_conv3	63.32	59.74	63.32	49.80	63.32	78.62

Mặc dù mức cải thiện trên HAM10000 không lớn như trên VNCancer, kết quả này vẫn cho thấy chọn lọc entropy giúp tăng tính ổn định của biểu diễn đặc trưng trong bài toán phân loại đa lớp phức tạp và có mức mất cân bằng lớp cao hơn.

3) Phân tích ROC trước và sau khi sử dụng entropy

Hình 2.15 minh họa đường cong ROC trên bộ dữ liệu HAM10000 trước và sau khi sử dụng chọn lọc entropy.

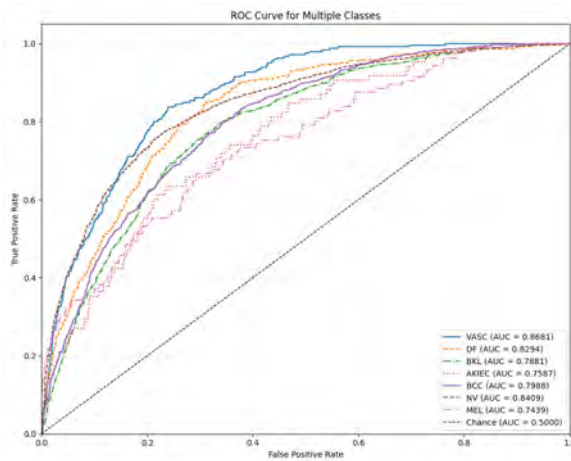
Trong thiết lập không sử dụng entropy, các giá trị AUC có sự khác biệt giữa các lớp, trong đó một số lớp như MEL và AKIEC có AUC thấp hơn. Sau khi sử dụng entropy, AUC của nhiều lớp được cải thiện, lớp VASC tăng từ 0.8681 lên 0.9179, lớp DF tăng từ 0.8294 lên 0.8440, lớp AKIEC tăng từ 0.7587 lên 0.7727, và lớp NV tăng từ 0.8409 lên 0.8465.

Kết quả này ghi nhận sự thay đổi khả năng phân biệt đối với một số lớp tổn thương

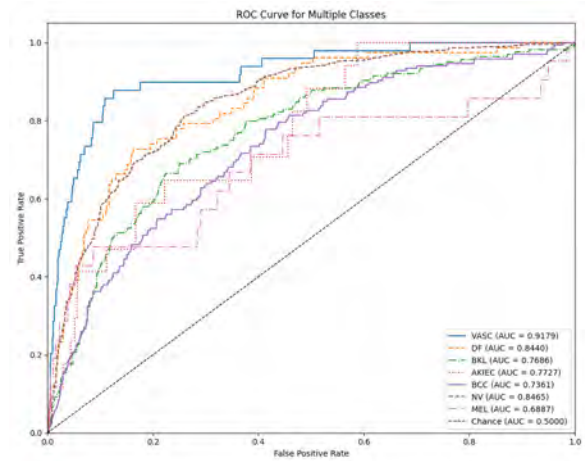
Bảng 2.9: Kết quả trên bộ dữ liệu HAM10000 khi sử dụng chọn lọc entropy

Block	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
block1_conv1	67.00	65.00	67.00	64.00	67.00	84.82
block1_conv2	70.00	66.00	70.00	64.00	70.00	85.31
block2_conv1	70.00	67.00	70.00	65.00	70.00	85.62
block2_conv2	68.00	66.00	68.00	65.00	68.00	84.94
block3_conv1	68.00	65.00	68.00	64.00	68.00	84.72
block3_conv2	69.00	66.00	69.00	64.00	69.00	85.06
block3_conv3	66.00	64.00	66.00	64.00	66.00	84.15
block4_conv1	66.00	64.00	66.00	64.00	66.00	84.06
block4_conv2	67.00	65.00	67.00	65.00	67.00	84.63
block4_conv3	67.00	63.00	67.00	63.00	67.00	83.92
block5_conv1	67.00	63.00	67.00	63.00	67.00	83.81
block5_conv2	66.00	62.00	66.00	62.00	66.00	83.24
block5_conv3	65.00	64.00	65.00	64.00	65.00	82.95

khó, mặc dù các lớp như MEL vẫn còn nhiều thách thức do số lượng mẫu hạn chế và mức độ tương đồng hình thái với các lớp khác.



(a) Không sử dụng entropy, block2_conv2



(b) Sử dụng entropy, block2_conv1

Hình 2.15: So sánh ROC trên bộ dữ liệu HAM10000 trước và sau khi sử dụng chọn lọc entropy.

2.6.2. Đánh giá kết quả hợp nhất đặc trưng đa nguồn

Mục này đánh giá kết quả của phương pháp hợp nhất đặc trưng đa nguồn trong BlockNet. Trong mô hình đề xuất, vector đặc trưng cuối cùng được xây dựng bằng cách kết hợp đặc trưng học sâu sau chọn lọc entropy từ các bản đồ đặc trưng của từng lớp tích chập trong VGG16 với các đặc trưng thủ công phản ánh màu sắc, kết cấu, hình dạng và độ phức tạp hình học của tổn thương da.

Các thực nghiệm được thực hiện trên hai bộ dữ liệu VNCancer và HAM10000. Các mô hình cơ sở sử dụng đặc trưng được trích xuất từ các backbone học sâu phổ biến và được huấn luyện bằng cùng bộ phân loại MLP. Do đó, phép so sánh tập trung vào chất

lượng biểu diễn đặc trưng thay vì sự khác biệt của bộ phân loại.

2.6.2.1. Đánh giá hiệu năng trên bộ dữ liệu VNCancer

Bộ dữ liệu VNCancer gồm ba lớp ung thư da chính: BCC, MM và SCC. Đây là bộ dữ liệu chuyên biệt, được sử dụng để đánh giá khả năng phân loại ảnh tổn thương da trong điều kiện ảnh thực tế.

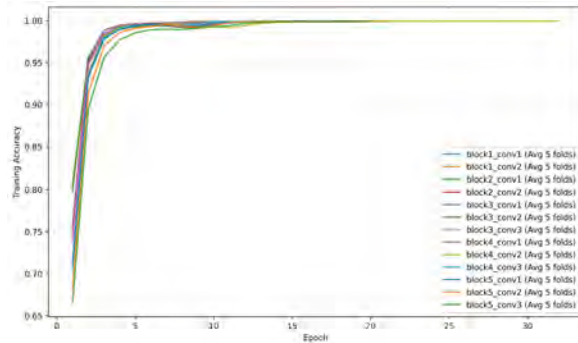
1) So sánh hiệu năng giữa BlockNet và các mô hình SOTA

Bảng 2.10 trình bày kết quả so sánh giữa BlockNet và các mô hình cơ sở trên bộ dữ liệu VNCancer. Kết quả cho thấy các mô hình VGG16, VGG19 và các biến thể ResNet đạt Accuracy khoảng 92% – 93%. Trong đó, ResNet101 là mô hình cơ sở cao nhất với Accuracy đạt 93.59%. Một số mô hình như InceptionV3 và ViT-B16 cho kết quả thấp hơn trên bộ dữ liệu này.

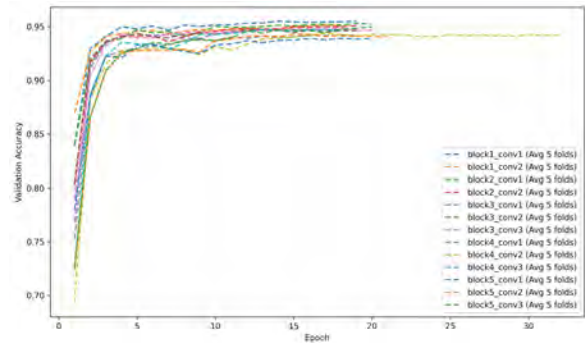
BlockNet đạt kết quả cao hơn các mô hình cơ sở trên toàn bộ các chỉ số đánh giá. Cấu hình cao nhất sử dụng đặc trưng từ lớp block1_conv1, đạt Accuracy bằng $95.58\% \pm 0.69\%$, F1-score bằng $95.59\% \pm 0.69\%$ và Specificity bằng $97.53\% \pm 0.37\%$. Kết quả này cho thấy việc hợp nhất đặc trưng học sâu sau chọn lọc entropy với đặc trưng thủ công giúp nâng cao chất lượng biểu diễn đầu vào cho bộ phân loại.

2) Phân tích theo từng lớp tích chập

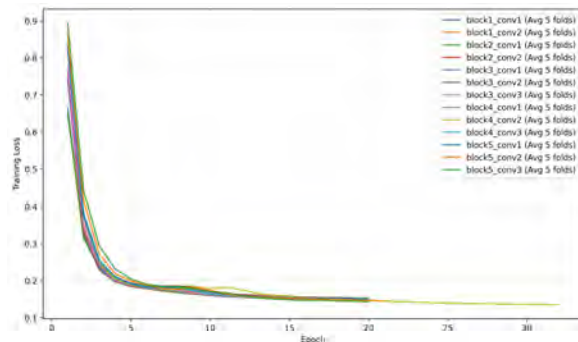
Để đánh giá ảnh hưởng của độ sâu đặc trưng, nghiên cứu khảo sát toàn bộ 13 lớp tích chập của VGG16 trong BlockNet. Hình 2.16 trình bày động học huấn luyện của mô hình trên VNCancer thiết lập 200 epochs.



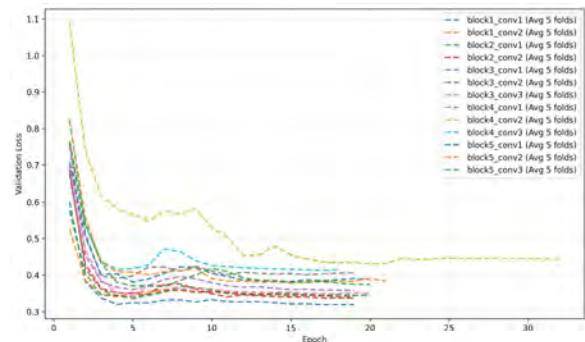
(a) Accuracy trên tập huấn luyện



(b) Accuracy trên tập xác thực



(c) Loss trên tập huấn luyện



(d) Loss trên tập xác thực

Hình 2.16: Động học huấn luyện của BlockNet trên bộ dữ liệu VNCancer.

Các lớp tích chập nông như block1_conv1 và block1_conv2 đạt các chỉ số cao nhất trong thiết lập này. Kết quả cho thấy biểu diễn được tạo từ tập bản đồ sau sàng lọc

Bảng 2.10: So sánh hiệu năng trên bộ dữ liệu VNCancer sử dụng kiểm định chéo 5 folds. Kết quả được trình bày dưới dạng trung bình $\pm$ độ lệch chuẩn (%).

Mô hình / Lớp	Accuracy (%)	Precision (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
Các mô hình cơ sở					
VGG16	92.19 $\pm$ 0.48	92.16 $\pm$ 0.49	92.15 $\pm$ 0.47	92.19 $\pm$ 0.48	89.78 $\pm$ 1.54
VGG19	92.21 $\pm$ 0.81	92.25 $\pm$ 0.79	92.18 $\pm$ 0.83	92.21 $\pm$ 0.81	89.25 $\pm$ 1.79
ResNet50	92.83 $\pm$ 0.72	92.83 $\pm$ 0.68	92.81 $\pm$ 0.70	92.83 $\pm$ 0.72	90.97 $\pm$ 0.38
ResNet152	92.86 $\pm$ 0.55	92.93 $\pm$ 0.43	92.86 $\pm$ 0.48	92.86 $\pm$ 0.55	91.06 $\pm$ 1.08
ResNet101	93.59 $\pm$ 0.78	93.56 $\pm$ 0.71	93.53 $\pm$ 0.76	93.59 $\pm$ 0.78	91.13 $\pm$ 0.44
InceptionV3	80.68 $\pm$ 1.14	79.66 $\pm$ 1.26	80.02 $\pm$ 1.24	80.68 $\pm$ 1.14	72.76 $\pm$ 2.61
ViT-B16	85.64 $\pm$ 0.90	85.20 $\pm$ 0.81	85.31 $\pm$ 0.76	85.64 $\pm$ 0.90	79.66 $\pm$ 0.70
MobileNetV2	88.96 $\pm$ 0.83	88.76 $\pm$ 0.81	88.78 $\pm$ 0.78	88.96 $\pm$ 0.83	85.74 $\pm$ 1.68
EfficientNetB7	92.55 $\pm$ 0.54	92.65 $\pm$ 0.54	92.57 $\pm$ 0.54	92.55 $\pm$ 0.54	91.15 $\pm$ 1.64
EfficientNetB0	91.80 $\pm$ 0.77	91.80 $\pm$ 0.77	91.78 $\pm$ 0.77	91.80 $\pm$ 0.77	89.80 $\pm$ 1.22
DenseNet121	89.06 $\pm$ 1.04	89.01 $\pm$ 1.23	88.97 $\pm$ 1.11	89.06 $\pm$ 1.04	86.71 $\pm$ 1.96
DenseNet169	89.19 $\pm$ 1.33	88.97 $\pm$ 1.36	89.00 $\pm$ 1.33	89.19 $\pm$ 1.33	85.20 $\pm$ 1.95
DenseNet201	88.26 $\pm$ 1.15	88.03 $\pm$ 1.14	88.04 $\pm$ 1.11	88.26 $\pm$ 1.15	83.96 $\pm$ 1.58
Mô hình BlockNet đề xuất					
block1_conv1	95.58 $\pm$ 0.69	95.59 $\pm$ 0.69	95.59 $\pm$ 0.69	95.58 $\pm$ 0.69	97.53 $\pm$ 0.37
block1_conv2	95.21 $\pm$ 0.81	95.30 $\pm$ 0.80	95.22 $\pm$ 0.81	95.21 $\pm$ 0.81	97.32 $\pm$ 0.43
block2_conv1	94.90 $\pm$ 0.19	95.00 $\pm$ 0.16	94.90 $\pm$ 0.19	94.90 $\pm$ 0.19	97.15 $\pm$ 0.14
block2_conv2	94.52 $\pm$ 0.28	94.55 $\pm$ 0.30	94.52 $\pm$ 0.30	94.52 $\pm$ 0.28	96.95 $\pm$ 0.13
block3_conv1	94.70 $\pm$ 0.30	94.72 $\pm$ 0.30	94.70 $\pm$ 0.29	94.70 $\pm$ 0.30	97.04 $\pm$ 0.20
block3_conv2	94.90 $\pm$ 0.26	94.91 $\pm$ 0.26	94.90 $\pm$ 0.26	94.90 $\pm$ 0.26	97.17 $\pm$ 0.16
block3_conv3	94.67 $\pm$ 0.51	94.68 $\pm$ 0.51	94.67 $\pm$ 0.51	94.67 $\pm$ 0.51	97.04 $\pm$ 0.29
block4_conv1	94.05 $\pm$ 0.37	94.07 $\pm$ 0.37	94.05 $\pm$ 0.37	94.05 $\pm$ 0.37	96.72 $\pm$ 0.15
block4_conv2	93.82 $\pm$ 0.61	93.83 $\pm$ 0.62	93.81 $\pm$ 0.61	93.82 $\pm$ 0.61	96.62 $\pm$ 0.32
block4_conv3	94.33 $\pm$ 0.82	94.33 $\pm$ 0.83	94.33 $\pm$ 0.82	94.33 $\pm$ 0.82	96.88 $\pm$ 0.45
block5_conv1	93.75 $\pm$ 0.78	93.76 $\pm$ 0.79	93.75 $\pm$ 0.78	93.75 $\pm$ 0.78	96.53 $\pm$ 0.41
block5_conv2	93.69 $\pm$ 0.56	93.70 $\pm$ 0.57	93.68 $\pm$ 0.56	93.69 $\pm$ 0.56	96.52 $\pm$ 0.31
block5_conv3	94.66 $\pm$ 0.79	94.68 $\pm$ 0.78	94.67 $\pm$ 0.78	94.66 $\pm$ 0.79	97.00 $\pm$ 0.47

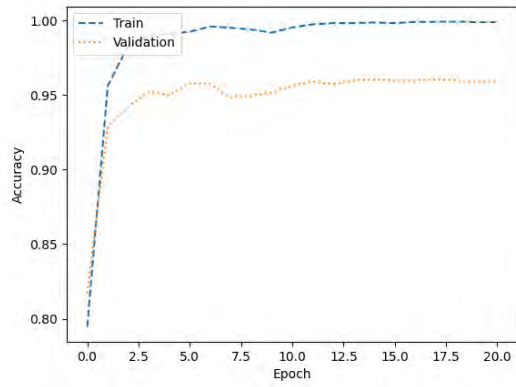
entropy và đặc trưng thủ công phù hợp với bài toán phân loại trên VNCancer; nhận định này áp dụng cho biểu diễn kết hợp, không quy trực tiếp khả năng phân biệt cho entropy của từng bản đồ.

Ngoài ra, các đường huấn luyện ghi nhận sự ổn định sau một số epoch đầu. Accuracy trên tập huấn luyện và tập xác thực tăng theo xu hướng tương đồng, đồng thời khoảng cách giữa hai đường này không lớn. Trong thiết lập thực nghiệm hiện tại, chưa ghi nhận dấu hiệu quá khớp qua hai đường accuracy; nhận định này chỉ áp dụng trong phạm vi bộ dữ liệu, cách chia tập và quy trình đánh giá đã sử dụng trong luận án.

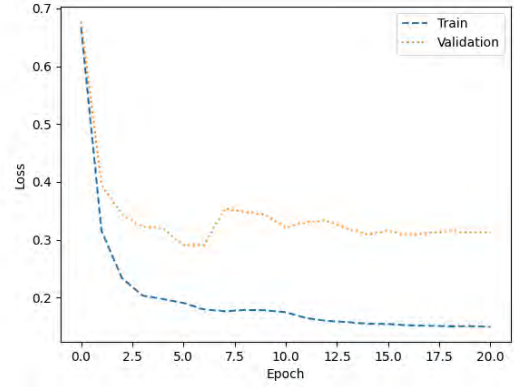
3) Phân tích cấu hình cao nhất

Hình 2.17 trình bày phân tích chi tiết cấu hình cao nhất của BlockNet trên VNCancer, sử dụng đặc trưng từ lớp block1_conv1.

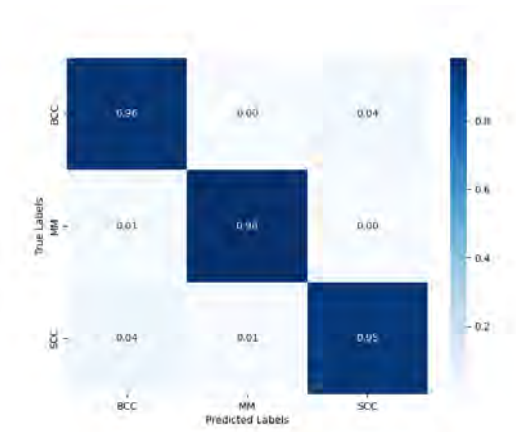
Ma trận nhầm lẫn ghi nhận kết quả phân loại trên ba lớp BCC, MM và SCC. Đồng thời, các đường cong ROC có giá trị AUC đều lớn hơn 99%.



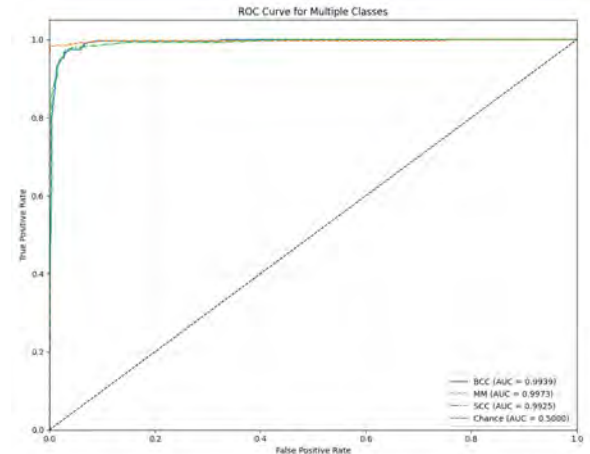
(a) Accuracy



(b) Loss



(c) Ma trận nhầm lẫn



(d) Đường cong ROC

Hình 2.17: Phân tích chi tiết cấu hình BlockNet cao nhất trên bộ dữ liệu VNCancer.

2.6.2.2. Đánh giá hiệu năng trên bộ dữ liệu HAM10000

HAM10000 là bộ dữ liệu khó hơn do bao gồm bảy lớp tổn thương da và có mức độ chồng lấn thị giác lớn giữa nhiều lớp khác nhau.

1) So sánh hiệu năng giữa BlockNet và các mô hình SOTA

Bảng 2.11 trình bày kết quả so sánh giữa BlockNet và các mô hình cơ sở trên HAM10000. Kết quả cho thấy các mô hình cơ sở có kết quả cao nhất, như ResNet50 và ResNet101, đạt Accuracy khoảng 80.8%, trong khi các mô hình khác như VGG16, DenseNet và MobileNetV2 đạt kết quả thấp hơn.

BlockNet cải thiện trên bài toán phân loại đa lớp phức tạp này. Cấu hình cao nhất sử dụng lớp `block1_conv2`, đạt Accuracy bằng $96.02\% \pm 0.24\%$, F1-score bằng $96.04\% \pm 0.23\%$ và Specificity bằng $99.21\% \pm 0.04\%$. Trong thiết lập thực nghiệm của luận án, kết quả này cao hơn mô hình cơ sở cao nhất là ResNet50 khoảng 15 điểm phần trăm theo Accuracy.

2) Phân tích theo từng lớp tích chập

Hình 2.18 trình bày động học huấn luyện của BlockNet trên bộ dữ liệu HAM10000.

Khác với VNCancer, bài toán HAM10000 yêu cầu biểu diễn đặc trưng sâu hơn để xử lý mức độ chồng lấn giữa các lớp tổn thương. Điều này thể hiện qua việc `block1_conv2`

Bảng 2.11: So sánh hiệu năng trên bộ dữ liệu HAM10000 sử dụng kiểm định chéo 5 folds. Kết quả được trình bày dưới dạng trung bình $\pm$ độ lệch chuẩn (%).

Mô hình / Lớp	Accuracy (%)	Precision (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
Các mô hình cơ sở					
ResNet50	80.81 $\pm$ 0.44	81.13 $\pm$ 0.42	80.92 $\pm$ 0.41	80.81 $\pm$ 0.44	87.06 $\pm$ 1.00
ResNet101	80.80 $\pm$ 0.81	81.15 $\pm$ 0.64	80.91 $\pm$ 0.73	80.80 $\pm$ 0.81	87.06 $\pm$ 0.90
EfficientNetB7	78.79 $\pm$ 0.75	80.37 $\pm$ 0.71	79.43 $\pm$ 0.71	78.79 $\pm$ 0.75	89.55 $\pm$ 0.66
EfficientNetB0	76.14 $\pm$ 0.96	78.25 $\pm$ 0.48	76.98 $\pm$ 0.77	76.14 $\pm$ 0.96	88.28 $\pm$ 0.57
VGG16	75.87 $\pm$ 0.61	76.69 $\pm$ 0.43	76.21 $\pm$ 0.50	75.87 $\pm$ 0.61	84.92 $\pm$ 0.84
DenseNet201	74.34 $\pm$ 0.66	78.45 $\pm$ 0.79	75.85 $\pm$ 0.65	74.34 $\pm$ 0.66	90.33 $\pm$ 1.05
VGG19	74.57 $\pm$ 0.73	75.47 $\pm$ 0.43	74.94 $\pm$ 0.52	74.57 $\pm$ 0.73	84.08 $\pm$ 0.76
DenseNet169	72.36 $\pm$ 1.38	77.94 $\pm$ 0.42	74.30 $\pm$ 1.09	72.36 $\pm$ 1.38	90.90 $\pm$ 0.71
DenseNet121	72.25 $\pm$ 1.05	77.56 $\pm$ 0.56	74.09 $\pm$ 0.84	72.25 $\pm$ 1.05	90.28 $\pm$ 0.68
MobileNetV2	67.18 $\pm$ 1.29	71.78 $\pm$ 0.82	68.97 $\pm$ 0.62	67.18 $\pm$ 1.29	84.68 $\pm$ 2.45
ViT-B16	66.46 $\pm$ 1.54	75.91 $\pm$ 0.79	69.52 $\pm$ 1.22	66.46 $\pm$ 1.54	91.80 $\pm$ 0.90
InceptionV3	57.33 $\pm$ 0.33	70.23 $\pm$ 0.76	61.47 $\pm$ 0.38	57.33 $\pm$ 0.33	87.78 $\pm$ 1.01
Mô hình BlockNet đề xuất					
block1_conv2	96.02 $\pm$ 0.24	96.06 $\pm$ 0.23	96.04 $\pm$ 0.23	96.02 $\pm$ 0.24	99.21 $\pm$ 0.04
block1_conv1	95.66 $\pm$ 0.14	95.68 $\pm$ 0.14	95.67 $\pm$ 0.14	95.66 $\pm$ 0.14	99.04 $\pm$ 0.02
block2_conv1	95.86 $\pm$ 0.11	95.91 $\pm$ 0.12	95.88 $\pm$ 0.11	95.86 $\pm$ 0.11	99.20 $\pm$ 0.02
block2_conv2	95.76 $\pm$ 0.20	95.79 $\pm$ 0.20	95.77 $\pm$ 0.20	95.76 $\pm$ 0.20	99.18 $\pm$ 0.04
block3_conv2	95.68 $\pm$ 0.24	95.70 $\pm$ 0.24	95.69 $\pm$ 0.24	95.68 $\pm$ 0.24	99.17 $\pm$ 0.04
block3_conv3	95.74 $\pm$ 0.28	95.74 $\pm$ 0.28	95.73 $\pm$ 0.28	95.74 $\pm$ 0.28	99.20 $\pm$ 0.06
block3_conv1	95.58 $\pm$ 0.24	95.59 $\pm$ 0.25	95.58 $\pm$ 0.25	95.58 $\pm$ 0.24	99.16 $\pm$ 0.05
block4_conv3	95.47 $\pm$ 0.18	95.47 $\pm$ 0.19	95.47 $\pm$ 0.18	95.47 $\pm$ 0.18	99.14 $\pm$ 0.04
block5_conv3	95.38 $\pm$ 0.40	95.38 $\pm$ 0.40	95.36 $\pm$ 0.41	95.38 $\pm$ 0.40	99.14 $\pm$ 0.07
block5_conv2	95.28 $\pm$ 0.25	95.28 $\pm$ 0.26	95.26 $\pm$ 0.26	95.28 $\pm$ 0.25	99.12 $\pm$ 0.05
block5_conv1	95.24 $\pm$ 0.23	95.26 $\pm$ 0.23	95.25 $\pm$ 0.23	95.24 $\pm$ 0.23	99.09 $\pm$ 0.05
block4_conv1	95.40 $\pm$ 0.20	95.40 $\pm$ 0.21	95.38 $\pm$ 0.20	95.40 $\pm$ 0.20	99.14 $\pm$ 0.04
block4_conv2	95.22 $\pm$ 0.27	95.23 $\pm$ 0.27	95.20 $\pm$ 0.27	95.22 $\pm$ 0.27	99.11 $\pm$ 0.05

và block2_conv1 cho kết quả cao hơn so với block1_conv1. Tuy nhiên, các lớp quá sâu như block5_conv1 hoặc block5_conv2 có xu hướng giảm nhẹ hiệu năng. Điều này cho thấy bản đồ đặc trưng quá sâu có thể làm mất một phần thông tin cục bộ đối với bài toán phân loại tổn thương da.

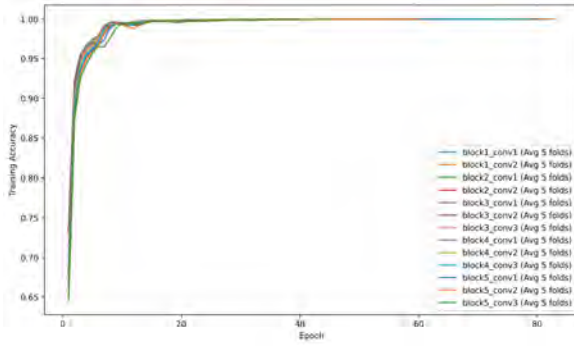
3) Phân tích cấu hình cao nhất

Hình 2.19 trình bày kết quả chi tiết của cấu hình cao nhất trên HAM10000.

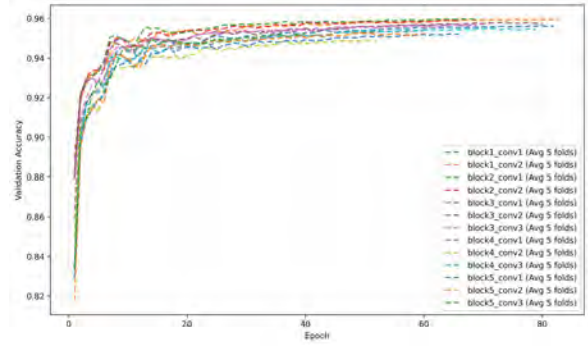
Ma trận nhầm lẫn cho thấy mô hình đạt hiệu năng cao trên cả bảy lớp tổn thương da. Đồng thời, các đường cong ROC cho thấy giá trị AUC đều lớn hơn 99%, bao gồm cả lớp MEL, phản ánh khả năng phân biệt cao đối với các tổn thương ác tính nguy hiểm.

4) So sánh với các nghiên cứu gần đây

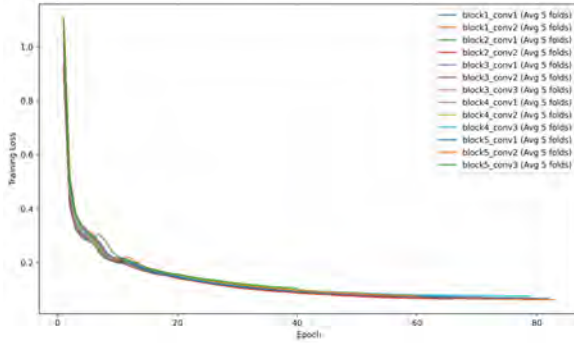
Bảng 2.12 trình bày kết quả so sánh giữa BlockNet và một số nghiên cứu gần đây trên HAM10000. Kết quả cho thấy BlockNet đạt Accuracy cao hơn một số nghiên cứu gần đây trong bảng so sánh, đồng thời cung cấp đầy đủ hơn các chỉ số đánh giá như Precision, F1-score, Sensitivity và Specificity.



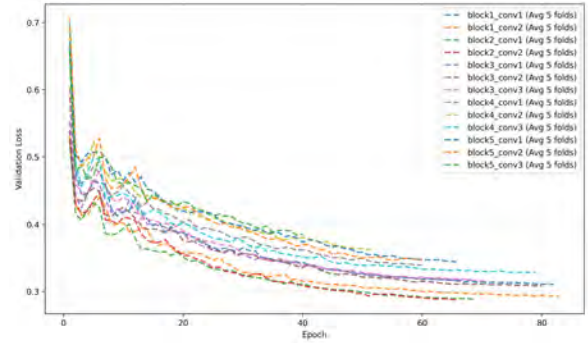
(a) Accuracy trên tập huấn luyện



(b) Accuracy trên tập xác thực



(c) Loss trên tập huấn luyện



(d) Loss trên tập xác thực

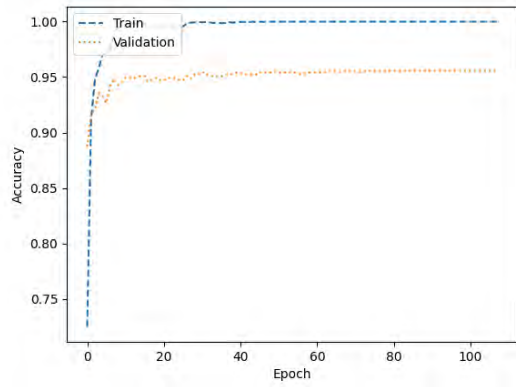
Hình 2.18: Động học huấn luyện của BlockNet trên bộ dữ liệu HAM10000.**Bảng 2.12:** So sánh giữa BlockNet và một số nghiên cứu gần đây trên HAM10000.

Phương pháp	Accuracy (%)	Năm
Yanchatua et al. [122]	90.34	2021
Jasil et al. [123]	95.00	2023
Shetty et al. [124]	95.18	2022
Wei et al. [125]	95.29	2023
BlockNet (block1_conv2)	96.02	—

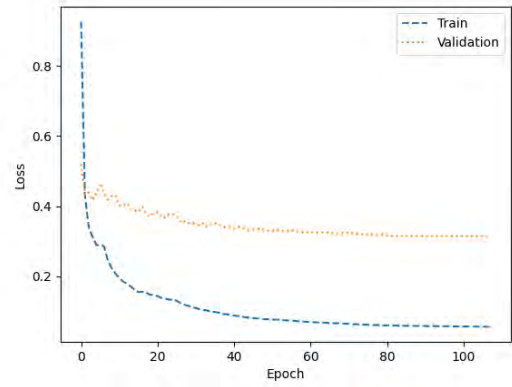
2.7. Phân tích và thảo luận kết quả thực nghiệm

Các kết quả thực nghiệm trên hai bộ dữ liệu VNCancer và HAM10000 ghi nhận sự khác biệt giữa BlockNet và các phương pháp sử dụng trực tiếp đặc trưng học sâu hoặc các backbone CNN phổ biến. Sự khác biệt được thể hiện qua Accuracy, F1-score, Sensitivity và AUC theo từng lớp.

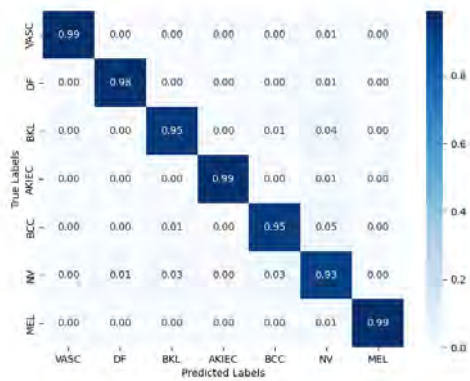
Các thực nghiệm ban đầu không được thiết kế như một nghiên cứu loại trừ toàn phần (full ablation study), trong đó lần lượt loại bỏ độc lập từng thành phần và giữ cố định tuyệt đối mọi thành phần còn lại. Tuy nhiên, các thiết lập đã thực hiện vẫn cung cấp những đối chứng thành phần có liên quan trực tiếp đến BlockNet: VGG16 đóng vai trò mô hình cơ sở chỉ sử dụng đặc trưng học sâu; hai thiết lập có và không chọn lọc entropy cho phép xem xét ảnh hưởng của tiêu chí chọn lọc; kết quả trên 13 lớp tích chập phản ánh ảnh hưởng của vị trí trích xuất; và so sánh BlockNet với các backbone chỉ dùng đặc trưng học sâu phản ánh hiệu quả tổng hợp của quy trình chọn lọc và hợp nhất. Vì vậy,



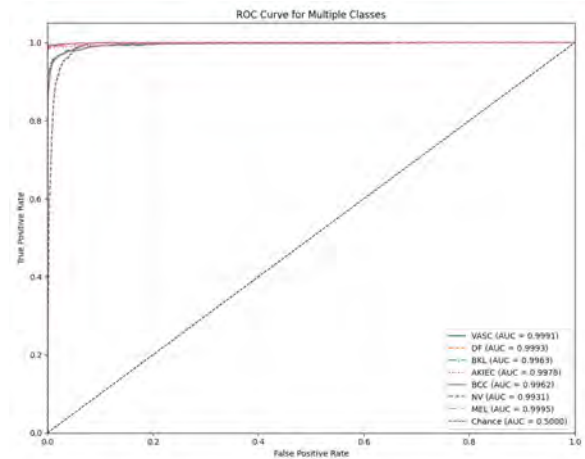
(a) Accuracy



(b) Loss



(c) Ma trận nhầm lẫn



(d) Đường cong ROC

Hình 2.19: Phân tích chi tiết cấu hình BlockNet cao nhất trên bộ dữ liệu HAM10000.

các mục tiếp theo phân tích vai trò của từng thành phần trong giới hạn của các đối chứng hiện có, không quy riêng mức cải thiện cho một thành phần nếu thiết kế thực nghiệm chưa tách độc lập thành phần đó.

2.7.1. Giải thích sự chênh lệch giữa hai nhóm kết quả trên HAM10000

Hai nhóm kết quả trên HAM10000 được trình bày với hai mục tiêu đánh giá khác nhau nên không thể xem là kết quả lặp lại của cùng một cấu hình. Nhóm thứ nhất, trong Bảng 2.8 và Bảng 2.9, là thí nghiệm chẩn đoán theo từng lớp tích chập, được sử dụng để quan sát ảnh hưởng của việc không chọn lọc và có chọn lọc bản đồ đặc trưng bằng entropy đối với biểu diễn đặc trưng học sâu. Trong nhóm này, cấu hình tốt nhất đạt Accuracy khoảng 69,07% khi không chọn lọc entropy và 70,00% khi có chọn lọc entropy.

Nhóm thứ hai trong Bảng 2.11 là kết quả của cấu hình BlockNet đầy đủ. Kết quả này được hình thành từ toàn bộ quy trình gồm chọn lọc bản đồ đặc trưng học sâu dựa trên entropy, hợp nhất với đặc trưng thủ công, cân bằng dữ liệu trong không gian đặc trưng trên tập huấn luyện và phân loại bằng mạng perceptron nhiều lớp (multilayer perceptron-MLP). Các chỉ số được tổng hợp qua kiểm định chéo phân tầng 5 folds (stratified five-fold cross-validation); cấu hình tốt nhất tại lớp block1_conv2 đạt Accuracy 96,02% $\pm$ 0,24%.

Do đó, chênh lệch lớn giữa hai nhóm chủ yếu phản ánh sự khác nhau về phạm vi cấu hình và mục đích thực nghiệm: nhóm thứ nhất khảo sát một cơ chế trên biểu diễn đặc trưng học sâu, còn nhóm thứ hai đánh giá hiệu quả tổng hợp của toàn bộ BlockNet. Mức chênh lệch này không được diễn giải là phần cải thiện riêng do entropy, đặc trưng thủ công hay cân bằng dữ liệu, bởi các thực nghiệm hiện có chưa tách độc lập hoàn toàn từng thành phần trong cùng một giao thức loại trừ.

2.7.2. Ảnh hưởng của chọn lọc bản đồ đặc trưng dựa trên entropy

Các kết quả trong Bảng 2.6, Bảng 2.7, Bảng 2.8 và Bảng 2.9 ghi nhận ảnh hưởng của cơ chế chọn lọc bản đồ đặc trưng dựa trên entropy đối với biểu diễn đặc trưng ở các block CNN.

Trên bộ dữ liệu VNCancer, Accuracy và F1-score tăng sau khi áp dụng chọn lọc entropy. Các lớp tích chập sâu như block4_conv3 và block5_conv3 có mức tăng lớn. Ví dụ, tại block4_conv3, Accuracy tăng từ 49.73% lên 86.85%, trong khi F1-score tăng từ 57.63% lên 86.04%. Kết quả này phù hợp với nhận định rằng các tầng CNN sâu có thể chứa nhiều bản đồ đặc trưng phản hồi yếu, dư thừa hoặc phân tán; việc sử dụng trực tiếp toàn bộ bản đồ đặc trưng có thể làm giảm chất lượng biểu diễn đầu vào.

Kết quả ghi nhận mức tăng của F1-score thường lớn hơn Accuracy, nhất là trên HAM10000. Kết quả này cho thấy chọn lọc entropy ảnh hưởng đến các lớp khó và các lớp thiểu số. Trong dữ liệu mất cân bằng, Accuracy có thể bị chi phối bởi các lớp phổ biến, trong khi Macro F1 phản ánh cân bằng hơn khả năng phân loại trên toàn bộ các lớp. Việc F1-score tăng cho thấy các bản đồ đặc trưng được giữ lại sau bước chọn lọc entropy có liên quan đến các đặc trưng phân biệt của các lớp ít mẫu hoặc có mức độ chồng lấn hình thái cao.

Ngoài ra, các kết quả ROC trong Hình 2.14 và Hình 2.15 ghi nhận thay đổi AUC sau khi áp dụng chọn lọc entropy, đối với các lớp như SCC, AKIEC hoặc VASC. Việc loại bỏ các bản đồ đặc trưng có entropy không phù hợp làm giảm nhiễu trong không gian biểu diễn.

Các kết quả này phù hợp với giả thuyết rằng không phải mọi bản đồ đặc trưng trung gian đều mang giá trị thông tin tương đương nhau đối với bài toán phân loại ảnh tổn thương da. Việc lựa chọn bản đồ đặc trưng dựa trên entropy giúp giảm thành phần dư thừa trong biểu diễn học sâu trước khi thực hiện hợp nhất đặc trưng đa nguồn.

2.7.3. Ảnh hưởng của các block CNN

Kết quả thực nghiệm ghi nhận sự thay đổi hiệu năng của BlockNet theo từng block CNN. Trên bộ dữ liệu VNCancer, các block nông như block1_conv1 và block1_conv2 cho kết quả cao nhất, trong khi các block sâu hơn có xu hướng giảm hiệu năng nếu không áp dụng chọn lọc entropy.

Các bản đồ đặc trưng ở các tầng nông chứa thông tin cục bộ liên quan đến bài toán phân loại tổn thương da, chẳng hạn màu sắc, biên tổn thương và kết cấu bề mặt. Đây là các đặc điểm có giá trị phân biệt đối với các tổn thương như BCC, MM hoặc SCC. Trong khi đó, các tầng quá sâu có xu hướng tập trung vào các biểu diễn ngữ nghĩa mức cao hơn và có thể làm mất một phần thông tin cục bộ đối với bài toán ảnh y tế.

Trên bộ dữ liệu HAM10000, các block trung gian như block1_conv2 và block2_conv1 cho kết quả cao hơn. Điều này phù hợp với bản chất của HAM10000 là bộ dữ liệu đa lớp phức tạp với mức độ chồng lấn thị giác lớn giữa nhiều lớp tổn thương. Trong trường hợp này, mô hình cần các biểu diễn giàu ngữ nghĩa hơn so với dữ liệu VNCancer để phân

biệt các lớp có đặc điểm gần giống nhau.

Các kết quả này hỗ trợ việc khai thác biểu diễn theo từng block CNN thay vì chỉ sử dụng tầng cuối của mạng trong bài toán phân loại ảnh tổn thương da. Đồng thời, mỗi mức biểu diễn của CNN có thể phù hợp với các đặc điểm dữ liệu khác nhau.

2.7.4. Hiệu quả của hợp nhất đặc trưng đa nguồn

Các kết quả trong Bảng 2.10 và Bảng 2.11 ghi nhận BlockNet có các chỉ số cao hơn so với các backbone CNN phổ biến như ResNet, DenseNet, EfficientNet và ViT.

Một nguyên nhân là mô hình không chỉ sử dụng đặc trưng học sâu mà còn kết hợp thêm đặc trưng thủ công phản ánh màu sắc, kết cấu và hình dạng tổn thương. Trong các bài toán ảnh tổn thương da, nhiều lớp có sự khác biệt tinh tế về phân bố màu hoặc cấu trúc biên tổn thương mà các bản đồ đặc trưng học sâu chưa chắc đã biểu diễn đầy đủ. Các handcrafted features giúp bổ sung thêm thông tin mang tính tri thức miền, trong các lớp khó hoặc các lớp có số lượng mẫu thấp.

Kết quả ghi nhận vai trò hỗ trợ của handcrafted features ngay cả khi sử dụng backbone CNN. Điều này phản ánh rằng các mô tả thủ công và đặc trưng học sâu có tính bổ trợ cho nhau thay vì hoàn toàn thay thế lẫn nhau. Trong khi deep features có khả năng học biểu diễn phức tạp từ dữ liệu, handcrafted features giúp duy trì các mô tả ổn định về hình thái và kết cấu của tổn thương da.

Ngoài ra, chọn lọc entropy trước bước hợp nhất là một thành phần của quy trình. So sánh thực nghiệm với cấu hình sử dụng toàn bộ bản đồ cho thấy biểu diễn sau sàng lọc đạt kết quả phân loại cao hơn trong thiết lập đang xét. Vì entropy kích hoạt không sử dụng nhân, kết quả này được diễn giải ở mức hiệu quả của tập bản đồ sau sàng lọc, không phải bằng chứng rằng mỗi bản đồ được giữ lại đều có khả năng phân biệt cao.

2.7.5. Ảnh hưởng của cân bằng trong không gian đặc trưng

Trong BlockNet, quá trình cân bằng dữ liệu được thực hiện trong không gian đặc trưng thay vì trên ảnh đầu vào. Đây là một thành phần của cấu hình BlockNet đầy đủ; các thực nghiệm hiện có chưa tách riêng duy nhất bước cân bằng trong khi giữ cố định toàn bộ các thành phần còn lại. Do đó, kết quả của cấu hình đầy đủ không được dùng để quy riêng mức cải thiện cho cơ chế cân bằng.

Trong các bài toán ảnh tổn thương da, tăng cường dữ liệu trực tiếp trên ảnh có thể tạo ra các biến đổi không phù hợp về màu sắc hoặc cấu trúc tổn thương, từ đó làm sai lệch đặc điểm bệnh học ban đầu của ảnh. Ngược lại, việc thực hiện cân bằng trên các vector đặc trưng đã được chọn lọc theo entropy giúp mở rộng biểu diễn của các lớp thiểu số trong không gian đặc trưng mà không làm thay đổi trực tiếp nội dung ảnh.

Các kết quả Sensitivity, F1-score, AUC theo lớp và ma trận nhầm lẫn trên hai bộ dữ liệu ghi nhận hiệu năng của toàn bộ quy trình đối với các lớp hiếm và các lớp khó phân loại. Đối với lớp melanoma, ký hiệu MM trên VNCancer và MEL trên HAM10000, các kết quả hiện có cũng thuộc cấu hình BlockNet đầy đủ, không phải hai thiết lập chỉ khác nhau ở việc có hoặc không cân bằng. Vì vậy, các kết quả này cho thấy khả năng phân loại lớp melanoma của toàn bộ mô hình, nhưng chưa đủ để kết luận mức thay đổi của riêng lớp này trước và sau bước cân bằng.

Việc chưa có đối chứng độc lập giữa class-weighted loss, focal loss, SMOTE và lấy mẫu tăng cường trong không gian đặc trưng, cũng như chưa tách riêng kết quả lớp melanoma trước và sau cân bằng, là giới hạn của thiết kế thực nghiệm hiện tại. Đánh giá

đầy đủ các cơ chế xử lý mất cân bằng trong cùng giao thức và trên cùng các fold là một hướng thực nghiệm tiếp theo.

Về thiết kế, cân bằng được đặt sau bước chọn lọc entropy để việc lấy mẫu lại chỉ thao tác trên các vector được xây dựng từ tập bản đồ đặc trưng đã chọn, thay vì lặp lại các biểu diễn trước chọn lọc. Đây là lý do tổ chức quy trình, không phải bằng chứng thực nghiệm về ưu thế tuyệt đối so với các chiến lược xử lý mất cân bằng khác.

2.7.6. Một số nhận xét tổng quát

Nhìn chung, các kết quả thực nghiệm cho thấy đóng góp của BlockNet không nằm ở việc sử dụng một backbone CNN mới, mà ở chiến lược tổ chức và tối ưu biểu diễn đặc trưng trước bước phân loại. Các thành phần chính gồm: (i) lựa chọn bản đồ đặc trưng dựa trên entropy, (ii) khai thác biểu diễn theo từng block CNN, (iii) hợp nhất deep features với handcrafted features, và (iv) cân bằng dữ liệu trong không gian đặc trưng, được tổ chức trong cùng một quy trình để tạo biểu diễn đầu vào cho bộ phân loại.

Kết quả này cho thấy vai trò của việc xử lý biểu diễn trước bước hợp nhất đối với bài toán phân loại ảnh tổn thương da đa nguồn, trong điều kiện dữ liệu mất cân bằng và có số lượng nhãn hạn chế.

Kết luận chương

Chương này đã trình bày mô hình BlockNet cho bài toán phân loại ảnh tổn thương da dựa trên sàng lọc bản đồ đặc trưng theo entropy kích hoạt và hợp nhất đặc trưng đa nguồn. Entropy được tính từ histogram giá trị kích hoạt để mô tả độ phân tán, không trực tiếp đo thông tin liên quan nhãn hay khả năng phân biệt lớp. Khoảng entropy theo từng block được sử dụng như quy tắc sàng lọc không giám sát trước khi xây dựng vector đặc trưng học sâu, còn giá trị của biểu diễn thu được được đánh giá bằng kết quả phân loại.

Các nội dung chính của chương bao gồm phát biểu bài toán, ý tưởng phương pháp đề xuất, kiến trúc tổng quan của BlockNet, cơ chế tính entropy và chọn lọc bản đồ đặc trưng, phương pháp hợp nhất đặc trưng đa nguồn, chiến lược cân bằng dữ liệu trong không gian đặc trưng, bộ phân loại và các thuật toán đề xuất. Các thành phần này tạo thành một quy trình thống nhất, trong đó chất lượng biểu diễn được xử lý và tối ưu trước khi đưa vào huấn luyện mô hình phân loại.

Kết quả thực nghiệm trên VNCancer và HAM10000 ghi nhận sự khác biệt giữa cấu hình dùng toàn bộ bản đồ và cấu hình áp dụng sàng lọc entropy. Mức tăng ở F1-score và AUC theo lớp cho thấy tập biểu diễn sau sàng lọc phù hợp hơn với nhiệm vụ phân loại trong giao thức thực nghiệm hiện tại. Kết quả không được dùng để quy entropy thành độ đo phân biệt của từng bản đồ riêng lẻ.

Bên cạnh đó, kết quả đánh giá hợp nhất đặc trưng đa nguồn ghi nhận sự khác biệt giữa mô hình kết hợp đặc trưng học sâu đã chọn lọc với đặc trưng thủ công và các backbone học sâu phổ biến. Các kết quả này phù hợp với vai trò bổ trợ giữa đặc trưng học sâu và đặc trưng thủ công trong bài toán phân loại tổn thương da. Trong khi đặc trưng học sâu có khả năng học biểu diễn phức tạp từ dữ liệu, các đặc trưng thủ công mô tả màu sắc, kết cấu và hình dạng tổn thương, đối với các lớp có mức độ chồng lấn hình thái cao.

Ngoài ra, chương trình bày việc thực hiện cân bằng dữ liệu trong không gian đặc trưng đối với dữ liệu mất cân bằng. Thay vì thực hiện tăng cường dữ liệu trực tiếp trên ảnh đầu vào, phương pháp thực hiện mở rộng biểu diễn trên các feature vector đã được

chọn lọc theo entropy. Cách xử lý này bổ sung mẫu trong không gian vector mà không làm thay đổi trực tiếp cấu trúc hình thái của ảnh tổn thương da.

Các kết quả của cấu hình đầy đủ chưa tách riêng ảnh hưởng của bước cân bằng và chưa cung cấp đối chứng riêng cho lớp melanoma trước và sau khi chỉ thay đổi thành phần này. Do đó, luận án giới hạn kết luận ở việc mô tả cơ chế và hiệu năng của toàn bộ quy trình, đồng thời xem đánh giá độc lập các chiến lược xử lý mất cân bằng là hướng nghiên cứu tiếp theo.

Nhìn chung, các kết quả trong chương này cho thấy vai trò của việc xử lý biểu diễn trước bước hợp nhất trong bài toán phân loại ảnh tổn thương da đa nguồn. Đóng góp chính của chương là đề xuất cơ chế chọn lọc đặc trưng có dẫn hướng bởi entropy kết hợp với hợp nhất đặc trưng đa nguồn để xây dựng biểu diễn đầu vào trong điều kiện dữ liệu đa dạng, mất cân bằng và phức tạp về mặt hình thái.

Tuy nhiên, phương pháp trong chương này vẫn chủ yếu thực hiện học trên từng mẫu dữ liệu một cách tương đối độc lập. Mô hình vẫn chưa khai thác tường minh quan hệ liên mẫu trong không gian dữ liệu. Đồng thời, dữ liệu chưa gán nhãn vẫn chưa được sử dụng trong quá trình huấn luyện, trong khi đây là nguồn dữ liệu lớn và có tiềm năng hỗ trợ nâng cao khả năng tổng quát hóa của mô hình.

Bên cạnh đó, cơ chế hợp nhất trong BlockNet vẫn mang tính tương đối tĩnh và chưa mô hình hóa trực tiếp sự phụ thuộc giữa các nguồn dữ liệu cũng như mối quan hệ giữa các mẫu. Điều này tạo động lực cho Chương 3, trong đó hướng nghiên cứu được mở rộng từ mức tối ưu hóa biểu diễn đặc trưng sang mức mô hình hóa quan hệ dữ liệu bằng học trên đồ thị và khai thác dữ liệu chưa gán nhãn thông qua cơ chế gán nhãn giả có kiểm soát độ tin cậy dựa trên entropy. Qua đó, luận án tiếp tục phát triển từ bài toán nâng cao chất lượng biểu diễn đầu vào sang bài toán học đa nguồn bán giám sát có khai thác quan hệ liên mẫu cho phân loại ảnh tổn thương da.

CHƯƠNG 3: MÔ HÌNH PHÂN LOẠI TỔN THƯƠNG DA DỰA TRÊN HỌC ĐA NGUỒN VÀ ĐỒ THỊ

Mở đầu

Chương 2 đã trình bày mô hình BlockNet nhằm nâng cao chất lượng biểu diễn ảnh tổn thương da thông qua trích xuất, đánh giá và chọn lọc bản đồ đặc trưng dựa trên entropy. Các kết quả thực nghiệm cho thấy việc lựa chọn bản đồ đặc trưng theo entropy giúp giảm thành phần dư thừa, hạn chế nhiễu trong không gian biểu diễn và cải thiện khả năng phân biệt của đặc trưng đầu vào. Đồng thời, việc hợp nhất đặc trưng học sâu với đặc trưng thủ công cho thấy hiệu quả trong bối cảnh dữ liệu tổn thương da có tính đa dạng và mất cân bằng cao. Tuy nhiên, mặc dù chất lượng biểu diễn của từng ảnh đã được cải thiện, mô hình trong Chương 2 vẫn chủ yếu thực hiện học trên từng mẫu dữ liệu theo cách tương đối độc lập. Nói cách khác, biểu diễn của mỗi ảnh chưa khai thác đầy đủ quan hệ giữa các mẫu trong không gian dữ liệu.

Trong thực tế, các mẫu tổn thương da thường không tồn tại hoàn toàn độc lập. Những ảnh có đặc điểm hình thái, màu sắc, kết cấu hoặc đặc trưng lâm sàng tương tự nhau có thể hỗ trợ lẫn nhau trong quá trình học. Điều này cần được xem xét đối với các lớp thiểu số hoặc các tổn thương khó phân biệt bằng thông tin thị giác đơn thuần. Vì vậy, sau khi giải quyết bài toán chất lượng biểu diễn ở mức đặc trưng trong Chương 2, Chương 3 chuyển trọng tâm từ “biểu diễn từng ảnh” sang “biểu diễn quan hệ giữa các mẫu”, thông qua mô hình hóa cấu trúc dữ liệu dưới dạng đồ thị và khai thác học trên đồ thị cho bài toán phân loại tổn thương da.

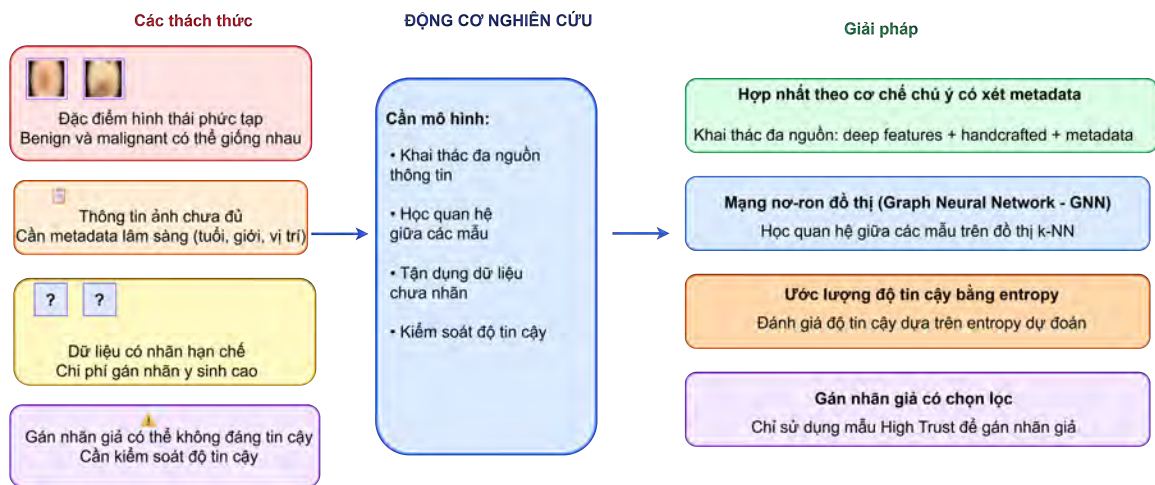
Bên cạnh đó, các nghiên cứu trước của Nghiên cứu sinh trong [CT2] đã khảo sát khả năng sử dụng nhiều kiến trúc nền CNN và mô hình ensemble cho dữ liệu tổn thương da. Kết quả cho thấy các mô hình học sâu khác nhau có thể khai thác các đặc trưng bổ sung cho nhau trong quá trình phân loại. Tuy nhiên, [CT2] vẫn xử lý từng mẫu dữ liệu theo cách tương đối độc lập và chưa khai thác được quan hệ tương đồng giữa các mẫu trong không gian biểu diễn.

Một thành phần khác trong bài toán phân loại tổn thương da là metadata lâm sàng. Trong nhiều trường hợp, hai ảnh tổn thương có biểu hiện thị giác gần giống nhau nhưng khác vị trí tổn thương, độ tuổi hoặc giới tính của người bệnh thì nguy cơ thuộc nhóm bệnh lý có thể khác nhau. Ví dụ, một tổn thương xuất hiện ở vùng da thường xuyên tiếp xúc ánh nắng trên bệnh nhân lớn tuổi có thể mang nguy cơ khác với tổn thương có hình thái tương tự ở nhóm bệnh nhân trẻ hơn hoặc ở vị trí ít phơi nhiễm ánh sáng. Do đó, metadata không thay thế thông tin ảnh, mà đóng vai trò bổ sung ngữ cảnh lâm sàng nhằm hỗ trợ điều chỉnh biểu diễn của ảnh trong quá trình phân loại. Đây là cơ sở để Chương 3 xây dựng biểu diễn đa nguồn, trong đó đặc trưng ảnh và metadata được khai thác đồng thời thay vì xử lý tách rời.

Trong thực tế, dữ liệu ảnh y tế có nhãn thường hạn chế do quá trình gán nhãn đòi hỏi chuyên gia và chi phí cao. Trong khi đó, dữ liệu chưa gán nhãn thường có số lượng lớn hơn và có tiềm năng hỗ trợ cải thiện hiệu quả học nếu được khai thác phù hợp. Tuy nhiên, học bán giám sát dựa trên gán nhãn giả cũng tiềm ẩn rủi ro. Nếu mô hình tạo ra các dự đoán sai nhưng vẫn có độ tự tin cao, các nhãn giả sai này có thể tiếp tục được đưa ngược vào quá trình huấn luyện và làm lan truyền sai số qua nhiều vòng lặp học. Điều

này đặc biệt nguy hiểm trong dữ liệu mất cân bằng, khi các lớp hiếm dễ bị lấn át bởi các lớp phổ biến. Vì vậy, việc sử dụng entropy trong Chương 3 không chỉ được xem như một thủ thuật kỹ thuật, mà đóng vai trò như một cơ chế kiểm soát rủi ro nhằm đánh giá độ tin cậy của dự đoán trước khi lựa chọn dữ liệu chưa nhãn cho quá trình tái huấn luyện.

Trong bối cảnh đó, học trên đồ thị trở thành một hướng tiếp cận phù hợp để khai thác đồng thời biểu diễn đa nguồn và quan hệ liên mẫu. Trong mô hình học trên đồ thị, mỗi mẫu dữ liệu được xem là một nút, trong khi quan hệ tương đồng giữa các mẫu được mô hình hóa thông qua các cạnh của đồ thị. Việc lan truyền thông tin giữa các nút lân cận cho phép mô hình tận dụng thêm ngữ cảnh liên mẫu thay vì chỉ dựa trên biểu diễn của từng ảnh riêng lẻ. Đồng thời, như minh họa trong Hình 3.1, bài toán còn đối mặt với nhiều thách thức như dữ liệu có nhãn hạn chế, sự cần thiết của metadata lâm sàng và nguy cơ lan truyền sai số khi sử dụng gán nhãn giả.



Hình 3.1: Bối cảnh và động cơ nghiên cứu.

Xuất phát từ các vấn đề trên, chương này trình bày đóng góp thứ hai của luận án, với nội dung chính là đề xuất một khung học bán giám sát đa nguồn trên đồ thị cho phân loại ảnh tổn thương da. So với công trình [CT4], chương này trình bày bổ sung phát biểu bài toán, quy trình thuật toán, thiết lập kiểm soát rò rỉ dữ liệu, phân tích kết quả theo từng chỉ số và thảo luận phạm vi áp dụng của mô hình. Chương 3 kế tiếp Chương 2 về hướng nghiên cứu kết hợp thông tin thị giác và khai thác entropy, nhưng không tái sử dụng vector BlockNet hoặc bước chọn lọc bản đồ đặc trưng theo entropy của Chương 2. Theo đúng thiết lập thực nghiệm của [CT4], đặc trưng học sâu trong Chương 3 là các vector đầu ra sau pooling của VGG19, DenseNet121 và MobileNet; đặc trưng thủ công gồm histogram RGB, histogram HSV và đặc trưng fractal. Các nhóm đặc trưng này được lựa chọn bằng phân tích Chi-square trong chính Chương 3, sau đó kết hợp với metadata lâm sàng để xây dựng biểu diễn nút đa nguồn. Từ các biểu diễn này, đồ thị tương đồng giữa các mẫu được xây dựng và mô hình GNN được sử dụng để học quan hệ giữa các mẫu. Sau giai đoạn huấn luyện có giám sát, entropy của phân phối xác suất dự đoán được sử dụng để đánh giá độ tin cậy, từ đó chỉ lựa chọn các mẫu chưa gán nhãn có độ tin cậy cao để gán nhãn giả và huấn luyện lại mô hình.

Khác với Chương 2, nơi entropy được sử dụng để đánh giá và chọn lọc bản đồ đặc trưng, trong Chương 3 entropy được sử dụng ở mức dự đoán của mô hình nhằm ước lượng độ bất định và phân tầng độ tin cậy của các mẫu. Điều này cho thấy entropy được

khai thác ở hai cấp độ khác nhau trong luận án: cấp độ biểu diễn đặc trưng ở Chương 2 và cấp độ độ tin cậy dự đoán ở Chương 3.

Trên cơ sở các vấn đề đã phân tích, chương này đề xuất một khung học bán giám sát đa nguồn trên đồ thị cho phân loại ảnh tổn thương da, trong đó đặc trưng ảnh, đặc trưng thủ công và metadata được hợp nhất thành biểu diễn nút; quan hệ giữa các mẫu được mô hình hóa bằng đồ thị; và entropy dự đoán được sử dụng để chọn lọc các mẫu chưa gán nhãn có độ tin cậy cao cho quá trình gán nhãn giả.

Nội dung còn lại của chương được tổ chức như sau. Mục 3.1 phát biểu bài toán phân loại ảnh tổn thương da trong bối cảnh học đa nguồn và bán giám sát. Mục 3.2 trình bày ý tưởng tổng quát của phương pháp đề xuất. Mục 3.3 mô tả chi tiết các thành phần của mô hình, bao gồm trích xuất đặc trưng đa nguồn, hợp nhất đặc trưng có xét metadata, xây dựng đồ thị tương đồng, huấn luyện GNN có giám sát, đánh giá độ tin cậy bằng entropy và gán nhãn giả có chọn lọc. Mục 3.4 trình bày thuật toán và quy trình thực hiện. Mục 3.5 mô tả thiết lập thực nghiệm trên bộ dữ liệu ISIC 2020. Mục 3.6 phân tích kết quả thực nghiệm và đánh giá hiệu quả của từng thành phần trong mô hình. Cuối cùng, Mục 3.7 tổng kết các kết quả chính của chương.

3.1. Phát biểu bài toán

Bài toán được xem xét trong chương này là phân loại ảnh tổn thương da trong bối cảnh dữ liệu đa nguồn và học bán giám sát. Thay vì xem dữ liệu là một thành phần riêng của bài toán, luận án phát biểu bài toán thông qua ba yếu tố chính: không gian đầu vào đa nguồn, không gian nhãn và mục tiêu học hàm phân loại trên đồ thị. Trong thiết lập này, mỗi mẫu có thể được mô tả bởi ảnh tổn thương da, đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng tương ứng.

Khác với bài toán đa lớp ở Chương 2, bài toán trong Chương 3 được thiết lập dưới dạng phân loại nhị phân lành tính/ác tính trên bộ dữ liệu ISIC 2020. Tập nhãn được ký hiệu là $\mathcal{Y} = \{0, 1\}$, trong đó nhãn 0 tương ứng với tổn thương lành tính và nhãn 1 tương ứng với tổn thương ác tính.

Với mẫu thứ i , biểu diễn đầu vào đa nguồn được ký hiệu theo Công thức 3.1:

$$\mathbf{x}_i = \left(I_i, \mathbf{x}_i^{deep}, \mathbf{x}_i^{hand}, \mathbf{x}_i^{meta} \right), \quad (3.1)$$

trong đó I_i là ảnh tổn thương da; $\mathbf{x}_i^{deep}$ là vector đặc trưng học sâu được trích xuất từ ảnh; $\mathbf{x}_i^{hand}$ là vector đặc trưng thủ công mô tả các thuộc tính thị giác như màu sắc, kết cấu hoặc hình dạng; và $\mathbf{x}_i^{meta}$ là vector metadata lâm sàng sau khi được chuẩn hóa và mã hóa. Công thức 3.1 chỉ mô tả cấu trúc các nguồn thông tin đầu vào trước bước hợp nhất, chưa phải là vector biểu diễn cuối cùng dùng làm nút của đồ thị.

Tập dữ liệu có nhãn được xác định theo Công thức 3.2:

$$D_L = \left\{ (\mathbf{x}_i, y_i) \mid \mathbf{x}_i = \left(I_i, \mathbf{x}_i^{deep}, \mathbf{x}_i^{hand}, \mathbf{x}_i^{meta} \right), y_i \in \mathcal{Y}, i = 1, 2, \dots, N_L \right\}, \quad (3.2)$$

trong đó D_L là tập dữ liệu có nhãn; $\mathbf{x}_i$ là biểu diễn đầu vào đa nguồn của mẫu thứ i ; y_i là nhãn lớp tương ứng; và N_L là số lượng mẫu có nhãn.

Tập dữ liệu chưa gán nhãn được xác định theo Công thức 3.3:

$$D_U = \left\{ \mathbf{x}_j \mid \mathbf{x}_j = \left(I_j, \mathbf{x}_j^{deep}, \mathbf{x}_j^{hand}, \mathbf{x}_j^{meta} \right), j = 1, 2, \dots, N_U \right\}, \quad (3.3)$$

trong đó D_U là tập dữ liệu chưa gán nhãn và N_U là số lượng mẫu chưa gán nhãn. Điểm khác biệt giữa D_L và D_U là các mẫu trong D_L có nhãn y_i , trong khi các mẫu trong D_U chưa có nhãn và chỉ được khai thác thông qua cơ chế gán nhãn giả có kiểm soát độ tin cậy.

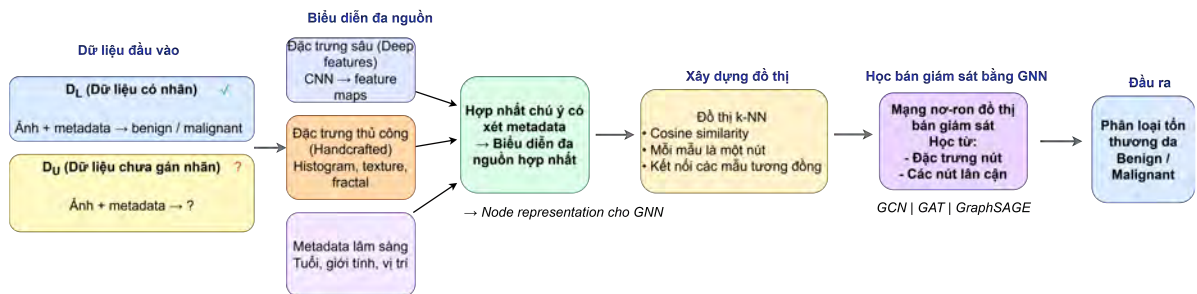
Quá trình hợp nhất ba nguồn thông tin được thực hiện theo cấu trúc phân cấp của [CT4]. Trước hết, đặc trưng học sâu và đặc trưng thủ công được ghép nối để tạo biểu diễn ảnh $\mathbf{x}_i^{img}$. Sau đó, biểu diễn ảnh được hợp nhất với metadata lâm sàng bằng hàm hợp nhất có xét metadata $\Phi(\cdot)$. Hai bước được xác định theo Công thức 3.4:

$$\begin{aligned} \mathbf{x}_i^{img} &= \left[\mathbf{x}_i^{deep}; \mathbf{x}_i^{hand} \right], \\ \mathbf{h}_i &= \Phi \left(\mathbf{x}_i^{img}, \mathbf{x}_i^{meta} \right). \end{aligned} \quad (3.4)$$

trong đó $\mathbf{x}_i^{img}$ là biểu diễn ảnh tổng hợp, $\mathbf{x}_i^{meta}$ là metadata lâm sàng đã được mã hóa và $\mathbf{h}_i$ là vector đặc trưng hợp nhất dùng làm biểu diễn nút trong mô hình học trên đồ thị. Công thức trên liệt kê đầy đủ ba nguồn thông tin được mô hình khai thác, đồng thời làm rõ rằng cơ chế chú ý hoạt động giữa hai đường biểu diễn là ảnh và metadata; đặc trưng học sâu và đặc trưng thủ công không phải hai đầu vào chú ý độc lập.

Trong thiết lập của Chương 3 và [CT4], đặc trưng học sâu $\mathbf{x}_i^{deep}$ được tạo bằng cách ghép nối ba vector đầu ra sau pooling của VGG19, DenseNet121 và MobileNet, có số chiều lần lượt là 512, 1024 và 1024. Đây là các biểu diễn CNN tiêu chuẩn được lựa chọn bằng phân tích Chi-square trong Chương 3; mô hình không đưa vector BlockNet và không thực hiện chọn lọc bản đồ đặc trưng theo entropy của Chương 2. Đặc trưng thủ công $\mathbf{x}_i^{hand}$ gồm đúng ba nhóm histogram RGB, histogram HSV và đặc trưng fractal. Hai vector $\mathbf{x}_i^{deep}$ và $\mathbf{x}_i^{hand}$ được ghép nối thành $\mathbf{x}_i^{img}$; biểu diễn ảnh này tiếp tục được hợp nhất với metadata lâm sàng để hình thành vector $\mathbf{h}_i$.

Hình 3.2 minh họa quá trình chuyển từ dữ liệu đa nguồn sang biểu diễn nút trên đồ thị. Trong đó, mỗi mẫu dữ liệu tương ứng với một nút và quan hệ tương đồng giữa các mẫu được biểu diễn bằng các cạnh.



Hình 3.2: Sơ đồ minh họa quá trình biểu diễn dữ liệu đa nguồn, xây dựng đồ thị và học bán giám sát cho phân loại ảnh tổn thương da.

Đồ thị dữ liệu được ký hiệu là $G = (V, E)$, trong đó V là tập các nút và E là tập các cạnh. Đồ thị được xác định theo Công thức 3.5:

$$G = (V, E), \quad (3.5)$$

mỗi nút $v_i \in V$ tương ứng với một mẫu dữ liệu và được gán biểu diễn nút $\mathbf{h}_i$. Các cạnh trong E được xây dựng dựa trên độ tương đồng giữa các biểu diễn nút. Trong chương này, độ tương đồng giữa hai mẫu i và j được tính bằng độ tương đồng cosine, được xác định theo Công thức 3.6:

$$s_{ij} = \frac{\mathbf{h}_i^T \mathbf{h}_j}{\|\mathbf{h}_i\|_2 \|\mathbf{h}_j\|_2}, \quad (3.6)$$

trong đó s_{ij} biểu diễn mức độ tương đồng giữa hai mẫu tương ứng với hai nút v_i và v_j . Dựa trên ma trận độ tương đồng, đồ thị được xây dựng theo chiến lược k -nearest neighbors với $k = 5$. Giá trị $k = 5$ được lựa chọn nhằm cân bằng giữa việc bảo toàn quan hệ cục bộ giữa các mẫu gần nhau trong không gian biểu diễn và tránh tạo ra đồ thị quá dày làm suy giảm khả năng phân biệt của cấu trúc đồ thị. Trong nghiên cứu này, giá trị k được cố định như một tham số trong cấu hình thực nghiệm.

Theo chiến lược này, mỗi nút được kết nối với năm nút có biểu diễn gần nhất trong không gian đặc trưng hợp nhất. Quá trình xây dựng đồ thị không sử dụng thông tin nhãn, nhờ đó hạn chế rò rỉ dữ liệu giữa các tập huấn luyện, kiểm định và kiểm tra.

Trên đồ thị G , mô hình mạng nơ-ron đồ thị được sử dụng để học biểu diễn và dự đoán nhãn cho từng nút. Với tham số mô hình θ , phân phối xác suất dự đoán của mẫu thứ i được xác định theo Công thức 3.7:

$$\hat{\mathbf{y}}_i = f_\theta(G, \mathbf{h}_i), \quad (3.7)$$

trong đó $\hat{\mathbf{y}}_i = [\hat{y}_{i,0}, \hat{y}_{i,1}]$ là vector xác suất dự đoán tương ứng với hai lớp lành tính và ác tính. Thành phần $\hat{y}_{i,c}$ biểu thị xác suất dự đoán mẫu thứ i thuộc lớp c , với $c = 0$ là lớp lành tính và $c = 1$ là lớp ác tính. Nhãn dự đoán của mẫu thứ i được xác định theo Công thức 3.8:

$$\hat{y}_i = \arg \max_{c \in \{0,1\}} \hat{y}_{i,c}. \quad (3.8)$$

Đối với các mẫu thuộc tập chưa gán nhãn D_U , phân phối xác suất dự đoán của mô hình được sử dụng để đánh giá độ bất định thông qua entropy. Entropy dự đoán của mẫu thứ i được xác định theo Công thức 3.9:

$$H(\hat{\mathbf{y}}_i) = - \sum_{c=0}^1 \hat{y}_{i,c} \log(\hat{y}_{i,c}). \quad (3.9)$$

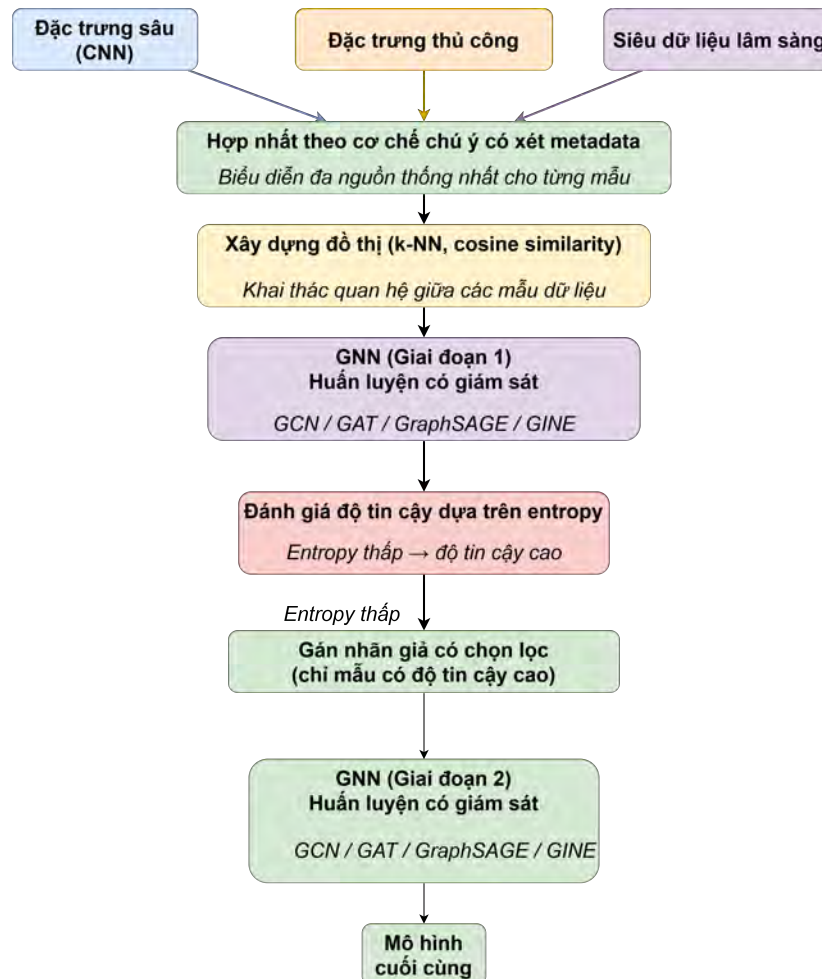
Trong Công thức 3.9, c là chỉ số lớp trong bài toán phân loại nhị phân, với $c = 0$ tương ứng với lớp lành tính và $c = 1$ tương ứng với lớp ác tính; $\hat{y}_{i,c}$ là xác suất dự đoán mẫu thứ i thuộc lớp c ; và $\log$ là logarit tự nhiên. Entropy thấp cho thấy phân phối xác suất dự đoán tập trung hơn vào một lớp, trong khi entropy cao phản ánh mức độ bất định

lớn hơn. Vì vậy, entropy được sử dụng để phân tầng độ tin cậy của các mẫu chưa gán nhãn trước khi đưa vào quá trình gán nhãn giả.

Trên cơ sở đó, bài toán của chương này được phát biểu như sau: với tập dữ liệu có nhãn D_L được xác định theo Công thức 3.2 và tập dữ liệu chưa gán nhãn D_U được xác định theo Công thức 3.3, cần xây dựng một mô hình học bán giám sát đa nguồn trên đồ thị có khả năng học biểu diễn nút $\mathbf{h}_i$ từ các nguồn đặc trưng khác nhau, khai thác quan hệ tương đồng giữa các mẫu trên đồ thị G , dự đoán nhãn tổn thương da theo Công thức 3.7, đồng thời sử dụng entropy dự đoán theo Công thức 3.9 để lựa chọn các mẫu chưa gán nhãn có độ tin cậy cao cho quá trình gán nhãn giả và huấn luyện lại mô hình.

3.2. Cơ sở đề xuất phương pháp

Như đã trình bày trong Mục 3.1, bài toán phân loại ảnh tổn thương da trong bối cảnh dữ liệu đa nguồn và học bán giám sát đặt ra các yêu cầu về biểu diễn dữ liệu, khai thác quan hệ giữa các mẫu và kiểm soát độ tin cậy khi sử dụng dữ liệu chưa gán nhãn. Trên cơ sở đó, chương này đề xuất một khung học bán giám sát đa nguồn trên đồ thị cho phân loại ảnh tổn thương da. Cơ sở tổng quát của phương pháp được minh họa trong Hình 3.3.



Hình 3.3: Cơ sở tổng quát của phương pháp đề xuất.

Thứ nhất, mỗi mẫu tổn thương da cần được biểu diễn bởi nhiều nguồn thông tin thay vì chỉ sử dụng ảnh đầu vào. Đặc điểm hình thái của tổn thương da có thể chồng lấn giữa các lớp, do đó một nguồn đặc trưng đơn lẻ chưa đủ phản ánh đầy đủ thông tin

phục vụ phân loại. Phương pháp đề xuất khai thác ba nguồn thông tin gồm đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng. Theo [CT4], đặc trưng học sâu là các vector đầu ra sau pooling của VGG19, DenseNet121 và MobileNet; đặc trưng thủ công gồm histogram RGB, histogram HSV và đặc trưng fractal; metadata lâm sàng cung cấp ngữ cảnh như tuổi, giới tính và vị trí tổn thương. Cấu hình này là thiết lập riêng của Chương 3, không phải đầu ra của BlockNet và không sử dụng chọn lọc bản đồ đặc trưng theo entropy của Chương 2.

Thứ hai, phương pháp mô hình hóa quan hệ giữa các mẫu dữ liệu dưới dạng đồ thị. Cách tiếp cận này xuất phát từ thực tế nhiều tổn thương da có biểu hiện thị giác gần nhau hoặc xuất hiện trong bối cảnh lâm sàng tương tự. Trước khi xây dựng đồ thị, các nguồn thông tin được hợp nhất để tạo biểu diễn đa nguồn thống nhất cho từng mẫu. Phương pháp sử dụng cơ chế hợp nhất theo chú ý có xét metadata lâm sàng, trong đó metadata được xem là nguồn thông tin bổ trợ giúp điều chỉnh mức độ đóng góp của các nguồn đặc trưng trong quá trình hợp nhất.

Sau khi xây dựng biểu diễn đa nguồn hợp nhất, mỗi mẫu dữ liệu được xem là một nút trên đồ thị, trong khi các cạnh được thiết lập dựa trên mức độ tương đồng giữa các biểu diễn nút. Trong nghiên cứu này, đồ thị được xây dựng theo chiến lược k -nearest neighbors với $k = 5$, trong đó độ tương đồng giữa các mẫu được tính bằng cosine similarity. Giá trị $k = 5$ được lựa chọn nhằm cân bằng giữa việc bảo toàn quan hệ cục bộ giữa các mẫu gần nhau trong không gian đặc trưng và tránh tạo ra đồ thị quá dày làm suy giảm khả năng phân biệt của cấu trúc đồ thị. Cách xây dựng đồ thị này cho phép mô hình kết nối các tổn thương da có đặc điểm tương tự nhau trong không gian biểu diễn hợp nhất, từ đó khai thác được thêm thông tin quan hệ giữa các mẫu dữ liệu.

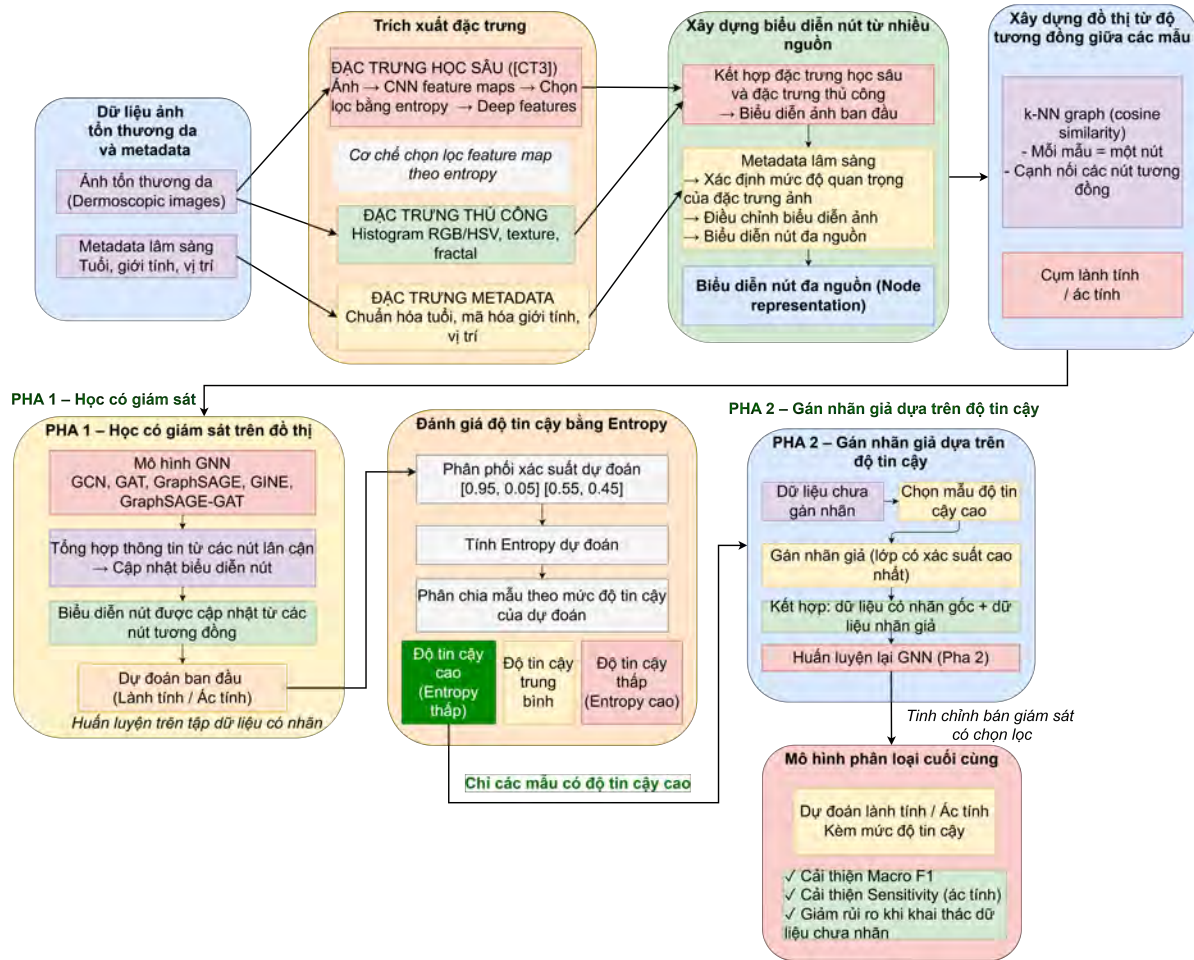
Trên đồ thị được xây dựng, mô hình mạng nơ-ron đồ thị được sử dụng để học biểu diễn nút thông qua cơ chế lan truyền thông tin giữa các nút lân cận. Thay vì chỉ học từ đặc trưng của chính mẫu dữ liệu, mỗi nút còn tiếp nhận thông tin từ các nút tương đồng trên đồ thị. Nhờ đó, mô hình có khả năng khai thác đồng thời thông tin cục bộ của từng mẫu và thông tin quan hệ giữa các mẫu trong tập dữ liệu. Trong nghiên cứu này, nhiều biến thể mạng nơ-ron đồ thị khác nhau được khảo sát, bao gồm GCN, GAT, GraphSAGE và GINE.

Thứ ba, dữ liệu chưa gán nhãn chỉ được khai thác khi dự đoán của mô hình có mức độ tin cậy đủ cao. Sau giai đoạn huấn luyện có giám sát ban đầu, mô hình được sử dụng để dự đoán trên tập dữ liệu chưa gán nhãn. Thay vì sử dụng toàn bộ dự đoán để tạo nhãn giả, phương pháp sử dụng entropy của phân phối xác suất dự đoán để ước lượng độ bất định. Các mẫu có entropy thấp được lựa chọn cho quá trình gán nhãn giả, còn các mẫu có entropy cao bị loại bỏ nhằm giảm nguy cơ lan truyền nhãn sai trong quá trình huấn luyện lại.

Khác với vai trò của entropy trong Chương 2, nơi entropy được sử dụng để đánh giá mức thông tin của bản đồ đặc trưng ở cấp độ biểu diễn đặc trưng, trong chương này entropy được sử dụng ở cấp độ dự đoán nhằm ước lượng mức độ bất định của phân phối xác suất đầu ra và phân tầng độ tin cậy của các mẫu chưa gán nhãn. Điều này cho thấy entropy được khai thác ở hai cấp độ khác nhau trong luận án: cấp độ chất lượng biểu diễn ở Chương 2 và cấp độ độ tin cậy dự đoán ở Chương 3.

Từ các cơ sở trên, mô hình tổng thể của phương pháp được xây dựng như minh họa trong Hình 3.4. Mô hình bao gồm hai giai đoạn chính. Giai đoạn thứ nhất thực hiện học có giám sát trên đồ thị bằng các mô hình mạng nơ-ron đồ thị nhằm xây dựng mô hình

phân loại ban đầu. Sau đó, entropy dự đoán được sử dụng để đánh giá mức độ bất định của các mẫu chưa gán nhãn và lựa chọn các mẫu có độ tin cậy cao cho quá trình gán nhãn giả. Trong giai đoạn thứ hai, dữ liệu gán nhãn giả được kết hợp với tập dữ liệu có nhãn ban đầu để huấn luyện lại mô hình, từ đó cải thiện khả năng tổng quát hóa và nâng cao hiệu quả phân loại.



Hình 3.4: Tổng quan mô hình đề xuất.

Trên cơ sở đó, phương pháp đề xuất hướng tới xây dựng một mô hình học bán giám sát đa nguồn trên đồ thị có khả năng khai thác đồng thời nhiều nguồn thông tin khác nhau, học quan hệ giữa các mẫu dữ liệu và tận dụng dữ liệu chưa gán nhãn theo hướng có kiểm soát độ tin cậy nhằm nâng cao hiệu quả phân loại tổn thương da lành tính và ác tính.

3.3. Phương pháp đề xuất

Từ phát biểu bài toán ở Mục 3.1 và cơ sở đề xuất phương pháp ở Mục 3.2, phần này trình bày chi tiết phương pháp học bán giám sát đa nguồn trên đồ thị cho phân loại ảnh tổn thương da. Phương pháp đề xuất được tổ chức thành sáu bước chính gồm: (1) trích xuất đặc trưng đa nguồn; (2) hợp nhất đặc trưng có xét metadata lâm sàng; (3) xây dựng đồ thị tương đồng giữa các mẫu; (4) huấn luyện mô hình mạng nơ-ron đồ thị có giám sát; (5) ước lượng entropy dự đoán để đánh giá độ tin cậy; và (6) gán nhãn giả có chọn lọc cho dữ liệu chưa gán nhãn.

Trong phương pháp đề xuất, mỗi mẫu dữ liệu được biểu diễn đồng thời bởi ba nguồn

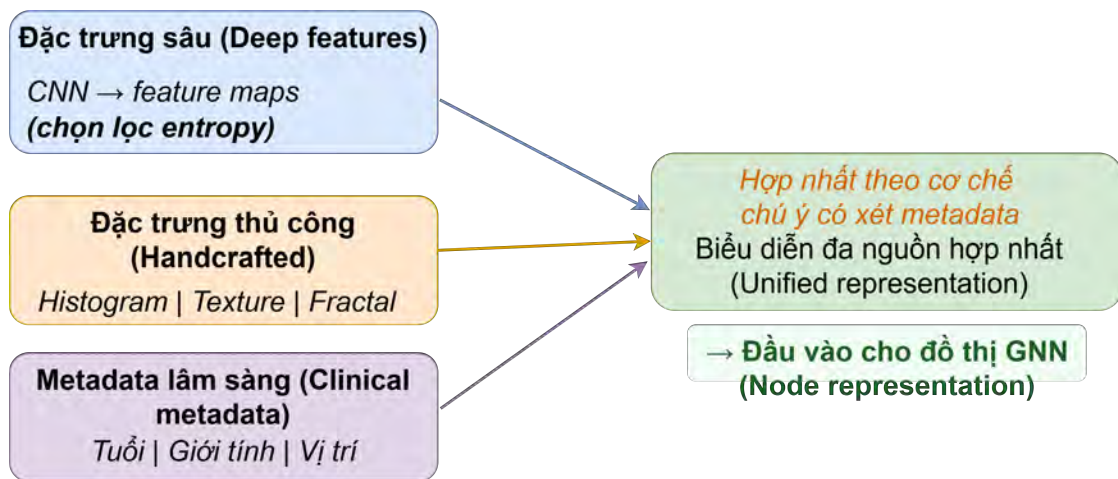
thông tin gồm đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng. Ba nguồn này được hợp nhất theo hai tầng: đặc trưng học sâu và đặc trưng thủ công được ghép nối thành biểu diễn ảnh $\mathbf{x}_i^{img}$; sau đó, biểu diễn ảnh được hợp nhất với metadata lâm sàng bằng cơ chế chú ý có xét metadata để tạo biểu diễn nút $\mathbf{h}_i$ như xác định trong Công thức 3.4. Từ các biểu diễn nút, đồ thị tương đồng được xây dựng nhằm mô hình hóa quan hệ giữa các mẫu dữ liệu trong không gian biểu diễn hợp nhất. Sau đó, mô hình mạng nơ-ron đồ thị được sử dụng để học biểu diễn nút và thực hiện phân loại tổn thương da.

Đối với dữ liệu chưa gán nhãn, entropy của phân phối xác suất dự đoán được sử dụng để đánh giá mức độ tin cậy của từng mẫu trước khi thực hiện gán nhãn giả. Chỉ các mẫu có entropy thấp mới được lựa chọn để bổ sung vào quá trình tái huấn luyện nhằm hạn chế nguy cơ lan truyền nhãn sai trong học bán giám sát.

Các thành phần của phương pháp đề xuất được trình bày chi tiết trong các mục tiếp theo, bao gồm trích xuất đặc trưng đa nguồn, hợp nhất đặc trưng có xét metadata lâm sàng, xây dựng đồ thị tương đồng, huấn luyện mạng nơ-ron đồ thị có giám sát, ước lượng entropy dự đoán và cơ chế gán nhãn giả có chọn lọc.

3.3.1. Trích xuất đặc trưng đa nguồn

Bước đầu tiên của phương pháp là xây dựng biểu diễn đặc trưng ban đầu cho từng mẫu tổn thương da từ nhiều nguồn thông tin khác nhau. Khác với các phương pháp chỉ sử dụng đặc trưng ảnh đơn thuần, nghiên cứu này khai thác đồng thời ba nguồn đặc trưng chính gồm đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng. Các nguồn đặc trưng này được trích xuất độc lập trước khi được đưa vào bước hợp nhất đặc trưng ở giai đoạn tiếp theo. Tổng quan các nguồn đặc trưng được sử dụng trong phương pháp được minh họa trong Hình 3.5.



Hình 3.5: Các nguồn đặc trưng đa nguồn và biểu diễn hợp nhất được sử dụng trong phương pháp đề xuất.

Trong ba nguồn thông tin được sử dụng, đặc trưng học sâu đóng vai trò là nguồn biểu diễn thị giác chính của mô hình. Theo cấu hình thực nghiệm trong [CT4], ba mạng CNN nền gồm VGG19, DenseNet121 và MobileNet được dùng làm bộ trích xuất đặc trưng: đầu phân loại của từng mạng được loại bỏ và vector đầu ra sau pooling được giữ lại. Ba vector này có số chiều lần lượt là 512, 1024 và 1024, được chuẩn hóa theo cùng quy trình trước khi ghép nối. Chương 3 không chọn các bản đồ đặc trưng trung gian bằng entropy và không sử dụng vector BlockNet của Chương 2.

Vector đặc trưng học sâu của mẫu i được xác định theo Công thức 3.10.

$$\mathbf{x}_i^{deep} = [\mathbf{x}_i^{VGG19}; \mathbf{x}_i^{DenseNet121}; \mathbf{x}_i^{MobileNet}] \in \mathbb{R}^{2560}, \quad (3.10)$$

trong đó $\mathbf{x}_i^{VGG19} \in \mathbb{R}^{512}$, $\mathbf{x}_i^{DenseNet121} \in \mathbb{R}^{1024}$ và $\mathbf{x}_i^{MobileNet} \in \mathbb{R}^{1024}$ là các vector đầu ra sau pooling của ba kiến trúc CNN. Vì vậy, $d_{deep} = 512 + 1024 + 1024 = 2560$. Việc lựa chọn ba loại đặc trưng này được thực hiện ở cấp nhóm đặc trưng bằng phân tích Chi-square tại Mục 3.5.3, khác với chọn lọc từng bản đồ đặc trưng theo entropy ở Chương 2.

Bên cạnh đặc trưng học sâu, đặc trưng thủ công được sử dụng nhằm bổ sung các mô tả có tính trực quan và dễ diễn giải hơn đối với tổn thương da. Theo [CT4] và cấu hình thực nghiệm của Chương 3, nhóm này chỉ gồm histogram màu RGB, histogram màu HSV và đặc trưng fractal. GLCM và HOG không được sử dụng trong Chương 3.

Vector đặc trưng thủ công của mẫu i được xác định theo Công thức 3.11.

$$\mathbf{x}_i^{hand} = [\mathbf{x}_i^{RGB}; \mathbf{x}_i^{HSV}; \mathbf{x}_i^{fractal}] \in \mathbb{R}^{1600}, \quad (3.11)$$

trong đó $\mathbf{x}_i^{RGB} \in \mathbb{R}^{768}$, $\mathbf{x}_i^{HSV} \in \mathbb{R}^{768}$ và $\mathbf{x}_i^{fractal} \in \mathbb{R}^{64}$. Do đó, $d_{hand} = 768 + 768 + 64 = 1600$, thống nhất với các độ dài vector được báo cáo trong Bảng 3.5.

Ngoài thông tin ảnh, metadata lâm sàng cũng được sử dụng nhằm bổ sung ngữ cảnh cho quá trình phân loại. Trong nghiên cứu này, metadata bao gồm tuổi, giới tính và vị trí tổn thương. Các thuộc tính phân loại như giới tính và vị trí tổn thương được mã hóa thành dạng số, trong khi tuổi được chuẩn hóa trước khi đưa vào mô hình nhằm bảo đảm tính nhất quán giữa các nguồn đặc trưng.

Vector metadata lâm sàng của mẫu i được xác định theo Công thức 3.12.

$$\mathbf{x}_i^{meta} = \Psi_{meta}(m_i), \quad (3.12)$$

trong đó $\Psi_{meta}(\cdot)$ là hàm tiền xử lý và mã hóa metadata lâm sàng. Đối với các trường metadata bị thiếu, dữ liệu được xử lý theo cùng một quy tắc thống nhất trong toàn bộ thực nghiệm nhằm bảo đảm tính nhất quán của biểu diễn đầu vào. Chi tiết về tỷ lệ thiếu dữ liệu và quy trình xử lý được trình bày trong phần thiết lập thực nghiệm của chương này.

Việc đưa metadata lâm sàng vào biểu diễn dữ liệu cho phép mô hình khai thác thêm các yếu tố ngữ cảnh không thể quan sát trực tiếp từ ảnh nhưng có thể liên quan đến nguy cơ tổn thương da. Ví dụ, cùng một biểu hiện thị giác nhưng xuất hiện ở các nhóm tuổi hoặc vị trí tổn thương khác nhau có thể mang mức độ nguy cơ khác nhau trong thực tế lâm sàng.

Sau khi hoàn thành quá trình trích xuất, mỗi mẫu dữ liệu được biểu diễn bởi ba vector đặc trưng gồm đặc trưng học sâu $\mathbf{x}_i^{deep}$, đặc trưng thủ công $\mathbf{x}_i^{hand}$ và metadata lâm sàng $\mathbf{x}_i^{meta}$. Các vector đặc trưng này được xây dựng độc lập nhằm giữ được đặc thù của từng nguồn thông tin trước khi thực hiện hợp nhất để tạo thành biểu diễn đa nguồn cho mô hình mạng nơ-ron đồ thị.

3.3.2. Hợp nhất đặc trưng đa nguồn có xét metadata

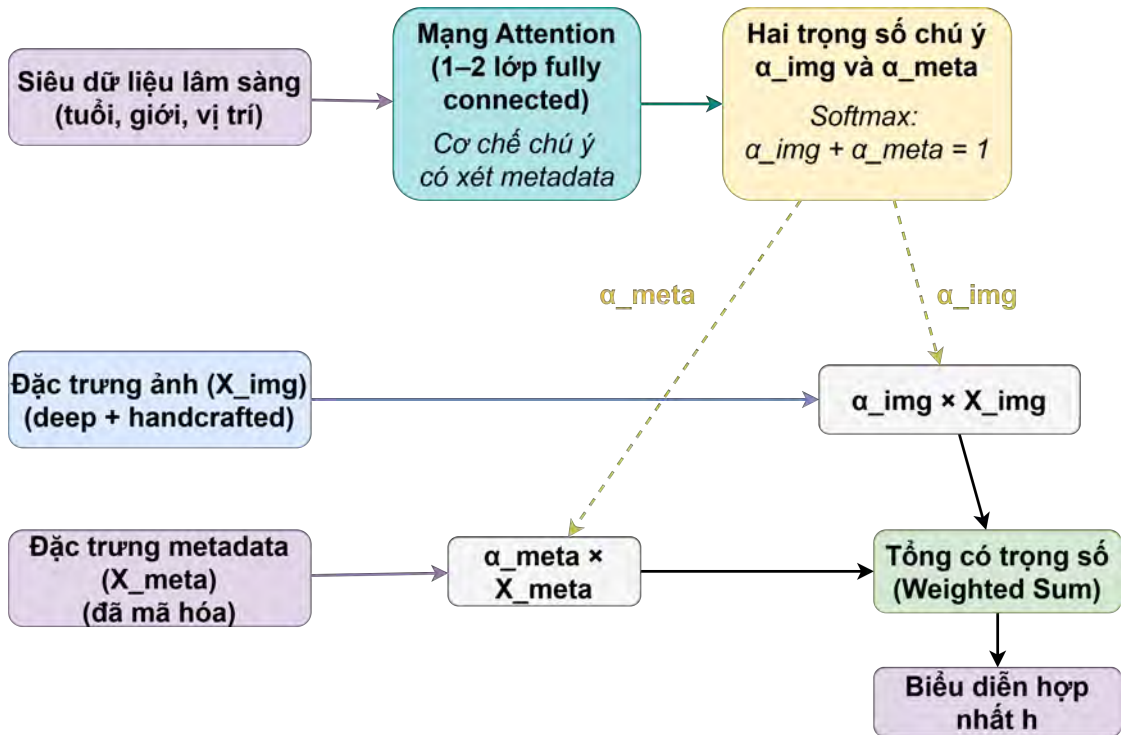
Sau khi hoàn thành bước trích xuất đặc trưng đa nguồn, mỗi mẫu dữ liệu được biểu diễn bởi ba thành phần gồm đặc trưng học sâu $\mathbf{x}_i^{deep}$, đặc trưng thủ công $\mathbf{x}_i^{hand}$ và metadata lâm sàng $\mathbf{x}_i^{meta}$. Theo cấu trúc phương pháp trong [CT4], phép ghép nối và cơ chế chú ý được sử dụng ở hai tầng có vai trò khác nhau. Phép ghép nối kết hợp hai nguồn đặc trưng thì giác là đặc trưng học sâu và đặc trưng thủ công để tạo biểu diễn ảnh. Cơ chế chú ý có xét metadata sau đó hợp nhất biểu diễn ảnh với metadata lâm sàng. Cách tổ chức này hạn chế việc metadata có số chiều thấp bị lấn át khi ghép trực tiếp với toàn bộ các thành phần thì giác, đồng thời cho phép xây dựng biểu diễn cuối cùng mang cả thông tin ảnh và ngữ cảnh lâm sàng.

Trong phương pháp đề xuất, trước hết đặc trưng học sâu và đặc trưng thủ công được kết hợp để tạo thành biểu diễn ảnh $\mathbf{x}_i^{img}$, được xác định theo Công thức 3.13.

$$\mathbf{x}_i^{img} = [\mathbf{x}_i^{deep}; \mathbf{x}_i^{hand}], \quad (3.13)$$

trong đó ký hiệu $[\cdot]$ biểu diễn phép nối vector. Biểu diễn $\mathbf{x}_i^{img}$ chứa thông tin thì giác tổng hợp từ đặc trưng học sâu và đặc trưng thủ công, trong khi $\mathbf{x}_i^{meta}$ biểu diễn thông tin metadata lâm sàng sau khi đã được mã hóa.

Sau khi $\mathbf{x}_i^{img}$ được hình thành, cơ chế chú ý có xét metadata thực hiện hợp nhất giữa hai đường biểu diễn $\mathbf{x}_i^{img}$ và $\mathbf{x}_i^{meta}$. Các thành phần $\mathbf{x}_i^{deep}$ và $\mathbf{x}_i^{hand}$ không được xem là hai nhánh chú ý riêng, mà đã nằm trong cùng biểu diễn ảnh. Cơ chế hợp nhất này được minh họa trong Hình 3.6.



Hình 3.6: Hợp nhất đặc trưng đa nguồn theo cơ chế chú ý có xét metadata.

Với metadata $\mathbf{x}_i^{meta}$, mạng attention tạo ra hai hệ số chú ý tương ứng với đặc trưng ảnh và metadata. Hai hệ số này được xác định theo Công thức 3.14.

$$\left[\alpha_i^{img}, \alpha_i^{meta} \right] = \text{softmax} \left(\mathbf{W}_a \mathbf{x}_i^{meta} + \mathbf{b}_a \right), \quad (3.14)$$

trong đó $\mathbf{W}_a$ và $\mathbf{b}_a$ là các tham số của mạng attention, α_i^{img} là trọng số chú ý ứng với đặc trưng ảnh và α_i^{meta} là trọng số chú ý ứng với metadata. Do được chuẩn hóa bằng hàm softmax, hai trọng số này thỏa mãn điều kiện trong Công thức 3.15.

$$\alpha_i^{img} + \alpha_i^{meta} = 1, \quad (3.15)$$

trong đó giá trị của các trọng số phản ánh mức độ đóng góp tương đối giữa thông tin thị giác và metadata lâm sàng trong quá trình xây dựng biểu diễn hợp nhất. Trong nghiên cứu này, cơ chế attention được sử dụng nhằm điều chỉnh tương tác giữa các nguồn biểu diễn trong không gian đặc trưng chung, thay vì giả định rằng mô hình có thể học được quan hệ nhân quả giữa metadata và ảnh tổn thương da.

Từ các trọng số chú ý này, biểu diễn hợp nhất của mẫu thứ i được xây dựng bằng tổng có trọng số giữa biểu diễn ảnh và metadata. Biểu diễn hợp nhất $\mathbf{h}_i$ được xác định theo Công thức 3.16.

$$\mathbf{h}_i = \alpha_i^{img} \mathbf{x}_i^{img} + \alpha_i^{meta} \mathbf{x}_i^{meta}, \quad (3.16)$$

trong đó biểu diễn cuối cùng không chỉ chứa thông tin thị giác của tổn thương da mà còn được điều chỉnh bởi ngữ cảnh lâm sàng tương ứng với từng mẫu dữ liệu.

Trong trường hợp hai vector $\mathbf{x}_i^{img}$ và $\mathbf{x}_i^{meta}$ có số chiều khác nhau, các phép biến đổi tuyến tính được sử dụng để ánh xạ chúng về cùng không gian biểu diễn trước khi thực hiện tổng có trọng số. Khi đó, Công thức 3.16 được viết lại dưới dạng tổng quát hơn theo Công thức 3.17.

$$\mathbf{h}_i = \alpha_i^{img} \mathbf{W}_{img} \mathbf{x}_i^{img} + \alpha_i^{meta} \mathbf{W}_{meta} \mathbf{x}_i^{meta}, \quad (3.17)$$

trong đó $\mathbf{W}_{img}$ và $\mathbf{W}_{meta}$ là các ma trận chiếu nhằm đưa đặc trưng ảnh và metadata lâm sàng về cùng không gian biểu diễn. Trọng số α_i^{img} áp dụng cho toàn bộ biểu diễn ảnh đã ghép nối, không áp dụng riêng cho đặc trưng học sâu; tương tự, α_i^{meta} áp dụng cho đường metadata. Vì vậy, Công thức 3.17 mô tả chú ý giữa hai đường ảnh và metadata, thống nhất với cấu trúc hợp nhất phân cấp trong Công thức 3.4 và [CT4].

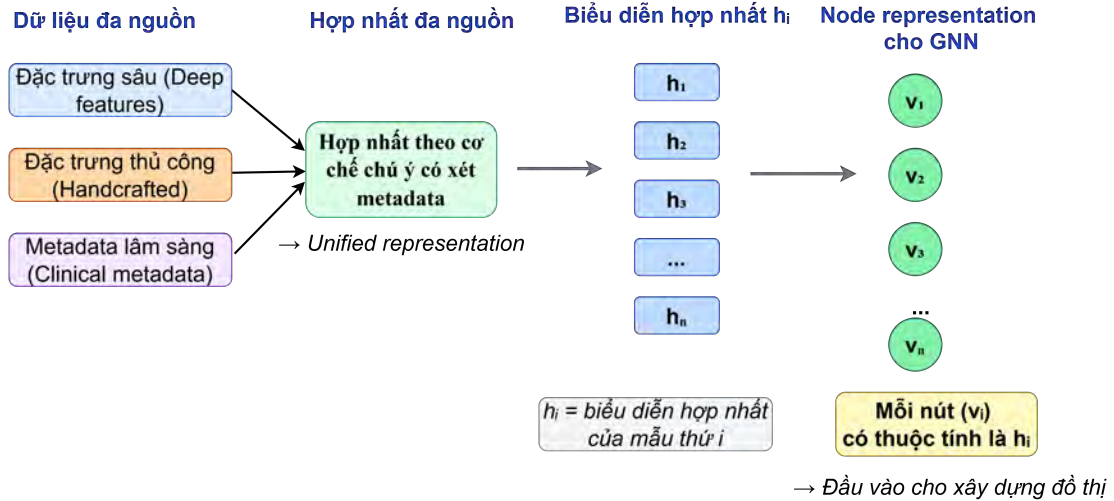
Biểu diễn hợp nhất $\mathbf{h}_i$ không chỉ được sử dụng cho quá trình phân loại mà còn đóng vai trò là biểu diễn nút đầu vào cho mô hình học trên đồ thị. Nói cách khác, sau bước hợp nhất, mỗi mẫu dữ liệu được xem như một nút với thuộc tính nút là $\mathbf{h}_i$. Các biểu diễn nút này tiếp tục được sử dụng để xác định quan hệ tương đồng giữa các mẫu và xây dựng đồ thị dữ liệu trong bước tiếp theo.

3.3.3. Xây dựng đồ thị từ biểu diễn hợp nhất

Sau khi hoàn thành bước hợp nhất đặc trưng đa nguồn, mỗi mẫu dữ liệu được biểu diễn bởi vector hợp nhất $\mathbf{h}_i$ như đã xác định trong Công thức 3.4. Các biểu diễn hợp nhất này tiếp tục được sử dụng để xây dựng đồ thị dữ liệu nhằm mô hình hóa quan hệ tương đồng giữa các mẫu trong không gian biểu diễn chung. Một điểm trong quá trình này là

cấu trúc đồ thị được xây dựng hoàn toàn từ đặc trưng dữ liệu; thông tin nhãn lớp không tham gia vào quá trình tạo cạnh. Điều này giúp hạn chế nguy cơ rò rỉ dữ liệu và bảo đảm rằng các quan hệ trên đồ thị phản ánh sự tương đồng tự nhiên giữa các mẫu thay vì phụ thuộc trực tiếp vào nhãn phân loại.

Trong mô hình đề xuất, mỗi mẫu tổn thương da được ánh xạ thành một nút v_i của đồ thị, trong đó thuộc tính nút được biểu diễn bởi vector hợp nhất $\mathbf{h}_i$. Quá trình chuyển từ biểu diễn hợp nhất sang biểu diễn nút cho mô hình mạng nơ-ron đồ thị được minh họa trong Hình 3.7.



Hình 3.7: Biểu diễn nút đa nguồn được xây dựng từ đặc trưng hợp nhất cho mô hình mạng nơ-ron đồ thị.

Từ các biểu diễn nút này, đồ thị dữ liệu được xây dựng nhằm mô hình hóa quan hệ tương đồng giữa các mẫu trong không gian biểu diễn hợp nhất. Định nghĩa đồ thị được sử dụng trong nghiên cứu này được trình bày trong Định nghĩa 3.1.

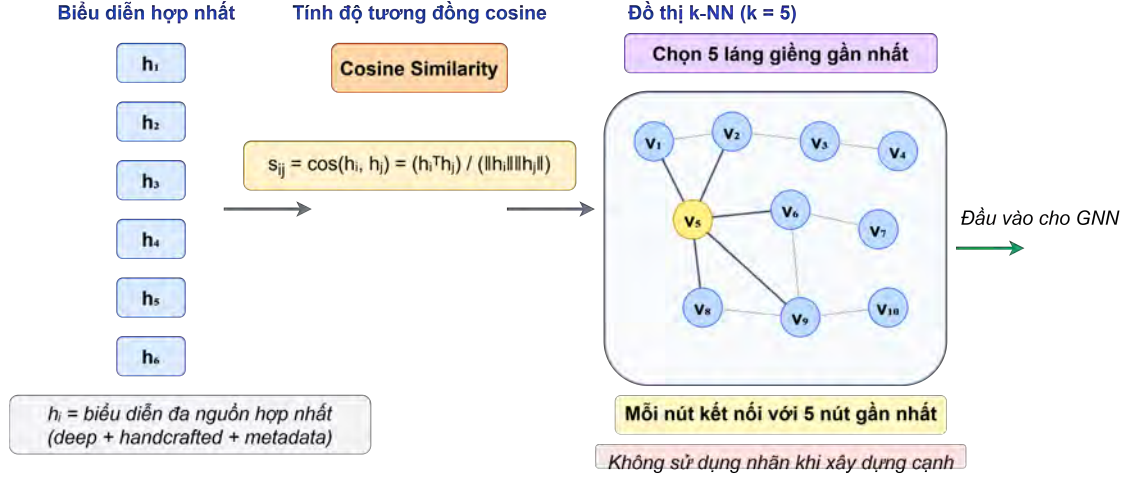
ĐỊNH NGHĨA 3.1 [Đồ thị đa nguồn dữ liệu [CT4]] Đồ thị dữ liệu được ký hiệu là $G = (V, E)$, trong đó V là tập nút của đồ thị và E là tập cạnh biểu diễn các kết nối giữa các nút. Mỗi nút $v_i \in V$ tương ứng với một mẫu tổn thương da, còn các cạnh trong tập E phản ánh quan hệ tương đồng giữa các mẫu dữ liệu trong không gian biểu diễn hợp nhất.

Một ví dụ về đồ thị k-NN được xây dựng từ các biểu diễn hợp nhất được trình bày trong Ví dụ 3.1. Trong ví dụ này, mỗi nút biểu diễn một mẫu dữ liệu và các cạnh được hình thành dựa trên độ tương đồng cosine giữa các biểu diễn nút.

VÍ DỤ 3.1 Hình 3.8 minh họa quá trình xây dựng đồ thị k-NN với $k = 5$ từ các biểu diễn hợp nhất $\mathbf{h}_i$. Đối với mỗi nút v_i , độ tương đồng cosine được tính với các nút còn lại và năm nút có độ tương đồng lớn nhất được lựa chọn để thiết lập cạnh. Các cạnh này biểu diễn quan hệ lân cận giữa các mẫu dữ liệu trong không gian đặc trưng hợp nhất.

Với mỗi mẫu thứ i , một nút v_i được tạo ra và được gán thuộc tính nút là $\mathbf{h}_i$. Do đó, cấu trúc đồ thị không chỉ lưu trữ thông tin biểu diễn của từng mẫu mà còn mô hình hóa quan hệ giữa các mẫu dữ liệu trong không gian biểu diễn đa nguồn.

Để xác định cạnh giữa các nút, độ tương đồng giữa hai biểu diễn nút $\mathbf{h}_i$ và $\mathbf{h}_j$ được tính bằng độ tương đồng cosine. Độ tương đồng này đã được xác định theo Công thức 3.6. Giá trị s_{ij} càng lớn cho thấy hai mẫu i và j càng gần nhau trong không gian biểu diễn hợp nhất.



Hình 3.8: Đồ thị k-NN với $k = 5$ được xây dựng từ biểu diễn hợp nhất dựa trên độ tương đồng cosine.

Độ tương đồng cosine được lựa chọn vì đây là một độ đo đơn giản, ổn định và được sử dụng phổ biến trong các bài toán dựa trên biểu diễn vector nhiều chiều. Khác với khoảng cách Euclidean, cosine similarity tập trung vào hướng của vector thay vì độ lớn tuyệt đối, nhờ đó phù hợp hơn trong bối cảnh các nguồn đặc trưng có thể có phân bố hoặc thang đo khác nhau sau quá trình hợp nhất. Tuy nhiên, nghiên cứu này không xem cosine similarity là lựa chọn tối ưu tuyệt đối cho mọi trường hợp, mà sử dụng như một độ đo phù hợp và ổn định cho quá trình xây dựng đồ thị trong thực nghiệm.

Từ ma trận độ tương đồng cosine, đồ thị được xây dựng theo chiến lược k -nearest neighbors với $k = 5$. Cụ thể, đối với mỗi nút v_i , năm nút có độ tương đồng cosine lớn nhất với v_i được lựa chọn làm các nút lân cận. Tập lân cận của nút v_i được xác định theo Công thức 3.18.

$$\mathcal{N}_5(v_i) = \text{TopK}_{j \neq i}(s_{ij}, 5), \quad (3.18)$$

trong đó $\mathcal{N}_5(v_i)$ là tập năm nút lân cận gần nhất của nút v_i theo độ tương đồng cosine. Ký hiệu $\text{TopK}_{j \neq i}(s_{ij}, 5)$ biểu diễn thao tác chọn năm nút khác v_i có giá trị s_{ij} lớn nhất.

Giá trị $k = 5$ được lựa chọn nhằm giữ được quan hệ cục bộ giữa các mẫu gần nhau trong không gian biểu diễn đồng thời hạn chế ảnh hưởng của các mẫu ở xa. Nếu k quá nhỏ, đồ thị có thể bị phân mảnh và khó lan truyền thông tin giữa các vùng dữ liệu. Ngược lại, nếu k quá lớn, đồ thị có thể trở nên quá dày và làm lan truyền thông tin từ các mẫu không thực sự tương đồng, từ đó làm giảm khả năng phân biệt của cấu trúc đồ thị. Trong nghiên cứu này, $k = 5$ được sử dụng như một cấu hình thực nghiệm cố định nhằm tạo ra cấu trúc đồ thị tương đối ổn định cho quá trình huấn luyện mô hình mạng nơ-ron đồ thị.

Tập cạnh của đồ thị được xác định dựa trên quan hệ lân cận này. Cụ thể, nếu v_j thuộc tập lân cận $\mathcal{N}_5(v_i)$ thì cạnh giữa v_i và v_j được thiết lập. Tập cạnh E được xác định theo Công thức 3.19.

$$E = \{(v_i, v_j) \mid v_j \in \mathcal{N}_5(v_i)\}, \quad (3.19)$$

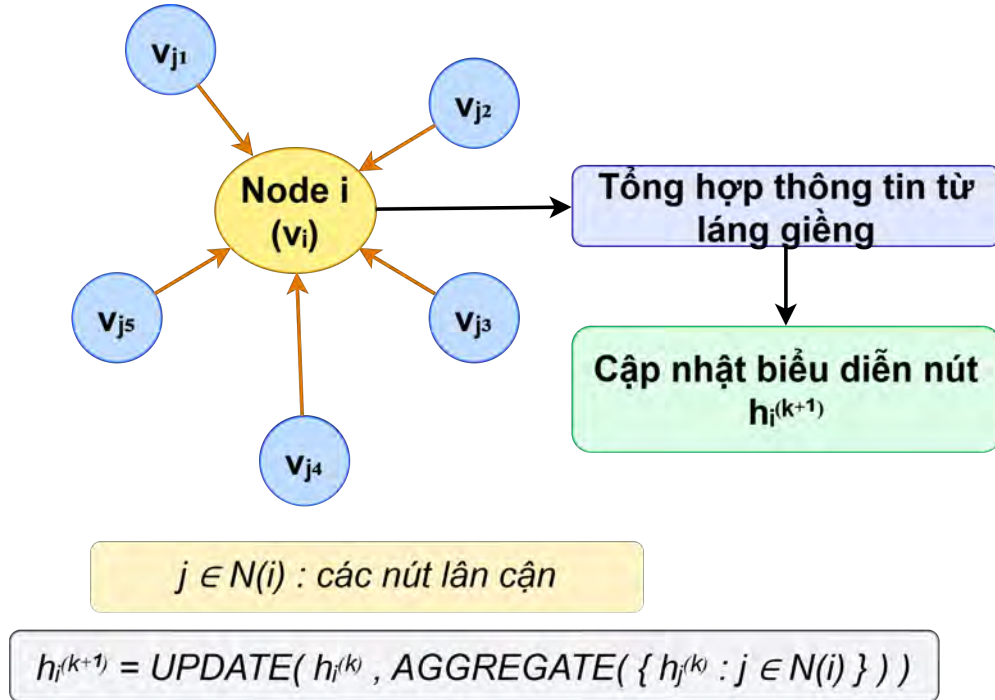
Một điểm trong quá trình xây dựng đồ thị là toàn bộ các cạnh chỉ được xác định từ độ tương đồng giữa các biểu diễn nút mà không sử dụng nhãn lớp của dữ liệu. Điều này giúp hạn chế nguy cơ rò rỉ thông tin trong quá trình xây dựng đồ thị và bảo đảm rằng quan hệ giữa các mẫu được hình thành từ đặc trưng dữ liệu thay vì từ thông tin nhãn. Nhờ đó, đồ thị thu được phản ánh cấu trúc lân cận tự nhiên trong không gian biểu diễn hợp nhất và có thể được sử dụng làm đầu vào cho mô hình mạng nơ-ron đồ thị ở pha huấn luyện tiếp theo.

3.3.4. Pha 1: Huấn luyện GNN có giám sát

Sau khi đồ thị k-NN được xây dựng từ các biểu diễn hợp nhất, pha đầu tiên của phương pháp thực hiện huấn luyện có giám sát mô hình mạng nơ-ron đồ thị trên tập dữ liệu có nhãn. Pha 1 không chỉ đóng vai trò xây dựng mô hình phân loại ban đầu, mà còn tạo cơ sở cho bước ước lượng entropy dự đoán ở Pha 2. Vì vậy, mục tiêu của pha này không đơn thuần là đạt hiệu năng phân loại cao trên dữ liệu có nhãn, mà còn hướng tới xây dựng một mô hình nền có khả năng tạo ra phân phối xác suất dự đoán đủ ổn định để đánh giá mức độ tin cậy của dữ liệu chưa gán nhãn trong giai đoạn tiếp theo.

Đầu vào của pha này gồm đồ thị $G = (V, E)$ đã được xác định theo Công thức 3.5, biểu diễn nút $\mathbf{h}_i$ của mỗi mẫu dữ liệu và nhãn y_i của các mẫu thuộc tập dữ liệu có nhãn D_L được xác định theo Công thức 3.2. Trên cơ sở đó, mô hình mạng nơ-ron đồ thị được huấn luyện để học hàm phân loại trên đồ thị, trong đó mỗi nút được dự đoán dựa trên cả đặc trưng của chính nó và thông tin lan truyền từ các nút lân cận.

Cơ chế học chính của mô hình mạng nơ-ron đồ thị là lan truyền thông tin trên đồ thị, thường được gọi là *message passing*. Với mỗi nút v_i , mô hình không chỉ sử dụng biểu diễn ban đầu $\mathbf{h}_i$ của chính nút đó, mà còn tổng hợp thông tin từ tập lân cận $\mathcal{N}_5(v_i)$ đã được xác định theo Công thức 3.18. Cơ chế message passing được minh họa trong Hình 3.9.



Hình 3.9: Cơ chế lan truyền thông tin trong mô hình mạng nơ-ron đồ thị.

Tại lớp thứ k của mô hình mạng nơ-ron đồ thị, biểu diễn của nút v_i được cập nhật

dựa trên biểu diễn hiện tại của chính nó và thông tin tổng hợp từ các nút lân cận. Quá trình này được mô tả tổng quát theo Công thức 3.20.

$$\mathbf{h}_i^{(k+1)} = \text{UPDATE}^{(k)} \left(\mathbf{h}_i^{(k)}, \text{AGGREGATE}^{(k)} \left(\left\{ \mathbf{h}_j^{(k)} : v_j \in \mathcal{N}_5(v_i) \right\} \right) \right), \quad (3.20)$$

trong đó $\mathbf{h}_i^{(k)}$ là biểu diễn của nút v_i tại lớp thứ k , $\mathcal{N}_5(v_i)$ là tập năm nút lân cận gần nhất của v_i , $\text{AGGREGATE}^{(k)}(\cdot)$ là hàm tổng hợp thông tin từ các nút lân cận và $\text{UPDATE}^{(k)}(\cdot)$ là hàm cập nhật biểu diễn nút.

Ở lớp đầu tiên, biểu diễn ban đầu của nút được xác định bởi Công thức 3.21.

$$\mathbf{h}_i^{(0)} = \mathbf{h}_i, \quad (3.21)$$

trong đó $\mathbf{h}_i$ là biểu diễn hợp nhất đa nguồn của mẫu thứ i đã được xác định theo Công thức 3.4. Sau K lớp mạng nơ-ron đồ thị, biểu diễn nút cuối cùng $\mathbf{h}_i^{(K)}$ được sử dụng cho bước phân loại nút. Xác suất dự đoán của mẫu thứ i được xác định theo Công thức 3.22.

$$\hat{y}_i = \text{softmax} \left(\mathbf{W}_o \mathbf{h}_i^{(K)} + \mathbf{b}_o \right), \quad (3.22)$$

trong đó $\mathbf{W}_o$ và $\mathbf{b}_o$ là các tham số của lớp phân loại, còn $\hat{y}_i = [\hat{y}_{i,0}, \hat{y}_{i,1}]$ là phân phối xác suất dự đoán tương ứng với hai lớp lành tính và ác tính. Công thức 3.22 cũng thống nhất với cách biểu diễn tổng quát của hàm dự đoán mạng nơ-ron đồ thị đã nêu trong Công thức 3.7.

Trong pha huấn luyện có giám sát, hàm mất mát chỉ được tính trên các nút có nhãn thuộc tập dữ liệu D_L . Các nút thuộc tập dữ liệu chưa gán nhãn không tham gia vào quá trình tính hàm mất mát trong Pha 1. Điều này giúp bảo đảm rằng mô hình ở giai đoạn đầu chỉ học từ thông tin nhãn đã biết, trong khi dữ liệu chưa gán nhãn chỉ được khai thác ở Pha 2 thông qua cơ chế gán nhãn giả có kiểm soát độ tin cậy.

Với bài toán phân loại nhị phân, hàm mất mát cross-entropy được xác định theo Công thức 3.23.

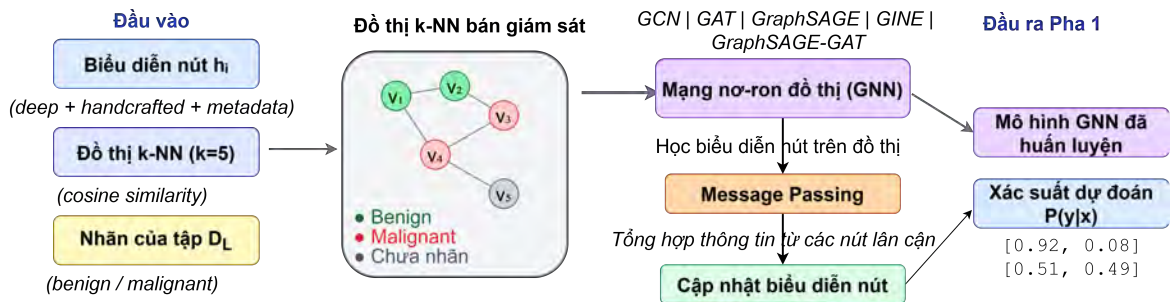
$$\mathcal{L}_{sup} = -\frac{1}{N_L} \sum_{i=1}^{N_L} \sum_{c=0}^1 \mathbb{I}(y_i = c) \log(\hat{y}_{i,c}), \quad (3.23)$$

trong đó $\mathbb{I}(y_i = c)$ là hàm chỉ thị nhận giá trị 1 nếu mẫu thứ i thuộc lớp c và nhận giá trị 0 trong trường hợp ngược lại. Việc tính hàm mất mát chỉ trên các nút có nhãn cho phép mô hình học có giám sát, trong khi cấu trúc đồ thị vẫn hỗ trợ lan truyền thông tin giữa các nút thông qua quan hệ lân cận.

Trong nghiên cứu này, nhiều biến thể mạng nơ-ron đồ thị khác nhau được khảo sát nhằm đánh giá ảnh hưởng của các cơ chế lan truyền thông tin khác nhau đối với bài toán phân loại tổn thương da. Mục tiêu của việc sử dụng nhiều kiến trúc nền không phải để tìm kiếm một kiến trúc tối ưu tuyệt đối, mà nhằm phân tích khả năng học quan hệ liên mẫu của các cơ chế tổng hợp và cập nhật biểu diễn nút khác nhau, từ đó lựa chọn mô hình phù hợp cho pha gán nhãn giả ở bước tiếp theo.

Cụ thể, GCN được sử dụng như một mô hình nền cho học bán giám sát trên đồ thị [20]; GAT được khảo sát nhằm đánh giá ảnh hưởng của cơ chế chú ý giữa các nút lân cận [22]; GraphSAGE được sử dụng để phân tích cơ chế học biểu diễn theo hướng quy nạp dựa trên tổng hợp lân cận [21]. Bên cạnh đó, GINE được khảo sát dựa trên khả năng biểu diễn của họ mô hình GIN và các mở rộng trong học biểu diễn đồ thị [91, 126]. Biến thể GraphSAGE–GAT được sử dụng nhằm kết hợp cơ chế tổng hợp lân cận của GraphSAGE với cơ chế chú ý của GAT [21, 22]. Mặc dù các kiến trúc nền này khác nhau về cách tổng hợp thông tin hoặc cách gán trọng số cho các nút lân cận, chúng đều tuân theo nguyên lý chung của message passing được mô tả trong Công thức 3.20.

Quy trình huấn luyện có giám sát trong Pha 1 được minh họa trong Hình 3.10. Từ biểu diễn nút $\mathbf{h}_i$, đồ thị k-NN với $k = 5$ và nhãn của tập dữ liệu D_L , mô hình mạng nơ-ron đồ thị được huấn luyện để học biểu diễn trên đồ thị và sinh phân phối xác suất dự đoán cho từng mẫu dữ liệu.



Hình 3.10: Pha 1: Huấn luyện có giám sát mô hình mạng nơ-ron đồ thị trên đồ thị k-NN.

Kết quả của Pha 1 là mô hình mạng nơ-ron đồ thị đã được huấn luyện cùng với phân phối xác suất dự đoán $\hat{y}_i$ cho các mẫu dữ liệu. Các xác suất dự đoán này không chỉ được sử dụng để xác định nhãn phân loại ban đầu theo Công thức 3.8, mà còn đóng vai trò là cơ sở cho bước ước lượng entropy dự đoán và đánh giá độ tin cậy của dữ liệu chưa gán nhãn trong Pha 2.

3.3.5. Pha 2: Gán nhãn giả có chọn lọc dựa trên độ tin cậy

Sau khi hoàn thành Pha 1, mô hình mạng nơ-ron đồ thị đã được huấn luyện được sử dụng để sinh phân phối xác suất dự đoán cho các mẫu dữ liệu chưa gán nhãn trong tập D_U . Khác với các phương pháp gán nhãn giả thông thường sử dụng toàn bộ dữ liệu chưa gán nhãn để mở rộng tập huấn luyện, phương pháp đề xuất chỉ lựa chọn các mẫu thỏa điều kiện độ tin cậy trước khi thực hiện gán nhãn giả. Đây là điểm khác biệt của phương pháp trong chương này.

Mục tiêu của Pha 2 là khai thác thêm thông tin từ dữ liệu chưa gán nhãn nhằm mở rộng tập huấn luyện, đồng thời hạn chế nguy cơ lan truyền sai số trong quá trình học bán giám sát. Nếu các mẫu có mức độ bất định cao cũng được đưa vào quá trình gán nhãn giả, các nhãn dự đoán sai có thể tiếp tục được sử dụng trong bước tái huấn luyện và làm suy giảm hiệu quả của mô hình. Vì vậy, phương pháp đề xuất sử dụng entropy dự đoán như một cơ chế đánh giá độ tin cậy trước khi lựa chọn mẫu cho gán nhãn giả.

Định nghĩa entropy dự đoán được trình bày trong Định nghĩa 3.2.

ĐỊNH NGHĨA 3.2 [Entropy dự đoán [CT4]] Với phân phối xác suất dự đoán $\hat{y}_i = [\hat{y}_{i,1}, \hat{y}_{i,2}, \dots, \hat{y}_{i,C}]$ trên C lớp của mẫu dữ liệu thứ i , entropy dự đoán được xác định theo Công thức 3.24. Giá trị entropy phản ánh mức độ bất định của mô hình đối với dự đoán

tương ứng, trong đó entropy càng lớn cho thấy mô hình càng không chắc chắn đối với kết quả phân loại.

$$H(\hat{\mathbf{y}}_i) = - \sum_{c=1}^C \hat{y}_{i,c} \log(\hat{y}_{i,c}), \quad (3.24)$$

trong đó $\hat{y}_{i,c}$ là xác suất dự đoán của mẫu thứ i thuộc lớp c .

Trong phương pháp đề xuất, entropy thấp thường tương ứng với phân phối xác suất dự đoán tập trung hơn vào một lớp và được xem như một chỉ báo cho mức độ tự tin cao hơn của mô hình. Ngược lại, entropy cao phản ánh phân phối dự đoán phân tán hơn và mức độ bất định lớn hơn trong quá trình phân loại. Tuy nhiên, entropy không được xem là một bảo đảm tuyệt đối về tính đúng đắn của nhãn dự đoán. Trong nghiên cứu này, entropy chỉ đóng vai trò như một đại lượng xấp xỉ đơn giản cho uncertainty của mô hình và vẫn có thể chịu ảnh hưởng bởi mức độ calibration của phân phối xác suất đầu ra.

Ví dụ về quá trình sinh xác suất dự đoán và tính entropy cho các mẫu dữ liệu chưa gán nhãn được trình bày trong Ví dụ 3.2.

VÍ DỤ 3.2 [Ví dụ tính entropy dự đoán] Sau khi mô hình mạng nơ-ron đồ thị được huấn luyện trên tập dữ liệu có nhãn D_L , mô hình được sử dụng để dự đoán trên các mẫu dữ liệu chưa gán nhãn thuộc tập D_U . Với mỗi mẫu dữ liệu, mô hình sinh ra phân phối xác suất dự đoán:

$$\hat{\mathbf{y}}_i = [\hat{y}_{i,1}, \hat{y}_{i,2}, \dots, \hat{y}_{i,C}],$$

sau đó entropy của từng mẫu được tính theo Công thức 3.9.

Quá trình mô hình mạng nơ-ron đồ thị sinh xác suất dự đoán và tính entropy cho các mẫu dữ liệu chưa gán nhãn được minh họa trong Hình 3.11.

Từ các giá trị entropy dự đoán, phương pháp xây dựng cơ chế phân tầng độ tin cậy nhằm chia các mẫu chưa gán nhãn thành ba nhóm gồm độ tin cậy cao, độ tin cậy trung bình và độ tin cậy thấp. Việc phân tầng này cho phép lựa chọn các mẫu có entropy thấp hơn cho quá trình gán nhãn giả, trong khi các mẫu có entropy cao hơn sẽ không được sử dụng để tái huấn luyện.

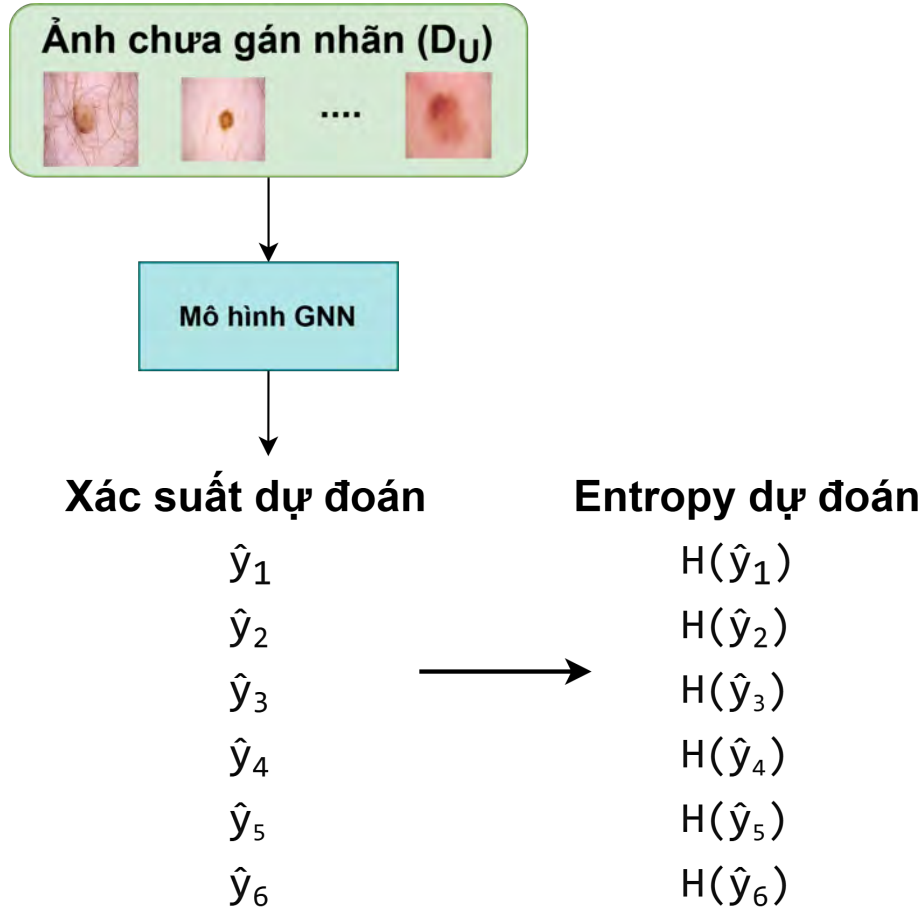
Giả sử tập entropy dự đoán của các mẫu chưa gán nhãn được ký hiệu là:

$$\mathcal{H} = \{H(\hat{\mathbf{y}}_1), H(\hat{\mathbf{y}}_2), \dots, H(\hat{\mathbf{y}}_{N_U})\}.$$

Trong nghiên cứu này, hai ngưỡng entropy được xác định từ phân bố entropy của tập dữ liệu chưa gán nhãn thông qua các phân vị của tập $\mathcal{H}$. Cụ thể, ngưỡng thứ nhất α_1 và ngưỡng thứ hai α_2 được xác định theo Công thức 3.25.

$$\alpha_1 = Q_{0.33}(\mathcal{H}), \quad \alpha_2 = Q_{0.66}(\mathcal{H}), \quad (3.25)$$

trong đó $Q_p(\mathcal{H})$ là phân vị bậc p của phân bố entropy $\mathcal{H}$ được tính trên tập dữ liệu chưa gán nhãn. Như vậy, α_1 tương ứng với ngưỡng phân chia khoảng 33% mẫu có entropy thấp nhất, còn α_2 tương ứng với ngưỡng phân chia khoảng 66% mẫu có entropy thấp hơn trong phân bố entropy của tập chưa gán nhãn.



Hình 3.11: Ví dụ quá trình sinh xác suất dự đoán và tính entropy cho các mẫu dữ liệu chưa gán nhãn.

Trên cơ sở hai ngưỡng này, hàm ánh xạ mức độ tin cậy $T(\cdot)$ được xác định theo Công thức 3.26.

$$T(\hat{y}_i) = \begin{cases} \text{High}, & H(\hat{y}_i) \leq \alpha_1, \\ \text{Medium}, & \alpha_1 < H(\hat{y}_i) \leq \alpha_2, \\ \text{Low}, & H(\hat{y}_i) > \alpha_2. \end{cases} \quad (3.26)$$

Trong ánh xạ này, nhóm độ tin cậy cao bao gồm các mẫu có entropy thấp nhất và được xem là các mẫu có mức độ tự tin cao hơn của mô hình. Nhóm độ tin cậy trung bình gồm các mẫu có entropy trung bình, còn nhóm độ tin cậy thấp gồm các mẫu có entropy cao hơn và mức độ bất định lớn hơn. Trong quy trình thực nghiệm của nghiên cứu này, chỉ các mẫu thuộc nhóm độ tin cậy cao được sử dụng cho quá trình gán nhãn giả và tái huấn luyện.

Ví dụ về quá trình xác định các ngưỡng entropy α_1 và α_2 được trình bày trong Ví dụ 3.3.

VÍ DỤ 3.3 [Ví dụ xác định ngưỡng entropy và mức độ tin cậy] Giả sử mô hình mạng nơ-ron đồ thị sau Pha 1 được sử dụng để dự đoán trên tập dữ liệu chưa gán nhãn D_U . Với mỗi mẫu dữ liệu, entropy dự đoán được tính theo Công thức 3.9.

Giả sử năm giá trị entropy thu được là:

$$H = [0.325, 0.325, 0.611, 0.611, 0.693].$$

Sau khi sắp xếp tăng dần:

$$H_{\text{sort}} = [0.325, 0.325, 0.611, 0.611, 0.693].$$

Ngưỡng thứ nhất được xác định theo phân vị 0.33:

$$p_1 = (n - 1) \times 0.33 = (5 - 1) \times 0.33 = 1.32.$$

Giá trị $p_1 = 1.32$ nằm giữa vị trí thứ hai và thứ ba trong dãy entropy đã sắp xếp:

$$H_2 = 0.325, \quad H_3 = 0.611.$$

Sử dụng nội suy tuyến tính:

$$\alpha_1 = 0.325 + 0.32 \times (0.611 - 0.325).$$

Suy ra:

$$\alpha_1 = 0.325 + 0.32 \times 0.286 = 0.325 + 0.09152 \approx 0.416.$$

Ngưỡng thứ hai được xác định theo phân vị 0.66:

$$p_2 = (n - 1) \times 0.66 = (5 - 1) \times 0.66 = 2.64.$$

Giá trị $p_2 = 2.64$ nằm giữa vị trí thứ ba và thứ tư:

$$H_3 = 0.611, \quad H_4 = 0.611.$$

Khi đó:

$$\alpha_2 = 0.611 + 0.64 \times (0.611 - 0.611) = 0.611.$$

Từ hai ngưỡng này, các mẫu có $H(\hat{\mathbf{y}}_i) \leq 0.416$ được xếp vào nhóm độ tin cậy cao, các mẫu có $0.416 < H(\hat{\mathbf{y}}_i) \leq 0.611$ được xếp vào nhóm độ tin cậy trung bình, còn các mẫu có $H(\hat{\mathbf{y}}_i) > 0.611$ được xếp vào nhóm độ tin cậy thấp.

Quá trình xác định ngưỡng entropy và phân tầng độ tin cậy được minh họa trong Hình 3.12.

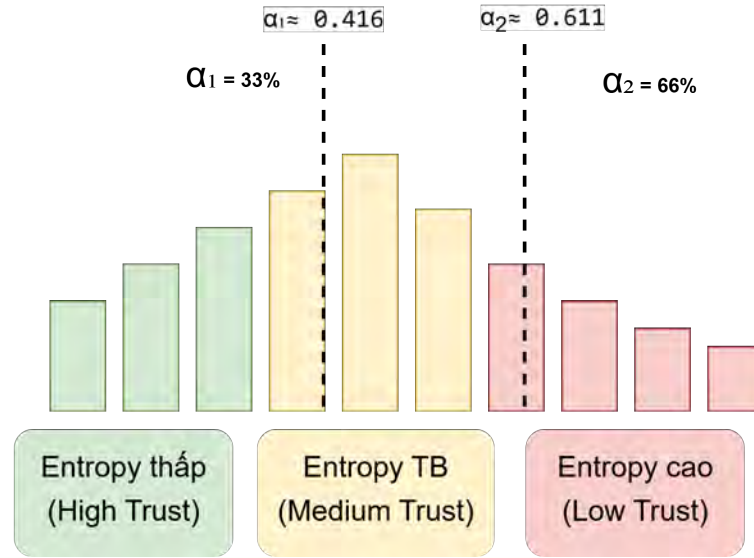
Sau khi xác định các ngưỡng entropy, mô hình mạng nơ-ron đồ thị được sử dụng để dự đoán trên các mẫu chưa gán nhãn thuộc tập D_U . Với mỗi mẫu chưa gán nhãn $x_j \in D_U$, mô hình sinh phân phối xác suất dự đoán $\hat{\mathbf{y}}_j$ và entropy tương ứng được tính theo Công thức 3.9. Ví dụ về quá trình đánh giá entropy cho một mẫu chưa gán nhãn được trình bày trong Ví dụ 3.4.

VÍ DỤ 3.4 [Ví dụ đánh giá entropy cho mẫu chưa gán nhãn] Giả sử một mẫu dữ liệu chưa gán nhãn $x_j \in D_U$ được đưa vào mô hình mạng nơ-ron đồ thị đã huấn luyện. Mô hình sinh phân phối xác suất dự đoán:

$$\hat{\mathbf{y}}_j = [0.80, 0.10].$$

Entropy dự đoán của mẫu được tính theo Công thức 3.9:

$$H(\hat{\mathbf{y}}_j) = -(0.80 \log(0.80) + 0.10 \log(0.10)).$$



Hình 3.12: Ví dụ xác định ngưỡng entropy và phân tầng độ tin cậy của dự đoán.

Kết quả thu được:

$$H(\hat{\mathbf{y}}_j) \approx 0.50.$$

Quá trình đánh giá entropy cho mẫu chưa gán nhãn được minh họa trong Hình 3.13.

So sánh với các ngưỡng $\alpha_1 = 0.416$ và $\alpha_2 = 0.611$, ta có:

$$0.416 < H(\hat{\mathbf{y}}_j) = 0.50 \leq 0.611.$$

Do đó, mẫu dữ liệu x_j thuộc nhóm độ tin cậy trung bình. Trong quy trình thực nghiệm của nghiên cứu này, các mẫu độ tin cậy trung bình không được sử dụng cho quá trình gán nhãn giả và chỉ được dùng cho mục đích phân tích độ tin cậy của mô hình.

Đối với các mẫu thuộc nhóm độ tin cậy cao, nhãn giả được xác định bằng lớp có xác suất dự đoán lớn nhất. Quy tắc gán nhãn giả có chọn lọc được xác định theo Công thức 3.27.

$$\tilde{y}_j = \arg \max_c \hat{y}_{j,c}, \quad \text{if } T(\hat{\mathbf{y}}_j) = \text{High}, \quad (3.27)$$

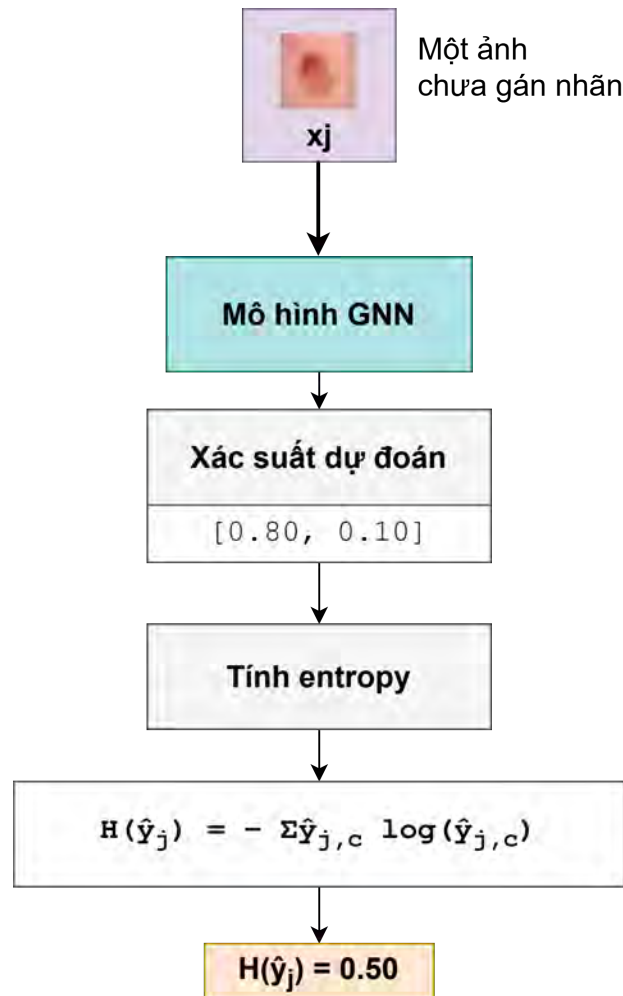
hay tương đương:

$$\tilde{y}_j = \arg \max_c \hat{y}_{j,c}, \quad \text{if } H(\hat{\mathbf{y}}_j) \leq \alpha_1. \quad (3.28)$$

Ví dụ về lựa chọn mẫu có độ tin cậy cao cho gán nhãn giả được trình bày trong Ví dụ 3.5.

VÍ DỤ 3.5 [Ví dụ lựa chọn mẫu có độ tin cậy cao cho gán nhãn giả] Giả sử một mẫu chưa gán nhãn có entropy dự đoán:

$$H(\hat{\mathbf{y}}_j) = 0.30.$$



Hình 3.13: Ví dụ tính entropy cho một mẫu dữ liệu chưa gán nhãn.

So sánh với ngưỡng entropy thứ nhất:

$$\alpha_1 = 0.416,$$

ta có:

$$H(\hat{\mathbf{y}}_j) = 0.30 < 0.416.$$

Do đó:

$$T(\hat{\mathbf{y}}_j) = \text{High}.$$

Mẫu này được lựa chọn cho quá trình gán nhãn giả.

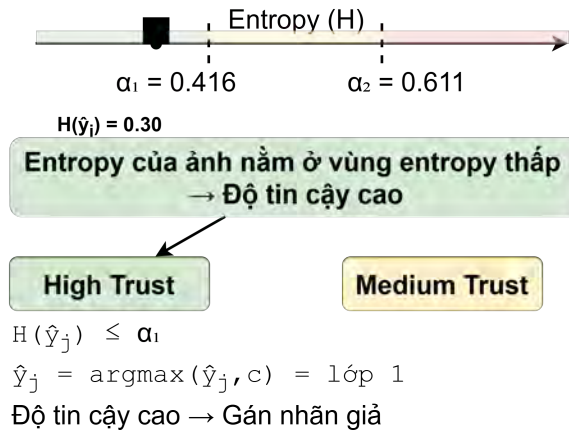
Nhãn giả của mẫu được xác định:

$$\tilde{y}_j = \arg \max_c \hat{y}_{j,c}.$$

Giả sử lớp có xác suất dự đoán lớn nhất là lớp 1, khi đó:

$$\tilde{y}_j = 1.$$

Quá trình lựa chọn mẫu có độ tin cậy cao và gán nhãn giả được minh họa trong Hình 3.14.

So sánh entropy của ảnh x_j với ngưỡng

Hình 3.14: Ví dụ lựa chọn mẫu có độ tin cậy cao cho quá trình gán nhãn giả.

Sau bước gán nhãn giả, tập huấn luyện mở rộng được hình thành bằng cách kết hợp tập dữ liệu có nhãn ban đầu với tập các mẫu nhãn giả thuộc nhóm độ tin cậy cao. Tập dữ liệu mở rộng được xác định theo Công thức 3.29.

$$D_{\text{new}} = D_L \cup D_{\text{pseudo}}^{\text{High}}, \quad (3.29)$$

trong đó $D_{\text{pseudo}}^{\text{High}}$ là tập các mẫu chưa gán nhãn được gán nhãn giả và thuộc nhóm độ tin cậy cao. Trong quy trình thực nghiệm của nghiên cứu này, chỉ các mẫu có độ tin cậy cao được đưa vào tập huấn luyện mở rộng; các mẫu độ tin cậy trung bình và độ tin cậy thấp không được sử dụng cho bước tái huấn luyện nhằm hạn chế lan truyền sai số từ các dự đoán có độ bất định cao hơn.

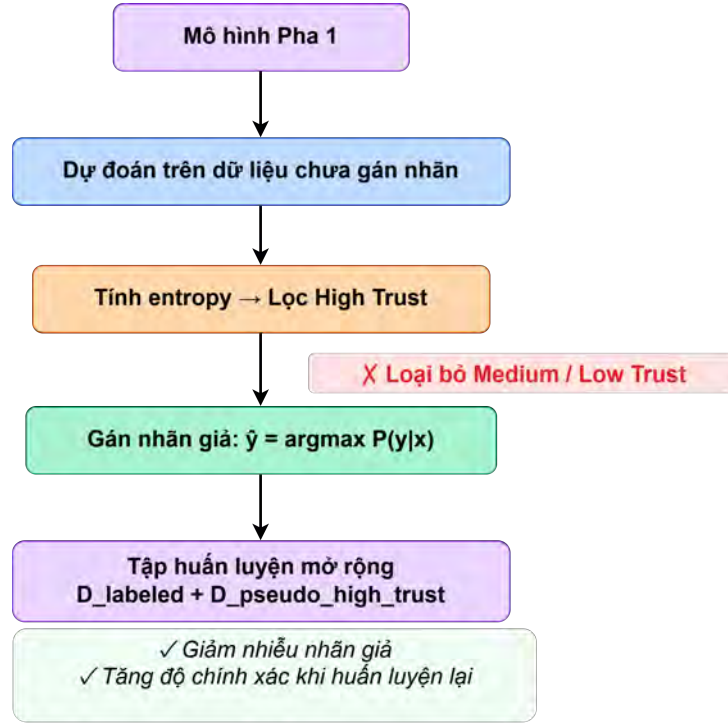
Quá trình mở rộng dữ liệu huấn luyện bằng gán nhãn giả được minh họa trong Hình 3.15.

Sau khi tập dữ liệu mở rộng được hình thành, mô hình mạng nơ-ron đồ thị được huấn luyện lại trên tập dữ liệu mới nhằm tinh chỉnh biểu diễn nút và cải thiện khả năng phân loại. Quá trình này không được mô tả như một vòng lặp self-training vô hạn, mà là một bước tái huấn luyện có chọn lọc dựa trên các mẫu có độ tin cậy cao. Cách tiếp cận này cho phép khai thác thêm thông tin từ dữ liệu chưa gán nhãn đồng thời giảm nguy cơ tích lũy sai số do sử dụng các nhãn giả không đáng tin cậy.

Mặc dù lọc entropy giúp giảm nguy cơ lan truyền nhãn sai, phương pháp vẫn có thể tồn tại các trường hợp dự đoán sai nhưng có entropy thấp do mô hình quá tự tin đối với một số mẫu dữ liệu. Vì vậy, hiệu quả của phương pháp không chỉ được đánh giá bằng Accuracy, mà còn được phân tích thông qua các chỉ số cân bằng lớp như Macro F1, Macro Recall, Macro AUC, confusion matrix và phân tích theo mức độ độ tin cậy ở phần thực nghiệm của chương này.

3.4. Thuật toán và quy trình thực hiện

Từ các thành phần đã trình bày trong Mục 3.3, quy trình thực hiện được mô tả bằng ba thuật toán. Thuật toán 3.1 xây dựng biểu diễn nút đa nguồn và đồ thị k-NN. Thuật toán 3.2 huấn luyện mô hình GNN trên tập có nhãn. Thuật toán 3.3 lựa chọn các mẫu chưa gán nhãn có entropy thấp để tạo nhãn giả và tái huấn luyện mô hình.



Hình 3.15: Mở rộng tập huấn luyện sử dụng các mẫu nhãn giả thuộc nhóm độ tin cậy cao.

Ba thuật toán tương ứng với các bước chính của phương pháp: xây dựng biểu diễn đa nguồn, tạo đồ thị tương đồng, huấn luyện có giám sát, ước lượng độ bất định bằng entropy, chọn mẫu có độ tin cậy cao và tái huấn luyện. Cách trình bày này tách rõ pha chuẩn bị dữ liệu, pha học có giám sát và pha học bán giám sát có kiểm soát độ tin cậy.

Phân tích thuật toán. Thuật toán 3.1 chuyển mỗi mẫu dữ liệu thành một nút của đồ thị theo đúng hai tầng hợp nhất. Ở tầng thứ nhất, $\mathbf{x}_i^{deep}$ và $\mathbf{x}_i^{hand}$ được ghép nối thành $\mathbf{x}_i^{img}$. Ở tầng thứ hai, cơ chế chú ý hợp nhất $\mathbf{x}_i^{img}$ với $\mathbf{x}_i^{meta}$ để tạo $\mathbf{h}_i$. Các cạnh được xây dựng bằng chiến lược k-NN trên không gian biểu diễn hợp nhất, không sử dụng nhãn lớp. Do đó, đồ thị phản ánh quan hệ tương đồng giữa các mẫu theo đặc trưng đầu vào thay vì theo nhãn phân loại.

Giả sử N là số mẫu, d là số chiều của biểu diễn nút $\mathbf{h}_i$. Bước hợp nhất đặc trưng có độ phức tạp $O(Nd)$. Nếu tính đầy đủ độ tương đồng giữa mọi cặp mẫu, bước xây dựng ma trận tương đồng có độ phức tạp $O(N^2d)$. Bước chọn k láng giềng cho từng nút phụ thuộc vào cách cài đặt; với cách sắp xếp trực tiếp, chi phí là $O(N^2 \log N)$, còn với các cấu trúc tìm kiếm lân cận gần đúng có thể giảm trong thực nghiệm. Trong thiết lập của luận án, phần chi phí chính đến từ bước tính tương đồng cặp nên độ phức tạp được viết gọn là $O(N^2d)$ khi d chi phối chi phí tính toán khoảng cách.

Sau khi Thuật toán 3.1 tạo đồ thị $G = (V, E)$ và các biểu diễn nút $\mathbf{h}_i$, bước tiếp theo là huấn luyện mô hình GNN trên các nút có nhãn. Nội dung này được trình bày trong Thuật toán 3.2.

Phân tích thuật toán. Thuật toán 3.2 chỉ sử dụng nhãn của tập huấn luyện D_L để tính hàm mất mát. Tập kiểm định D_{val} chỉ được dùng để lựa chọn tham số tốt nhất và dừng huấn luyện, không được dùng để cập nhật trọng số mô hình. Cách tách này giúp hạn chế rò rỉ thông tin trong quá trình đánh giá.

Thuật toán 3.1 Hợp nhất đa nguồn và xây dựng đồ thị k-NN [CT4]

Đầu vào: Tập mẫu $\mathcal{D}$; đặc trưng học sâu $\mathbf{x}_i^{deep}$; đặc trưng thủ công $\mathbf{x}_i^{hand}$; metadata lâm sàng $\mathbf{x}_i^{meta}$; số láng giềng $k = 5$.

Đầu ra: Đồ thị $G = (V, E)$; tập biểu diễn nút $\{\mathbf{h}_i\}_{i=1}^N$.

- 1: Khởi tạo $V \leftarrow \emptyset, E \leftarrow \emptyset$.
 - 2: **for** mỗi mẫu $i \in \mathcal{D}$ **do**
 - 3: Ghép đặc trưng ảnh: $\mathbf{x}_i^{img} = [\mathbf{x}_i^{deep}; \mathbf{x}_i^{hand}]$ theo Công thức (3.13).
 - 4: Tính trọng số chú ý theo metadata: $[\alpha_i^{img}, \alpha_i^{meta}] = \text{softmax}(\mathbf{W}_a \mathbf{x}_i^{meta} + \mathbf{b}_a)$ theo Công thức (3.14).
 - 5: Tạo biểu diễn nút hợp nhất: $\mathbf{h}_i = \alpha_i^{img} \mathbf{W}_{img} \mathbf{x}_i^{img} + \alpha_i^{meta} \mathbf{W}_{meta} \mathbf{x}_i^{meta}$ theo Công thức (3.17).
 - 6: Thêm nút v_i với thuộc tính nút $\mathbf{h}_i$ vào V .
 - 7: **end for**
 - 8: **for** mỗi cặp nút (v_i, v_j) với $i \neq j$ **do**
 - 9: Tính độ tương đồng cosine s_{ij} theo Công thức (3.6).
 - 10: **end for**
 - 11: **for** mỗi nút $v_i \in V$ **do**
 - 12: Xác định tập k láng giềng gần nhất $\mathcal{N}_k(v_i)$ theo Công thức (3.18).
 - 13: Thêm cạnh (v_i, v_j) vào E với mọi $v_j \in \mathcal{N}_k(v_i)$.
 - 14: **end for**
 - 15: Xác định tập cạnh E theo Công thức (3.19).
 - 16: **return** $G = (V, E)$ và $\{\mathbf{h}_i\}_{i=1}^N$.
-

Thuật toán 3.2 Pha 1: Huấn luyện GNN có giám sát

Đầu vào: Đồ thị $G = (V, E)$; tập biểu diễn nút $\{\mathbf{h}_i\}_{i=1}^N$; tập nút có nhãn D_L ; nhãn y_i với $i \in D_L$; tập kiểm định D_{val} ; số epoch tối đa $T_{\max}^{(1)}$.

Đầu ra: Tham số tốt nhất θ^* của mô hình GNN; xác suất dự đoán $\hat{y}_i$; biểu diễn nút sau học đồ thị $\mathbf{z}_i$.

- 1: Khởi tạo mô hình GNN f_θ với tham số ban đầu θ .
 - 2: **for** $t = 1$ đến $T_{\max}^{(1)}$ **do**
 - 3: Thực hiện lan truyền thông tin trên đồ thị theo Công thức (3.20).
 - 4: Tính xác suất dự đoán $\hat{y}_i$ theo Công thức (3.22).
 - 5: Tính hàm mất mát có giám sát $\mathcal{L}_{sup}$ trên các mẫu thuộc D_L theo Công thức (3.23).
 - 6: Cập nhật tham số θ bằng lan truyền ngược.
 - 7: Đánh giá mô hình trên D_{val} .
 - 8: **if** giá trị chỉ số kiểm định được chọn đạt tốt nhất **then**
 - 9: Lưu tham số $\theta^* \leftarrow \theta$.
 - 10: **end if**
 - 11: **end for**
 - 12: **return** $\theta^*, \hat{y}_i$ và $\mathbf{z}_i$.
-

Với N là số nút, $|E|$ là số cạnh, d là số chiều biểu diễn nút và $T_{\max}^{(1)}$ là số epoch huấn luyện, chi phí của một epoch phụ thuộc vào kiến trúc nền GNN. Với các mô hình

message passing thông dụng, chi phí mỗi epoch xấp xỉ $O(|E|d)$. Do đó, độ phức tạp của Thuật toán 3.2 là $O(T_{\max}^{(1)}|E|d)$. Với đồ thị k-NN cố định có $|E| = O(kN)$, độ phức tạp có thể viết là $O(T_{\max}^{(1)}kNd)$.

Sau pha huấn luyện có giám sát, mô hình f_{θ^*} được dùng để dự đoán trên tập chưa gán nhãn. Các dự đoán này không được xem là nhãn thật; chúng chỉ là cơ sở để tính entropy và lựa chọn mẫu có độ tin cậy cao cho pha gán nhãn giả trong Thuật toán 3.3.

Thuật toán 3.3 Pha 2: Đánh giá entropy, gán nhãn giả có chọn lọc và tái huấn luyện GNN [CT4]

Đầu vào: Đồ thị $G = (V, E)$; tập biểu diễn nút $\{\mathbf{h}_i\}_{i=1}^N$; mô hình đã huấn luyện f_{θ^*} ; tập có nhãn D_L ; tập chưa gán nhãn D_U ; tập kiểm định D_{val} ; số epoch tái huấn luyện $T_{\max}^{(2)}$; phân vị entropy $q_1 = 0.33$ và $q_2 = 0.66$.

Đầu ra: Tập nhãn giả độ tin cậy cao $D_{\text{pseudo}}^{\text{High}}$; tập huấn luyện mở rộng D_{new} ; mô hình sau tái huấn luyện $f_{\theta'^*}$.

```

1: Khởi tạo  $D_{\text{pseudo}}^{\text{High}} \leftarrow \emptyset$  và  $\mathcal{H} \leftarrow \emptyset$ .
2: for mỗi mẫu  $x_j \in D_U$  do
3:   Tính xác suất dự đoán  $\hat{\mathbf{y}}_j = f_{\theta^*}(G, \mathbf{h}_j)$ .
4:   Tính entropy  $H(\hat{\mathbf{y}}_j)$  theo Công thức (3.9).
5:   Thêm  $H(\hat{\mathbf{y}}_j)$  vào  $\mathcal{H}$ .
6: end for
7: Xác định ngưỡng entropy  $\alpha_1 = Q_{q_1}(\mathcal{H})$  và  $\alpha_2 = Q_{q_2}(\mathcal{H})$  theo Công thức (3.25).
8: for mỗi mẫu  $x_j \in D_U$  do
9:   Xác định mức độ tin cậy  $T(\hat{\mathbf{y}}_j)$  theo Công thức (3.26).
10:  if  $H(\hat{\mathbf{y}}_j) \leq \alpha_1$  then
11:    Tạo nhãn giả  $\tilde{y}_j = \arg \max_c \hat{y}_{j,c}$  theo Công thức (3.27).
12:    Thêm  $(x_j, \tilde{y}_j)$  vào  $D_{\text{pseudo}}^{\text{High}}$ .
13:  else
14:    Không đưa mẫu  $x_j$  vào tập huấn luyện mở rộng.
15:  end if
16: end for
17: Tạo tập huấn luyện mở rộng  $D_{\text{new}} = D_L \cup D_{\text{pseudo}}^{\text{High}}$  theo Công thức (3.29).
18: Khởi tạo mô hình Pha 2 từ tham số Pha 1:  $\theta' \leftarrow \theta^*$ .
19: for  $t = 1$  đến  $T_{\max}^{(2)}$  do
20:   Thực hiện lan truyền thông tin trên đồ thị  $G$ .
21:   Tính xác suất dự đoán  $\hat{\mathbf{y}}_i = f_{\theta'}(G, \mathbf{h}_i)$ .
22:   Tính hàm mất mát tái huấn luyện trên  $D_{\text{new}}$ .
23:   Cập nhật tham số  $\theta'$  bằng lan truyền ngược.
24:   Đánh giá mô hình trên  $D_{val}$ .
25:   if giá trị chỉ số kiểm định được chọn đạt tốt nhất then
26:     Lưu tham số  $\theta'^* \leftarrow \theta'$ .
27:   end if
28: end for
29: return  $D_{\text{pseudo}}^{\text{High}}$ ,  $D_{\text{new}}$  và  $f_{\theta'^*}$ .

```

Phân tích thuật toán. Thuật toán 3.3 sử dụng entropy của phân phối dự đoán để

lựa chọn mẫu chưa gán nhãn. Các mẫu có entropy thấp được đưa vào $D_{\text{pseudo}}^{\text{High}}$, trong khi các mẫu có entropy cao hơn không tham gia tái huấn luyện. Nhãn $\tilde{y}_j$ là nhãn giả do mô hình sinh ra, không phải nhãn tham chiếu. Vì vậy, tập ISIC 2020 Test không có nhãn công khai chỉ được dùng như nguồn dữ liệu chưa gán nhãn, không được dùng để báo cáo chỉ số đánh giá có giám sát.

Độ phức tạp của Thuật toán 3.3 gồm bốn phần: suy luận trên D_U , tính entropy và phân vị, chọn nhãn giả, và tái huấn luyện. Gọi $N_U = |D_U|$, C là số lớp, d là số chiều biểu diễn nút và $N_{\text{new}} = |D_{\text{new}}|$. Bước suy luận trên tập chưa gán nhãn phụ thuộc vào kiến trúc nền GNN; nếu xét theo đồ thị k-NN, chi phí xấp xỉ $O(kN_U d)$. Bước tính entropy có chi phí $O(N_U C)$. Bước tính phân vị entropy bằng sắp xếp đầy đủ có chi phí $O(N_U \log N_U)$. Bước tái huấn luyện có chi phí xấp xỉ $O(T_{\text{max}}^{(2)} k N_{\text{new}} d)$. Do đó, độ phức tạp tổng quát có thể viết là:

$$O(kN_U d + N_U C + N_U \log N_U + T_{\text{max}}^{(2)} k N_{\text{new}} d).$$

Trong trường hợp các phân vị entropy được tính bằng thuật toán chọn tuyến tính, thành phần $O(N_U \log N_U)$ có thể được thay bằng $O(N_U)$. Tuy nhiên, biểu thức trên phản ánh cách cài đặt phổ biến dựa trên sắp xếp.

Nhìn chung, quy trình tổng thể của phương pháp kết nối ba nhóm thao tác: xây dựng biểu diễn và đồ thị, huấn luyện có giám sát, và học bán giám sát bằng nhãn giả có lọc entropy. Cơ chế này cho phép khai thác dữ liệu chưa gán nhãn theo hướng có kiểm soát, đồng thời giới hạn việc tái huấn luyện ở các mẫu có mức độ bất định thấp.

3.5. Thiết lập thực nghiệm

Mục này trình bày thiết lập thực nghiệm được sử dụng để đánh giá phương pháp học bán giám sát đa nguồn trên đồ thị cho bài toán phân loại ảnh tổn thương da. Thực nghiệm được xây dựng nhằm đánh giá ba thành phần chính của phương pháp đề xuất, bao gồm biểu diễn đa nguồn, học trên đồ thị và gán nhãn giả có chọn lọc dựa trên entropy.

Để bảo đảm tính công bằng trong so sánh, các mô hình được huấn luyện trên cùng tập dữ liệu, cùng cấu hình đồ thị, cùng quy trình xây dựng đặc trưng và cùng các chỉ số đánh giá. Nội dung mục này bao gồm bộ dữ liệu sử dụng, phương pháp xây dựng đặc trưng đa nguồn, cấu hình đồ thị, cấu hình mô hình mạng nơ-ron đồ thị, các mô hình so sánh và các chỉ số đánh giá.

Để hạn chế rò rỉ dữ liệu, các bước chuẩn hóa metadata, chuẩn hóa đặc trưng, lựa chọn tham số tiền xử lý và cân bằng dữ liệu được thực hiện trên phần huấn luyện của từng lần chia dữ liệu. Tập kiểm định và tập kiểm tra không được sử dụng để ước lượng các tham số tiền xử lý. Đối với tập ISIC 2020 Test không có nhãn công khai, dữ liệu này chỉ được dùng như tập chưa gán nhãn trong pha gán nhãn giả; các nhãn giả không được xem là nhãn thật và không được dùng để báo cáo các chỉ số đánh giá có giám sát.

Quy trình đánh giá được tổ chức theo các lần chia dữ liệu lập phân tầng trên tập có nhãn. Trong mỗi lần chia, tập huấn luyện, tập kiểm định và tập kiểm tra được tách riêng trước khi thực hiện các bước chuẩn hóa, cân bằng lớp, lựa chọn mô hình và đánh giá. Các tham số tiền xử lý như trung bình, độ lệch chuẩn, bộ mã hóa metadata và các tham số chuẩn hóa đặc trưng được ước lượng từ tập huấn luyện, sau đó mới áp dụng cho tập kiểm định và tập kiểm tra. Cách tổ chức này nhằm bảo đảm rằng thông tin từ tập kiểm

định và tập kiểm tra không tham gia vào quá trình huấn luyện mô hình.

Trong pha gán nhãn giả, mô hình đã huấn luyện ở pha có giám sát chỉ sử dụng xác suất dự đoán trên tập chưa gán nhãn để tính entropy và lựa chọn mẫu. Nhãn của tập kiểm tra có nhãn, nếu có trong các lần kiểm chứng nội bộ, không được sử dụng để tạo nhãn giả hoặc chọn ngưỡng entropy. Các kết quả được báo cáo dưới dạng trung bình và độ lệch chuẩn trên các lần kiểm chứng để giảm phụ thuộc vào một lần chia dữ liệu cụ thể.

3.5.1. Bộ dữ liệu ISIC 2020

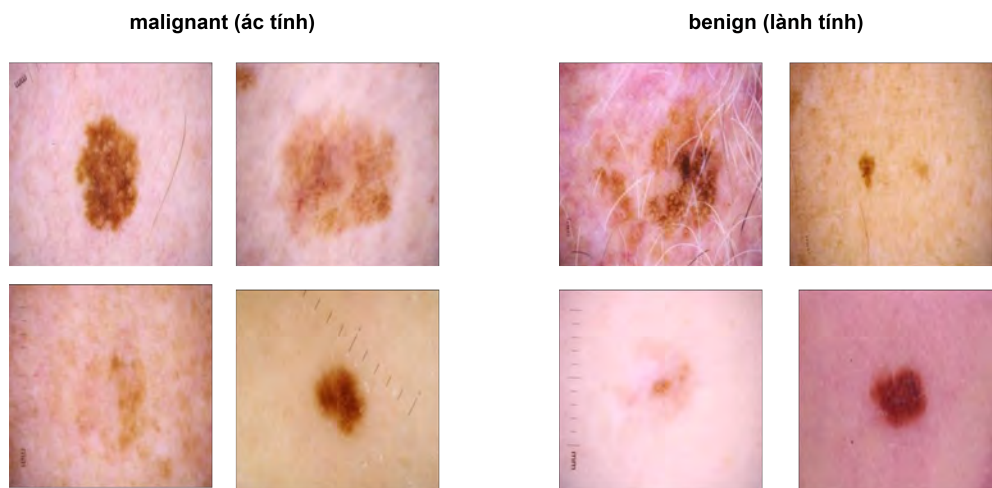
Các thực nghiệm trong chương này được thực hiện trên bộ dữ liệu ISIC 2020 [127], là một trong những bộ dữ liệu lớn và phổ biến trong nghiên cứu phân loại tổn thương da tự động. Bộ dữ liệu không chỉ cung cấp ảnh dermoscopy mà còn bao gồm metadata lâm sàng của bệnh nhân như tuổi, giới tính và vị trí tổn thương. Đây là một đặc điểm giúp ISIC 2020 phù hợp với hướng nghiên cứu học đa nguồn trong chương này, trong đó biểu diễn dữ liệu được xây dựng đồng thời từ ảnh và metadata lâm sàng thay vì chỉ sử dụng thông tin thị giác đơn thuần.

ISIC 2020 được tổ chức theo bài toán phân loại nhị phân giữa tổn thương lành tính và tổn thương ác tính. Bộ dữ liệu có mất cân bằng lớp giữa hai nhóm, do đó phù hợp để đánh giá các chỉ số cân bằng lớp và cơ chế gán nhãn giả có chọn lọc dựa trên entropy.

Tập huấn luyện có nhãn gồm 33,126 ảnh dermoscopy. Tập ISIC 2020 Test gồm 10,982 ảnh và không cung cấp nhãn phân loại công khai. Trong luận án, tập này được sử dụng như tập dữ liệu chưa gán nhãn D_U cho quá trình gán nhãn giả ở Pha 2; các nhãn giả chỉ phản ánh dự đoán của mô hình sau bước lọc độ tin cậy, không được xem là nhãn tham chiếu.

Các mẫu tổn thương trong ISIC 2020 có sự đa dạng lớn về màu sắc, kết cấu, hình dạng và kích thước, đồng thời tồn tại nhiều trường hợp có sự tương đồng thị giác giữa hai lớp lành tính và ác tính. Điều này làm tăng độ khó của bài toán phân loại tự động và tạo ra môi trường phù hợp để đánh giá khả năng khai thác quan hệ liên mẫu của mô hình mạng nơ-ron đồ thị.

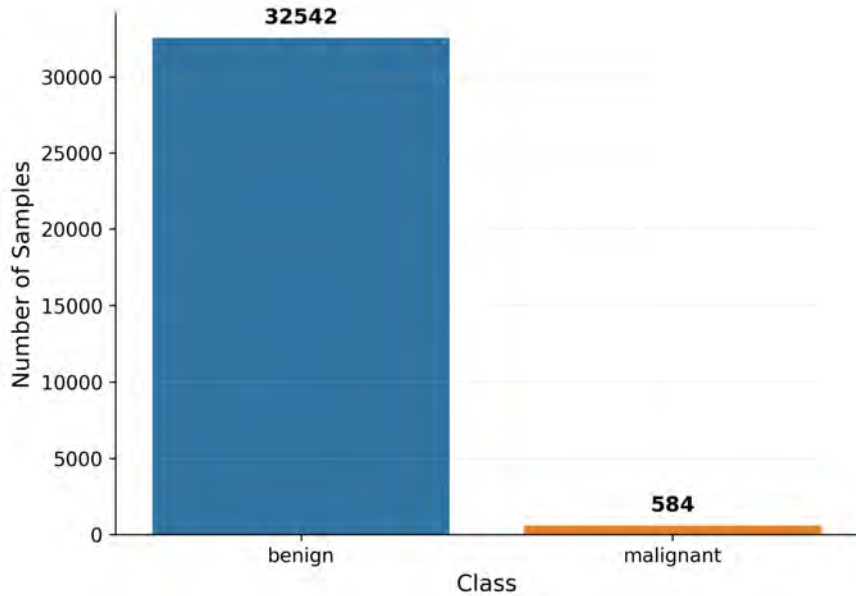
Hình 3.16 minh họa một số mẫu ảnh lành tính và ác tính trong bộ dữ liệu ISIC 2020.



Hình 3.16: Ví dụ các mẫu tổn thương lành tính và ác tính trong bộ dữ liệu ISIC 2020.

Sau khi chuẩn hóa nhãn, tập dữ liệu có nhãn gồm 32,542 mẫu lành tính và 584 mẫu ác tính. Có thể thấy bộ dữ liệu bị mất cân bằng lớn giữa hai lớp, điều thường gặp trong

các bài toán y sinh thực tế. Phân bố lớp của bộ dữ liệu được minh họa trong Hình 3.17.



Hình 3.17: Phân bố lớp lành tính và ác tính trong bộ dữ liệu ISIC 2020.

Bảng 3.1 trình bày số lượng mẫu của từng lớp.

Bảng 3.1: Phân bố dữ liệu có nhãn và chưa gán nhãn trong ISIC 2020

Tập dữ liệu	Lành tính	Ác tính	Tổng số mẫu
D_L (ISIC 2020 Train)	32,542	584	33,126
D_U (ISIC 2020 Test)	Chưa có nhãn		10,982

Do sự mất cân bằng lớp, các chỉ số đánh giá cân bằng lớp như Macro F1, Macro AUC, Sensitivity và Specificity được ưu tiên sử dụng thay vì chỉ dựa trên Accuracy nhằm phản ánh chính xác hơn hiệu quả phân loại của mô hình đối với cả lớp lành tính và ác tính.

3.5.2. Metadata lâm sàng

Bên cạnh đặc trưng ảnh, metadata lâm sàng được sử dụng như một nguồn thông tin bổ trợ trong mô hình đa nguồn của chương này. Trong thực tế, một số loại tổn thương da có xu hướng xuất hiện phổ biến hơn ở các nhóm tuổi, giới tính hoặc vị trí giải phẫu nhất định. Vì vậy, metadata lâm sàng có thể cung cấp thêm ngữ cảnh hỗ trợ cho quá trình phân loại, đặc biệt trong các trường hợp các mẫu ảnh có biểu hiện thị giác tương đối giống nhau nhưng thuộc các nhóm bệnh lý khác nhau.

Trong nghiên cứu này, metadata lâm sàng được sử dụng gồm ba trường chính: giới tính (*sex*), tuổi xấp xỉ của bệnh nhân (*age_approx*) và vị trí giải phẫu của tổn thương (*anatom_site_general_challenge*). Trong đó, giới tính và vị trí tổn thương là các thuộc tính phân loại, còn tuổi là thuộc tính liên tục dạng số thực. Các trường metadata này được trích xuất trực tiếp từ tệp metadata chính thức của bộ dữ liệu ISIC 2020 [127].

Metadata lâm sàng có số chiều nhỏ hơn đặc trưng ảnh và được sử dụng như nguồn thông tin bổ trợ. Hiệu quả của metadata phụ thuộc vào mức độ đầy đủ, độ đồng nhất và

khả năng liên hệ của các trường thông tin với nhân bệnh. Vì vậy, metadata được dùng để bổ sung ngữ cảnh cho đặc trưng ảnh, không thay thế vai trò của thông tin thị giác.

3.5.2.1. Chuẩn hóa metadata

Metadata lâm sàng được tiền xử lý theo một quy trình thống nhất nhằm bảo đảm tính nhất quán giữa các tập dữ liệu và tránh rò rỉ thông tin trong thực nghiệm. Quy trình này bao gồm phân tích dữ liệu thiếu, xử lý giá trị thiếu, chuẩn hóa thuộc tính liên tục và mã hóa các thuộc tính phân loại trước khi thực hiện hợp nhất đa nguồn.

Một phân tích định lượng về mức độ đầy đủ của metadata được thực hiện riêng cho tập huấn luyện và tập kiểm tra của bộ dữ liệu ISIC 2020. Các thống kê giá trị thiếu được tính trực tiếp từ các tệp metadata chính thức và được trình bày trong Bảng 3.2. Trong tập huấn luyện, tỷ lệ thiếu của thuộc tính giới tính là 7.3%, thuộc tính vị trí tổn thương là 10.5% và thuộc tính tuổi là 2.4%. Tập kiểm tra có xu hướng thiếu dữ liệu tương tự, với tỷ lệ thiếu của vị trí tổn thương và tuổi cao hơn nhẹ so với tập huấn luyện.

Các thống kê này được tính độc lập trên từng tập dữ liệu nhằm bảo đảm rằng quá trình tiền xử lý không sử dụng thông tin từ tập kiểm tra hoặc tập chưa gán nhãn trong giai đoạn huấn luyện.

Bảng 3.2: Thống kê metadata thiếu trong bộ dữ liệu ISIC 2020

Metadata field	Training set (%)	Test set (%)
Sex	7.3	7.1
Anatomical site	10.5	11.2
Age	2.4	2.6

Trong nghiên cứu này, các thuộc tính phân loại bị thiếu gồm giới tính và vị trí tổn thương không bị loại bỏ khỏi dữ liệu mà được ánh xạ vào một nhóm riêng có tên unknown. Chiến lược này giúp giữ lại toàn bộ mẫu dữ liệu trong quá trình huấn luyện và tránh làm giảm số lượng mẫu khi kết hợp với dữ liệu ảnh. Đồng thời, việc sử dụng một nhóm riêng cho giá trị thiếu giúp mô hình vẫn có thể học được thông tin từ các trường metadata còn lại thay vì loại bỏ hoàn toàn mẫu dữ liệu tương ứng.

Đối với thuộc tính tuổi là dữ liệu liên tục, các giá trị thiếu được thay thế bằng giá trị tuổi trung bình được tính riêng trên tập huấn luyện. Giá trị trung bình này sau đó được sử dụng thống nhất cho tập kiểm định và tập kiểm tra nhằm tránh rò rỉ thông tin giữa các tập dữ liệu.

Sau bước xử lý thiếu dữ liệu, thuộc tính tuổi được chuẩn hóa bằng Z-score normalization theo Công thức 3.30.

$$z_i = \frac{x_i - \mu}{\sigma}, \quad (3.30)$$

trong đó x_i là giá trị tuổi ban đầu của mẫu thứ i , μ là giá trị tuổi trung bình của tập huấn luyện và σ là độ lệch chuẩn tương ứng. Việc sử dụng Z-score normalization giúp đưa thuộc tính tuổi về phân bố có trung bình bằng 0 và độ lệch chuẩn bằng 1, từ đó hạn chế ảnh hưởng của khác biệt thang đo giữa metadata và các nguồn đặc trưng khác.

Các thuộc tính phân loại gồm giới tính và vị trí tổn thương được mã hóa bằng one-

hot encoding nhằm chuyển đổi dữ liệu ký hiệu thành vector số học có kích thước cố định. Phương pháp này cho phép duy trì tính nhất quán của biểu diễn metadata giữa các tập dữ liệu và hỗ trợ quá trình hợp nhất đa nguồn trong bước tiếp theo.

3.5.2.2. Biểu diễn metadata lâm sàng

Bảng 3.3 trình bày một số ví dụ metadata lâm sàng thô trong bộ dữ liệu ISIC 2020. Các trường metadata gồm giới tính, tuổi xấp xỉ và vị trí giải phẫu của tổn thương.

Bảng 3.3: Ví dụ metadata lâm sàng thô trong bộ dữ liệu ISIC 2020

Image ID	Sex	Age (Approx.)	Anatomical Site
ISIC_2637011	male	45	head/neck
ISIC_0015719	female	45	upper extremity
ISIC_0052212	female	50	lower extremity
ISIC_0075914	male	30	torso
ISIC_0082934	male	65	torso

Sau bước chuẩn hóa và one-hot encoding, metadata được chuyển thành vector đặc trưng số học có kích thước cố định. Bảng 3.4 trình bày các ví dụ vector metadata sau mã hóa.

Bảng 3.4: Metadata lâm sàng sau chuẩn hóa và one-hot encoding

Image ID	Sex (One-Hot)	Site (One-Hot)	Age (Z-score)	Encoded Feature Vector
ISIC_2637011	[1,0]	[1,0,0,0]	-0.50	[1,0,1,0,0,0,-0.50]
ISIC_0015719	[0,1]	[0,1,0,0]	-0.50	[0,1,0,1,0,0,-0.50]
ISIC_0052212	[0,1]	[0,0,1,0]	0.00	[0,1,0,0,1,0,0.00]
ISIC_0075914	[1,0]	[0,0,0,1]	-2.00	[1,0,0,0,0,1,-2.00]
ISIC_0082934	[1,0]	[0,0,0,1]	1.50	[1,0,0,0,0,1,1.50]

Sau quá trình mã hóa, metadata lâm sàng được biểu diễn dưới dạng vector số học có kích thước cố định và được kết hợp với đặc trưng học sâu cùng đặc trưng thủ công để tạo thành biểu diễn đa nguồn cho từng mẫu dữ liệu. Các vector metadata này sau đó được sử dụng trong bước hợp nhất đặc trưng có xét metadata lâm sàng và đóng vai trò như một thành phần của biểu diễn nút trong mô hình học trên đồ thị.

3.5.3. Lựa chọn đặc trưng

Trong chương này, biểu diễn đa nguồn của mỗi mẫu dữ liệu được xây dựng từ ba nhóm thông tin chính gồm đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng. Việc kết hợp đồng thời nhiều nguồn đặc trưng nhằm khai thác tính bổ trợ giữa biểu diễn thị giác học được từ mô hình học sâu, các mô tả hình thái dựa trên tri thức miền và thông tin ngữ cảnh lâm sàng của bệnh nhân.

Đối với đặc trưng học sâu, các ứng viên được đánh giá trong chính thiết lập thực nghiệm của Chương 3 bằng giá trị Chi-square trung bình. Ba nhóm được lựa chọn là vector đầu ra sau pooling của VGG19 [50], DenseNet121 [52] và MobileNet [128], có số chiều tương ứng 512, 1024 và 1024. Đây là lựa chọn ở cấp loại đặc trưng CNN;

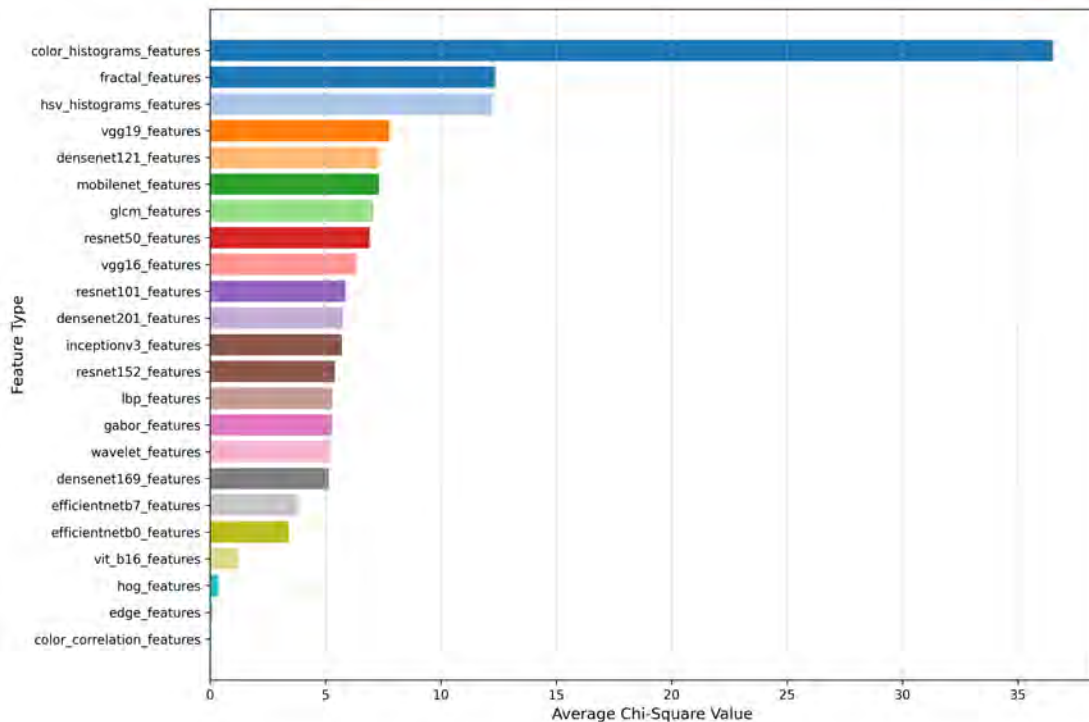
Chương 3 không áp dụng BlockNet và không chọn lọc bản đồ đặc trưng theo entropy như Chương 2.

Trong nhóm đặc trưng thủ công, nghiên cứu sử dụng đúng ba nhóm histogram màu RGB [37], histogram màu HSV [40] và đặc trưng fractal [41]. GLCM và HOG không thuộc cấu hình thực nghiệm của Chương 3. Các đặc trưng được chọn phản ánh phân bố màu và độ phức tạp hình học của tổn thương, đồng thời bổ sung cho các biểu diễn học sâu ở mức mô tả thị giác truyền thống.

Ngoài ra, metadata lâm sàng gồm tuổi, giới tính và vị trí tổn thương cũng được đưa vào biểu diễn đa nguồn nhằm cung cấp thêm thông tin ngữ cảnh cho quá trình phân loại. Các metadata này được chuẩn hóa và mã hóa như đã trình bày trong Mục 3.5.2.

Sau khi trích xuất các nhóm đặc trưng, nghiên cứu tiến hành đánh giá mức độ liên quan của từng nhóm đặc trưng bằng phân tích thống kê Chi-square [35]. Phương pháp này cho phép đo lường khả năng phân biệt của từng nhóm đặc trưng đối với bài toán phân loại lành tính và ác tính.

Hình 3.18 trình bày kết quả đánh giá mức độ của các nhóm đặc trưng theo giá trị Chi-square trung bình. Có thể quan sát rằng các đặc trưng thủ công như histogram màu RGB, đặc trưng fractal và histogram HSV đạt giá trị Chi-square cao nhất, cho thấy khả năng phân biệt đối với dữ liệu tổn thương da. Trong nhóm đặc trưng học sâu, VGG19, DenseNet121 và MobileNet đạt hiệu quả cao hơn so với nhiều kiến trúc nền CNN khác.



Hình 3.18: Đánh giá mức độ của các nhóm đặc trưng bằng phân tích Chi-square.

Dựa trên kết quả đánh giá này, nghiên cứu lựa chọn ba nhóm đặc trưng học sâu gồm VGG19, DenseNet121 và MobileNet, đồng thời lựa chọn ba nhóm đặc trưng thủ công gồm histogram màu RGB, histogram màu HSV và đặc trưng fractal cho quá trình xây dựng biểu diễn đa nguồn. Các đặc trưng được lựa chọn sau đó được hợp nhất với metadata lâm sàng để tạo thành biểu diễn nút đầu vào cho mô hình mạng nơ-ron đồ thị.

Bảng 3.5 trình bày các nhóm đặc trưng được sử dụng trong thực nghiệm.

Bảng 3.5: Các nhóm đặc trưng được lựa chọn trong nghiên cứu

Nhóm đặc trưng	Loại đặc trưng	Độ dài vector	Mô tả
VGG19 [50]	Đặc trưng học sâu	512	Biểu diễn học sâu từ kiến trúc nền CNN VGG19
DenseNet121 [52]	Đặc trưng học sâu	1024	Biểu diễn học sâu từ DenseNet121
MobileNet [128]	Đặc trưng học sâu	1024	Biểu diễn học sâu từ MobileNet
Color histogram [37]	Đặc trưng thủ công	768	Phân bố màu RGB
HSV histogram [40]	Đặc trưng thủ công	768	Phân bố màu HSV
Fractal feature [41]	Đặc trưng thủ công	64	Độ phức tạp hình học tổng hợp
Metadata lâm sàng	Metadata	10	Tuổi, giới tính, vị trí tổn thương

3.5.4. Chỉ số đánh giá

Các chỉ số đánh giá trong chương này được kế thừa từ phần cơ sở đánh giá mô hình phân loại đã trình bày ở Chương 1. Cụ thể, Accuracy, Precision, Recall và F1-score lần lượt được sử dụng để đánh giá tỷ lệ dự đoán đúng tổng thể, độ chính xác của các dự đoán dương tính, khả năng phát hiện đúng mẫu thuộc lớp quan tâm và sự cân bằng giữa Precision và Recall. Các chỉ số này đã được xác định tại các Công thức 1.41, 1.42, 1.43 và 1.46.

Bên cạnh các chỉ số phân loại cơ bản, Sensitivity và Specificity được sử dụng do có ý nghĩa lâm sàng trong bài toán phân loại tổn thương da. Trong đó, Sensitivity phản ánh khả năng phát hiện đúng các mẫu ác tính, còn Specificity phản ánh khả năng nhận diện đúng các mẫu lành tính. Hai chỉ số này được kế thừa từ các Công thức 1.44 và 1.45 ở Chương 1.

Để đánh giá khả năng phân biệt của mô hình trên toàn bộ miền ngưỡng quyết định, nghiên cứu sử dụng AUC và Macro AUC. Các chỉ số này đặc biệt phù hợp với bài toán y sinh, trong đó ngưỡng quyết định có thể thay đổi tùy theo yêu cầu sàng lọc hoặc hỗ trợ chẩn đoán. Khái niệm ROC, True Positive Rate, False Positive Rate và AUC đã được trình bày trong Chương 1 tại các Công thức 1.47, 1.48 và phần mô tả đường cong ROC.

Do bộ dữ liệu ISIC 2020 có mức mất cân bằng lớp giữa lành tính và ác tính, Accuracy có thể tạo ra đánh giá thiên lệch về hiệu năng mô hình. Ví dụ, một mô hình có xu hướng dự đoán phần lớn mẫu thuộc lớp lành tính vẫn có thể đạt Accuracy cao nhưng lại không phát hiện tốt các trường hợp ác tính. Vì vậy, trong các thực nghiệm của chương này, luận án ưu tiên sử dụng các chỉ số cân bằng lớp và có ý nghĩa lâm sàng như Macro F1, Macro AUC, Sensitivity và Specificity để phản ánh đầy đủ hơn khả năng phân loại của mô hình trên cả hai lớp dữ liệu.

Ngoài ra, do phương pháp đề xuất bao gồm học bán giám sát và gán nhãn giả có chọn lọc dựa trên entropy, các chỉ số đánh giá cũng được sử dụng để phân tích ảnh hưởng của quá trình tái huấn luyện bằng dữ liệu nhãn giả đối với hiệu năng phân loại. Các kết quả được báo cáo dưới dạng giá trị trung bình và độ lệch chuẩn qua các fold thực nghiệm nhằm đánh giá mức độ ổn định của mô hình trong các điều kiện chia dữ liệu khác nhau. Cách thiết lập này cũng phù hợp với thực nghiệm trong công trình [CT4], nơi hiệu năng của mô hình được phân tích chủ yếu thông qua Macro F1-score, Macro AUC, Sensitivity và Specificity.

3.5.5. Các mô hình so sánh

Để đánh giá toàn diện hiệu quả của phương pháp đề xuất, nghiên cứu sử dụng nhiều nhóm mô hình khác nhau làm mô hình đối sánh và đối tượng so sánh. Các mô hình được tổ chức thành bốn nhóm chính gồm: mô hình học máy truyền thống, mô hình mạng nơ-ron đồ thị không sử dụng metadata lâm sàng, mô hình mạng nơ-ron đồ thị có sử dụng metadata và mô hình học bán giám sát sau gán nhãn giả.

Nhóm thứ nhất gồm các mô hình học máy truyền thống được huấn luyện trực tiếp trên các vector đặc trưng đã trích xuất. Các mô hình được khảo sát gồm Logistic Regression, Support Vector Machine (SVM), Random Forest và XGBoost. Logistic Regression đại diện cho nhóm mô hình tuyến tính cơ bản, trong khi SVM cho phép xây dựng siêu phẳng phân tách trong không gian đặc trưng phi tuyến. Random Forest và XGBoost thuộc nhóm ensemble learning, cho phép khai thác các quan hệ phi tuyến giữa các đặc trưng thông qua tập hợp nhiều cây quyết định. Các mô hình này đóng vai trò mô hình đối sánh nhằm đánh giá hiệu quả của học trên đồ thị so với các phương pháp phân loại truyền thống.

Nhóm thứ hai gồm các kiến trúc nền mạng nơ-ron đồ thị chỉ sử dụng đặc trưng ảnh, bao gồm đặc trưng học sâu và đặc trưng thủ công. Các kiến trúc nền được khảo sát gồm GCN [129], GAT [22], GraphSAGE [21], GINE [91] và mô hình kết hợp GraphSAGE-GAT. Trong nhóm thực nghiệm này, metadata lâm sàng chưa được đưa vào quá trình học nhất đặc trưng nhằm đánh giá riêng ảnh hưởng của học quan hệ liên mẫu trên đồ thị.

Nhóm thứ ba mở rộng các kiến trúc nền mạng nơ-ron đồ thị bằng cách bổ sung metadata lâm sàng vào biểu diễn đa nguồn. Trong thiết lập này, mỗi nút của đồ thị được biểu diễn đồng thời bởi đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng. Các thực nghiệm của nhóm này nhằm đánh giá ảnh hưởng của học nhất có xét metadata đối với hiệu quả phân loại tổn thương da.

Nhóm thứ tư gồm các mô hình học bán giám sát sau quá trình gán nhãn giả ở Pha 2. Trong nhóm này, mô hình mạng nơ-ron đồ thị được tái huấn luyện bằng tập dữ liệu mở rộng bao gồm dữ liệu có nhãn ban đầu và các mẫu nhãn giả được lựa chọn từ tập dữ liệu chưa gán nhãn. Đây là nhóm thực nghiệm chính nhằm đánh giá ảnh hưởng của ước lượng độ tin cậy dựa trên entropy đối với quá trình khai thác dữ liệu chưa gán nhãn.

Để phân tích rõ hơn đóng góp của cơ chế lọc entropy, các thực nghiệm học bán giám sát được chia thành hai kịch bản. Kịch bản thứ nhất sử dụng gán nhãn giả trực tiếp trên toàn bộ dữ liệu chưa gán nhãn mà không thực hiện lọc entropy. Kịch bản thứ hai chỉ lựa chọn các mẫu thuộc nhóm độ tin cậy cao dựa trên entropy dự đoán trước khi đưa vào quá trình tái huấn luyện. Việc so sánh hai kịch bản này cho phép đánh giá trực tiếp vai trò của lọc độ tin cậy dựa trên entropy trong việc hạn chế lan truyền nhãn sai và cải thiện hiệu quả học bán giám sát.

Nhìn chung, các nhóm mô hình so sánh được thiết kế theo hướng tăng dần mức độ phức tạp của phương pháp: từ mô hình học máy truyền thống, sang học trên đồ thị, sau đó mở rộng với metadata lâm sàng và cuối cùng là học bán giám sát có chọn lọc dựa trên entropy. Thiết lập này giúp đánh giá riêng ảnh hưởng của từng thành phần trong phương pháp đề xuất đối với hiệu quả phân loại ảnh tổn thương da.

3.5.6. Thiết lập môi trường thực nghiệm

Các thực nghiệm trong nghiên cứu được triển khai bằng ngôn ngữ Python trên nền tảng PyTorch và PyTorch Geometric. Các kiến trúc nền CNN pretrained được sử dụng

thông qua thư viện TensorFlow/Keras và torchvision. Các mô hình học máy truyền thống được xây dựng bằng thư viện Scikit-learn, trong khi XGBoost được triển khai bằng thư viện XGBoost chính thức.

Toàn bộ thực nghiệm được thực hiện trên máy trạm có GPU NVIDIA hỗ trợ CUDA nhằm tăng tốc quá trình huấn luyện các mô hình học sâu và mô hình GNN. Hệ điều hành sử dụng là Ubuntu Linux, kết hợp với môi trường Python và các thư viện tính toán khoa học phổ biến như NumPy, Pandas và Matplotlib.

Bảng 3.6 trình bày cấu hình môi trường thực nghiệm được sử dụng trong nghiên cứu.

Bảng 3.6: Cấu hình môi trường thực nghiệm

Thành phần	Cấu hình
Hệ điều hành	Ubuntu Linux
Ngôn ngữ lập trình	Python
Framework học sâu	PyTorch, TensorFlow/Keras
Framework đồ thị	PyTorch Geometric
Thư viện học máy	Scikit-learn, XGBoost
GPU	NVIDIA CUDA-enabled GPU
CUDA	CUDA Toolkit

3.5.7. Thiết lập tham số mô hình

Sau bước hợp nhất đặc trưng đa nguồn, mỗi mẫu dữ liệu được biểu diễn bởi vector hợp nhất $\mathbf{h}_i$. Các vector này được sử dụng để xây dựng đồ thị dữ liệu cho mô hình GNN. Độ tương đồng giữa các nút được tính bằng cosine similarity, sau đó đồ thị được xây dựng theo chiến lược k -nearest neighbors với $k = 5$. Mỗi nút được kết nối với năm nút có độ tương đồng cosine cao nhất trong không gian biểu diễn hợp nhất.

Việc xây dựng cạnh hoàn toàn không sử dụng nhãn lớp nhằm tránh rò rỉ thông tin trong quá trình học đồ thị. Nhờ đó, cấu trúc đồ thị phản ánh quan hệ tự nhiên giữa các mẫu dữ liệu thay vì phụ thuộc vào nhãn phân loại.

Bảng 3.7 trình bày cấu hình xây dựng đồ thị được sử dụng trong nghiên cứu.

Bảng 3.7: Cấu hình xây dựng đồ thị

Tham số	Giá trị
Độ đo tương đồng	Cosine similarity
Chiến lược xây dựng đồ thị	k -nearest neighbors
Số láng giềng k	5
Loại đồ thị	Đồ thị vô hướng
Sử dụng nhãn khi tạo cạnh	Không
Biểu diễn nút	Vector hợp nhất $\mathbf{h}_i$

Các kiến trúc nền GNN được huấn luyện với cùng cấu hình nhằm bảo đảm tính công bằng trong so sánh thực nghiệm. Tất cả các mô hình đều sử dụng hai lớp GNN với hidden dimension bằng 128 và hàm kích hoạt ReLU. Dropout được đặt bằng 0.5 nhằm

giảm hiện tượng quá khớp trong quá trình huấn luyện.

Quá trình tối ưu được thực hiện bằng Adam optimizer với learning rate ban đầu bằng 10^{-3} . Batch size được đặt bằng 64 và số epoch tối đa là 150. Ngoài ra, cơ chế early stopping với patience bằng 10 epoch được sử dụng nhằm tránh huấn luyện quá mức khi validation loss không còn cải thiện. Cơ chế ReduceLRonPlateau cũng được áp dụng để tự động giảm learning rate trong các giai đoạn hội tụ chậm.

Bảng 3.8 trình bày các tham số huấn luyện cho các kiến trúc nền GNN.

Bảng 3.8: Các tham số huấn luyện của mô hình GNN

Tham số	Giá trị
Số lớp GNN	2
Hidden dimension	128
Hàm kích hoạt	ReLU
Dropout	0.5
Optimizer	Adam
Learning rate	10^{-3}
Batch size	64
Số epoch tối đa	150
Early stopping patience	10
Learning rate scheduler	ReduceLRonPlateau

Đối với các mô hình học máy truyền thống, các tham số được thiết lập theo cấu hình phổ biến thường được sử dụng trong các nghiên cứu phân loại ảnh y sinh. Logistic Regression sử dụng bộ tối ưu LBFGS, SVM sử dụng kernel RBF, Random Forest sử dụng tập hợp nhiều cây quyết định, trong khi XGBoost sử dụng boosting tuần tự nhằm tăng khả năng học quan hệ phi tuyến giữa các đặc trưng.

Bảng 3.9 trình bày các tham số của các mô hình học máy truyền thống.

Bảng 3.9: Các tham số của mô hình học máy truyền thống

Mô hình	Tham số chính	Giá trị
Logistic Regression	Solver	LBFGS
SVM	Kernel	RBF
Random Forest	Number of trees	100
XGBoost	Number of estimators	100
XGBoost	Learning rate	0.1

3.6. Kết quả thực nghiệm

3.6.1. Kết quả của pha huấn luyện có giám sát

Pha huấn luyện có giám sát được thực hiện nhằm đánh giá khả năng học biểu diễn của các mô hình khi chỉ sử dụng dữ liệu có nhãn. Các cấu hình được khảo sát gồm nhóm mô hình học máy truyền thống sử dụng đặc trưng hợp nhất và nhóm mô hình học trên đồ thị sử dụng biểu diễn đa nguồn. Do bộ dữ liệu ISIC 2020 có mức mất cân bằng lớp, việc

đánh giá tập trung vào các chỉ số cân bằng lớp và có ý nghĩa lâm sàng, gồm Sensitivity, Specificity, Macro F1 và Macro AUC.

Kết quả định lượng của các mô hình ở pha huấn luyện có giám sát được tổng hợp trong Bảng 3.10.

Bảng 3.10: So sánh kết quả pha huấn luyện có giám sát giữa các mô hình học máy truyền thống và các kiến trúc nền GNN trên bộ dữ liệu ISIC 2020. Các giá trị được báo cáo dưới dạng trung bình $\pm$ độ lệch chuẩn trên các lần kiểm chứng lặp phân tầng.

Mô hình	Sensitivity (%)	Specificity (%)	Macro F1 (%)	Macro AUC (%)
Nhóm học máy truyền thống				
Logistic Regression	53.8 ± 2.9	99.1 ± 0.1	75.6 ± 1.2	92.6 ± 1.1
SVM	39.3 ± 3.6	99.6 ± 0.1	74.4 ± 2.4	95.4 ± 0.7
Random Forest	3.0 ± 1.7	100.0 ± 0.0	52.4 ± 1.6	95.4 ± 0.8
XGBoost	51.3 ± 3.7	99.7 ± 0.1	80.6 ± 1.6	96.0 ± 0.8
Nhóm GNN có tích hợp metadata				
GCN	56.7 ± 20.4	70.8 ± 8.4	44.2 ± 2.5	71.2 ± 3.6
GAT	44.2 ± 13.9	77.4 ± 4.8	46.5 ± 1.3	72.2 ± 3.2
GraphSAGE	57.8 ± 18.9	70.7 ± 6.7	44.3 ± 1.9	73.7 ± 5.5
GINE	76.0 ± 22.7	61.2 ± 13.7	40.8 ± 5.0	75.9 ± 5.2
GraphSAGE–GAT	70.2 ± 17.7	65.8 ± 6.9	42.8 ± 2.1	74.8 ± 5.7

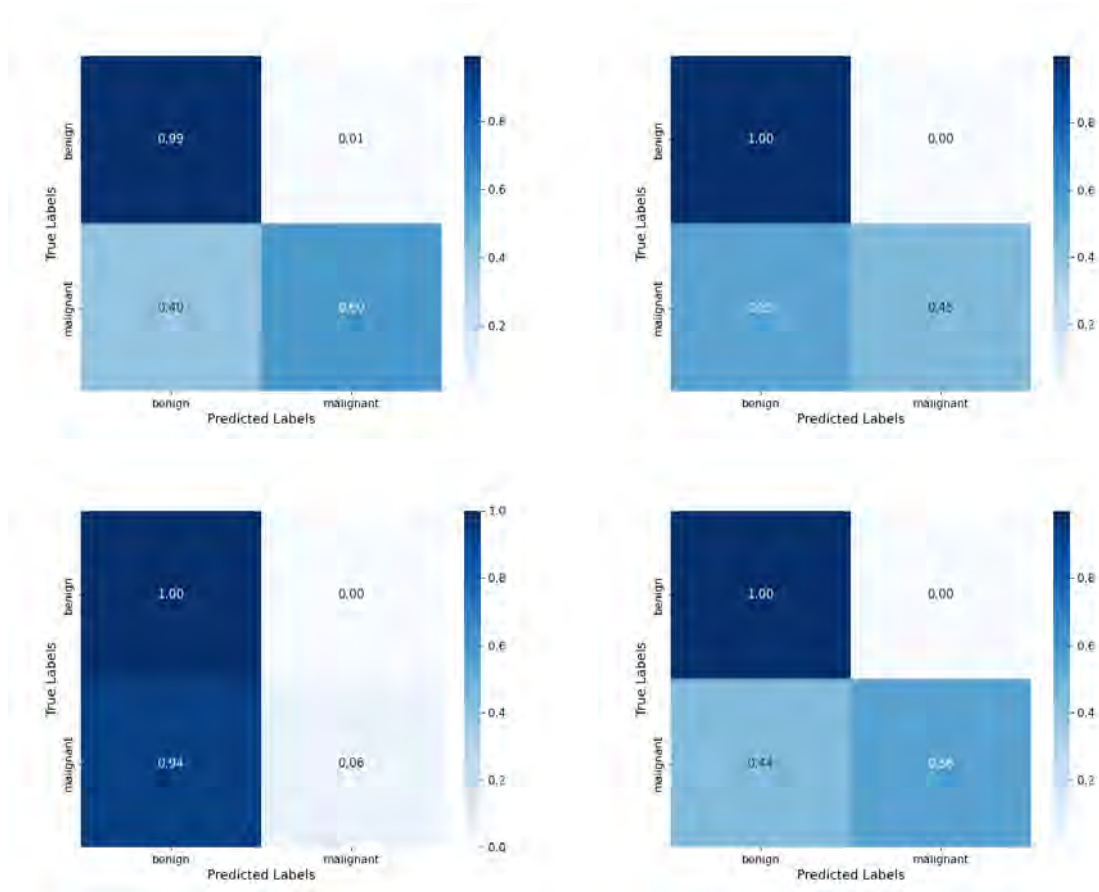
Bảng 3.10 cho thấy các mô hình học máy truyền thống đạt Specificity cao, đặc biệt là Random Forest và XGBoost. Điều này phản ánh khả năng nhận diện lớp lành tính tốt. Tuy nhiên, Sensitivity của nhóm mô hình này lại thấp hơn, nhất là ở Random Forest và SVM, cho thấy tỷ lệ bỏ sót lớp ác tính còn lớn. Trong bối cảnh dữ liệu lệch lớp, sự khác biệt giữa Specificity và Sensitivity cho thấy độ chính xác tổng thể cao chưa phản ánh đầy đủ chất lượng phân loại.

Ngược lại, các kiến trúc nền GNN cho thấy hành vi dự đoán cân bằng hơn giữa hai lớp. Mặc dù Macro F1 và Macro AUC của nhóm GNN ở pha đầu chưa vượt nhóm học máy truyền thống, nhưng một số kiến trúc như GINE và GraphSAGE–GAT đạt Sensitivity cao hơn. Điều này cho thấy việc khai thác quan hệ liên mẫu trên đồ thị có tác dụng đối với khả năng nhận diện lớp ác tính, vốn là lớp thiểu số nhưng có ý nghĩa lâm sàng.

Để làm rõ hơn hành vi dự đoán của từng nhóm phương pháp, các ma trận nhầm lẫn chuẩn hóa của nhóm mô hình học máy truyền thống và nhóm GNN được trình bày lần lượt trong Hình 3.19 và Hình 3.20.

Các kết quả trong Hình 3.19 cho thấy nhóm học máy truyền thống có xu hướng thiên lệch về lớp lành tính. Điều này đặc biệt rõ ở Random Forest, khi gần như toàn bộ mẫu được gán về lớp chiếm ưu thế. Trong khi đó, Hình 3.20 cho thấy các kiến trúc nền GNN giảm hiện tượng này, nhất là ở GINE và GraphSAGE–GAT, nhờ khai thác thêm ngữ cảnh quan hệ giữa các mẫu trong không gian đồ thị.

Từ các kết quả trên có thể rút ra ba nhận xét. Thứ nhất, các mô hình học máy truyền thống vẫn là đối sánh nếu chỉ xét độ chính xác tổng thể và độ đặc hiệu. Thứ hai, các kiến trúc nền GNN cho thấy ưu thế rõ hơn về độ nhạy, qua đó phù hợp hơn với mục tiêu phát hiện lớp ác tính trong bối cảnh dữ liệu lệch lớp. Thứ ba, trong số các kiến trúc GNN



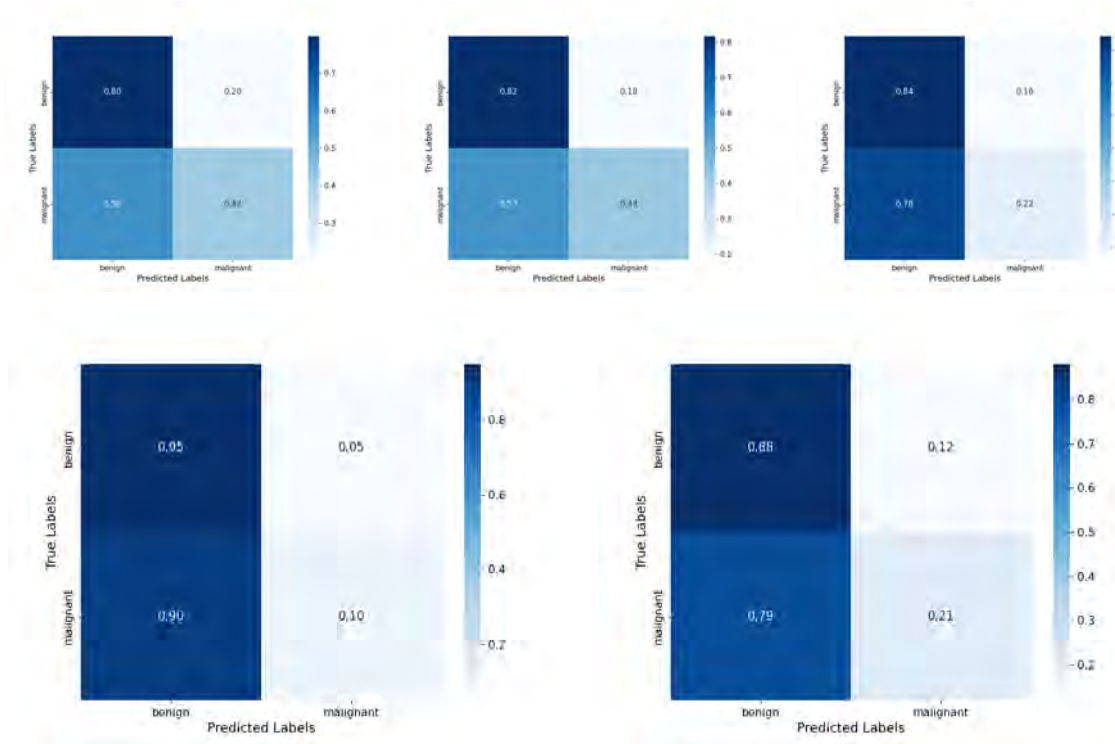
Hình 3.19: Ma trận nhầm lẫn chuẩn hóa của các mô hình học máy truyền thống ở pha huấn luyện có giám sát. Các mô hình này có Specificity cao nhưng còn bỏ sót mẫu ác tính ở một số cấu hình.

được khảo sát, GINE và GraphSAGE–GAT tạo ra điểm làm việc cân bằng hơn giữa các chỉ số chính, do đó thích hợp để tiếp tục xem xét trong pha gán nhãn giả dựa trên độ tin cậy.

Các đường cong trong Hình 3.21–3.24 cho thấy sự khác biệt về động học hội tụ giữa các kiến trúc nền GNN. Trên tập huấn luyện, GAT đạt độ chính xác cao hơn và duy trì ở mức ổn định sau giai đoạn đầu, nhưng đồng thời hàm mất mát trên tập xác thực lại dừng ở mức cao hơn so với các mô hình còn lại. Điều này cho thấy biểu diễn học được của GAT phù hợp tốt với dữ liệu huấn luyện nhưng khả năng khái quát hóa chưa thật sự cao.

GCN và GraphSAGE có xu hướng hội tụ sớm và duy trì quỹ đạo tương đối mượt trên cả tập huấn luyện và tập xác thực, nhưng mức hiệu năng xác thực cuối cùng không cao. GraphSAGE–GAT thể hiện khả năng cải thiện nhanh ở giai đoạn đầu, song các đường cong vẫn còn dao động trước khi ổn định.

Trong số các mô hình được khảo sát, GINE cho thấy đặc tính ổn định hơn về mặt tối ưu. Mặc dù độ chính xác trên tập huấn luyện không cao nhất, hàm mất mát trên cả tập huấn luyện và tập xác thực giảm đều hơn và đạt mức thấp nhất trong các kiến trúc nền được xét. Đồng thời, đường cong độ chính xác trên tập xác thực của GINE ít dao động hơn ở giai đoạn sau huấn luyện và duy trì xu hướng cải thiện ổn định. Đặc điểm này cho thấy GINE tạo ra một quá trình học cân bằng hơn giữa khả năng khớp dữ liệu



Hình 3.20: Ma trận nhầm lẫn chuẩn hóa của các kiến trúc nền GNN có tích hợp meta-data ở pha huấn luyện có giám sát. Một số mô hình GNN làm thay đổi cân bằng giữa khả năng nhận diện lớp lành tính và ác tính.

huấn luyện và khả năng khái quát hóa.

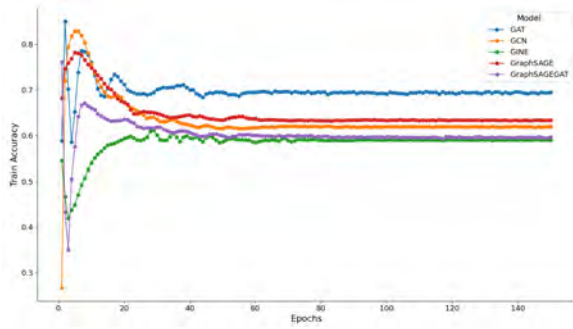
Kết hợp với các kết quả định lượng trong Bảng 3.10 và phân tích ma trận nhầm lẫn ở Hình 3.20, GINE được lựa chọn là kiến trúc nền ổn định nhất ở pha huấn luyện có giám sát. Việc lựa chọn này không dựa riêng trên một chỉ số đơn lẻ, mà dựa trên đồng thời ba căn cứ: khả năng phát hiện lớp ác tính ở mức cao, quỹ đạo tối ưu ổn định trên tập xác thực, và mức hàm mất mát cuối cùng thấp hơn so với các kiến trúc nền còn lại. Trên cơ sở đó, GINE được sử dụng làm mô hình nền cho pha gán nhãn giả dựa trên độ tin cậy ở bước tiếp theo.

3.6.2. Kết quả của pha gán nhãn giả dựa trên độ tin cậy

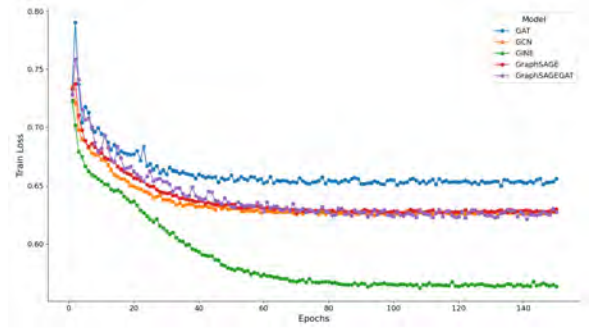
Sau pha huấn luyện có giám sát, mô hình GINE được lựa chọn làm kiến trúc nền cho bước bán giám sát dựa trên các kết quả ở pha một. Trên cơ sở mô hình nền này, pha thứ hai được thực hiện bằng cách ước lượng entropy dự đoán trên tập mẫu chưa gán nhãn, chọn lọc các mẫu có độ tin cậy cao theo Công thức (3.27) và tạo tập huấn luyện mở rộng theo Công thức (3.29), sau đó huấn luyện lại mô hình trên tập dữ liệu này.

Kết quả định lượng trước và sau pha gán nhãn giả được tổng hợp trong Bảng 3.11. Bảng này cho phép đánh giá trực tiếp mức cải thiện của mô hình khi chuyển từ huấn luyện có giám sát sang huấn luyện bán giám sát có kiểm soát bằng độ tin cậy.

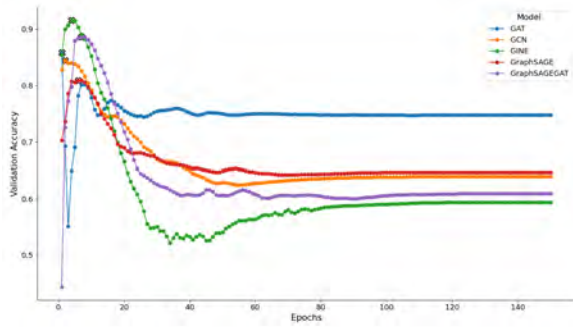
Bảng 3.11 cho thấy pha gán nhãn giả dựa trên độ tin cậy mang lại cải thiện nhất quán trên hầu hết các chỉ số chính. Trong đó, Accuracy tăng từ 78.5% lên 89.2%, Macro AUC tăng từ 72.6% lên 88.5%, còn Macro F1 tăng từ 43.0% lên 60.9%. Đặc biệt, Macro Recall và Recall đều tăng, cho thấy mô hình sau tinh chỉnh có khả năng nhận diện lớp ác tính tốt hơn so với pha huấn luyện thuần giám sát. Đây là điểm có ý nghĩa thực tế, vì



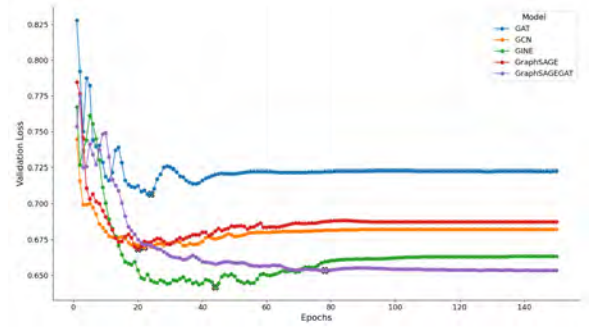
Hình 3.21: Diễn biến độ chính xác trên tập huấn luyện của các kiến trúc nền GNN ở pha huấn luyện có giám sát.



Hình 3.22: Diễn biến hàm mất mát trên tập huấn luyện của các kiến trúc nền GNN ở pha huấn luyện có giám sát.



Hình 3.23: Diễn biến độ chính xác trên tập xác thực của các kiến trúc nền GNN ở pha huấn luyện có giám sát.



Hình 3.24: Diễn biến hàm mất mát trên tập xác thực của các kiến trúc nền GNN ở pha huấn luyện có giám sát.

trong bối cảnh dữ liệu lệch lớp, việc giảm bỏ sót các trường hợp ác tính có giá trị hơn so với việc chỉ gia tăng độ chính xác tổng thể.

Một điểm cần lưu ý là Precision giảm nhẹ sau pha hai. Tuy nhiên, sự suy giảm này đi kèm với mức tăng của Recall, Macro Recall và Macro F1. Điều đó cho thấy pha bán giám sát điều chỉnh biên quyết định theo hướng tăng Recall cho lớp thiểu số, đồng thời làm Precision giảm nhẹ. Trong bối cảnh sàng lọc, sự đánh đổi giữa Precision và Recall cần được diễn giải cùng mục tiêu sử dụng cụ thể.

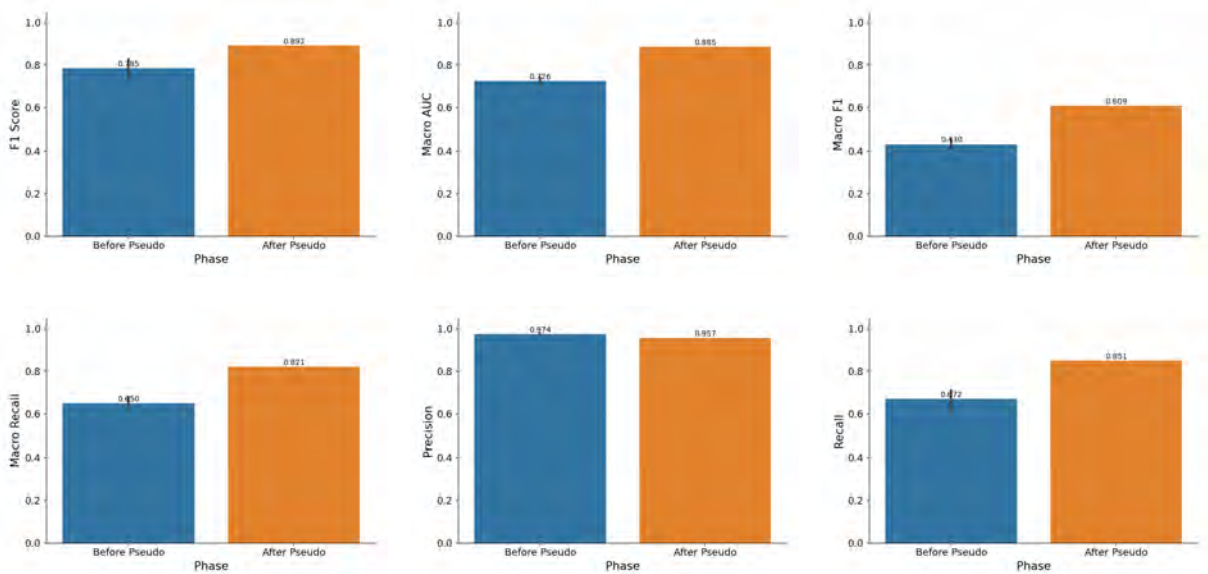
Để quan sát trực quan mức cải thiện của từng chỉ số, các biểu đồ cột trước và sau pha gán nhãn giả được trình bày trong Hình 3.25. Các hình này cho thấy xu hướng cải thiện không chỉ xuất hiện trên một chỉ số đơn lẻ mà được duy trì trên nhiều chỉ số cân bằng lớp khác nhau.

Các biểu đồ trong Hình 3.25 xác nhận rằng pha bán giám sát không chỉ cải thiện Accuracy, mà còn cải thiện đồng thời các chỉ số phản ánh chất lượng phân loại cân bằng theo lớp, đặc biệt là Macro F1, Macro Recall và Macro AUC. Vì vậy, mức cải thiện đạt được không thể xem là hệ quả của việc mô hình thích nghi tốt hơn với lớp đa số, mà phản ánh một sự cải thiện về năng lực phân biệt của mô hình sau khi khai thác thêm các mẫu chưa gán nhãn có độ tin cậy cao.

Tác động của pha gán nhãn giả cũng được phản ánh qua động học huấn luyện. Hình 3.26 trình bày đường cong Accuracy và Loss của mô hình GINE trước và sau khi áp dụng cơ chế gán nhãn có lọc theo độ tin cậy.

Bảng 3.11: So sánh kết quả của mô hình GINE trước và sau pha gán nhãn giả dựa trên độ tin cậy. Các giá trị được báo cáo dưới dạng trung bình $\pm$ độ lệch chuẩn trên các lần kiểm chứng.

Chỉ số	Pha 1 (%)	Pha 2 (%)
Accuracy	78.5 ± 1.8	89.2 ± 1.1
Macro AUC	72.6 ± 2.1	88.5 ± 1.4
Macro F1	43.0 ± 1.9	60.9 ± 1.7
Macro Recall	65.0 ± 2.3	82.1 ± 1.6
Precision	97.4 ± 0.6	95.7 ± 0.7
Recall	67.2 ± 2.5	85.1 ± 1.8

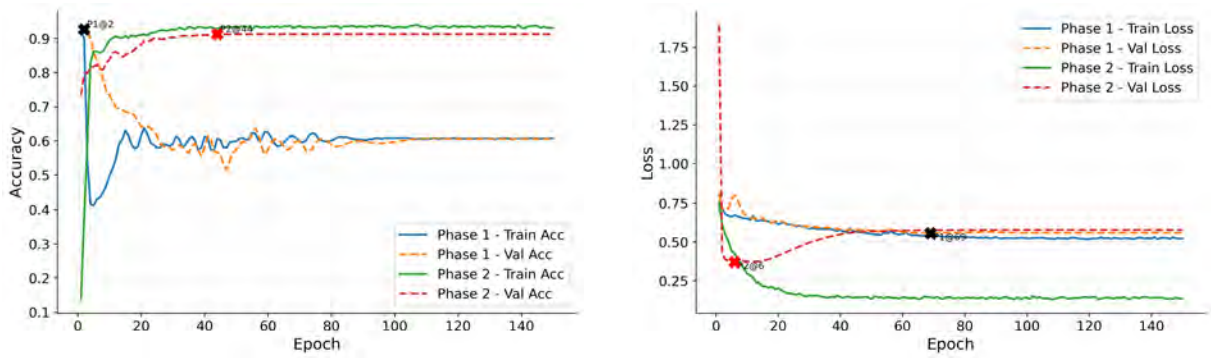


Hình 3.25: So sánh các chỉ số phân loại của mô hình GINE trước và sau pha gán nhãn giả dựa trên độ tin cậy.

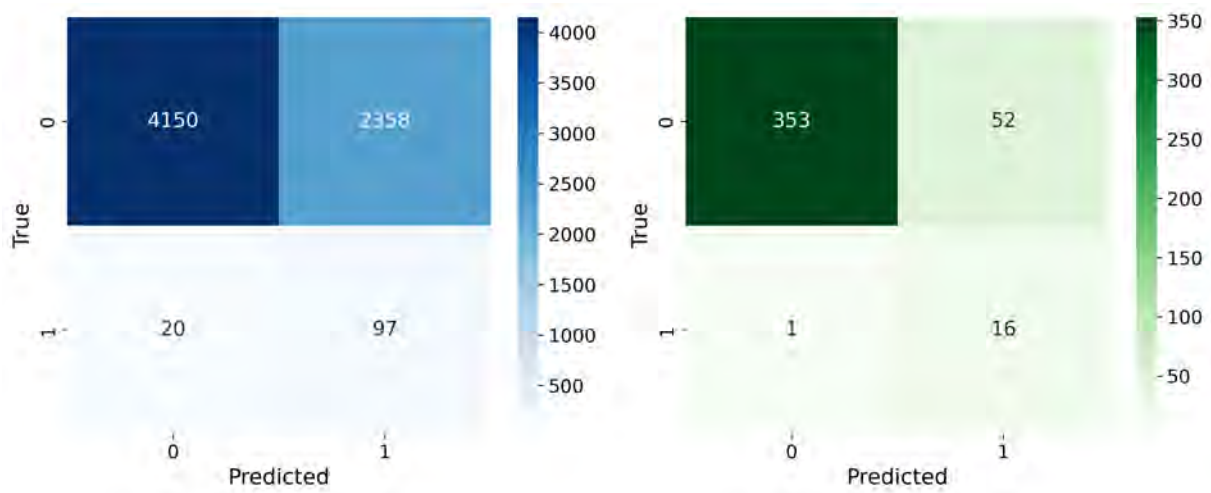
Hình 3.26 cho thấy sau khi bổ sung các mẫu giả nhãn có độ tin cậy cao, mô hình đạt mức Accuracy cao hơn và đồng thời giảm được Loss ở cả quá trình huấn luyện lẫn xác thực. Điều này cho thấy dữ liệu giả nhãn được chọn không làm khuếch đại nhiễu, mà ngược lại góp phần làm giàu không gian biểu diễn và giúp tối ưu hóa bộ phân loại theo hướng ổn định hơn. Nói cách khác, cơ chế entropy không chỉ đóng vai trò lọc mẫu, mà còn là điều kiện cần để bảo đảm rằng thông tin bổ sung từ dữ liệu chưa gán nhãn không làm lệch quá trình học.

Sự thay đổi về biên quyết định của mô hình sau pha hai được thể hiện rõ hơn qua Hình 3.27, trong đó trình bày ma trận nhầm lẫn trước và sau khi tinh chỉnh bằng dữ liệu giả nhãn.

Hình 3.27 cho thấy sau pha hai, số lượng mẫu ác tính bị phân loại nhầm sang lành tính giảm đi. Điều này nhất quán với mức tăng của Recall và Macro Recall trong Bảng 3.11. Mặc dù số lượng dự đoán dương tính có thể tăng nhẹ, sự thay đổi này giúp mô hình đạt được một điểm làm việc phù hợp hơn với yêu cầu phát hiện bệnh trong thực tế.



Hình 3.26: Diễn biến độ chính xác (trái) và hàm mất mát (phải) của mô hình GINE ở pha huấn luyện có giám sát và pha gán nhãn giả dựa trên độ tin cậy.



Hình 3.27: Ma trận nhầm lẫn của mô hình GINE trước và sau pha gán nhãn giả dựa trên độ tin cậy. Pha hai làm thay đổi số lượng dự đoán sai ở lớp thiểu số và biên quyết định giữa hai lớp.

Để đánh giá mức cải thiện có ổn định trên các lần chia dữ liệu hay không, kết quả kiểm định cặp giữa mô hình GINE ở pha một và mô hình GINE sau tinh chỉnh ở pha hai được trình bày trong Bảng 3.12.

Bảng 3.12 cho thấy các giá trị p đều rất nhỏ so với ngưỡng thông thường, cho thấy mức cải thiện của pha hai là có ý nghĩa thống kê. Kết quả này củng cố thêm nhận định rằng hiệu quả của cơ chế gán nhãn không xuất phát từ dao động ngẫu nhiên giữa các fold, mà phản ánh sự cải thiện ổn định của mô hình khi khai thác thêm dữ liệu chưa gán nhãn theo cách có kiểm soát.

Một điểm cần lưu ý là pha hai không sử dụng nhãn thật của tập chưa gán nhãn trong quá trình tạo giả nhãn. Toàn bộ quy trình lựa chọn mẫu đều dựa trên entropy của phân phối xác suất đầu ra. Vì vậy, phần cải thiện quan sát được không bắt nguồn từ rò rỉ thông tin, mà từ việc mô hình tận dụng được một tập con các mẫu chưa gán nhãn có mức độ tin cậy cao để điều chỉnh lại không gian quyết định.

Các kết quả trên cho thấy cơ chế gán nhãn giả dựa trên độ tin cậy là thành phần có ảnh hưởng đến kết quả pha hai. Đóng góp của pha này không chỉ nằm ở việc tăng số lượng mẫu huấn luyện, mà còn ở việc giới hạn tập bổ sung vào các mẫu có entropy thấp.

Bảng 3.12: Kết quả kiểm định cặp giữa mô hình GINE ở pha huấn luyện có giám sát và mô hình GINE sau pha gán nhãn giả dựa trên độ tin cậy trên 15 lần kiểm chứng. Hiệu trung bình được tính theo dạng (Pha 2 – Pha 1).

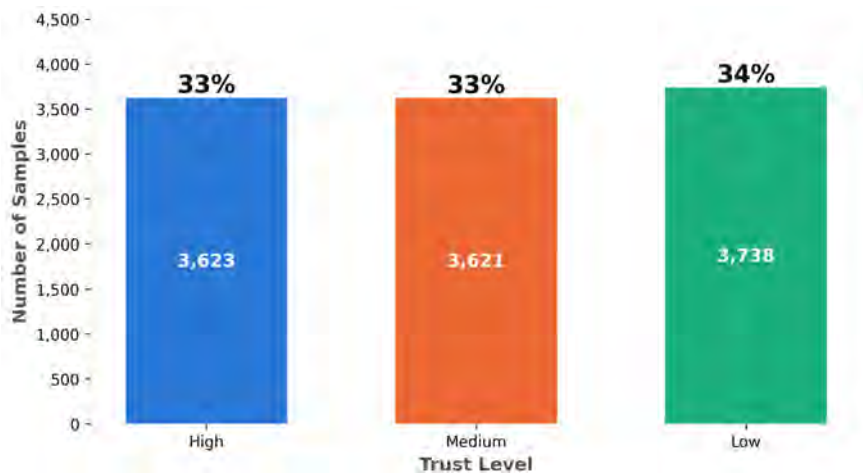
Chỉ số	Mean (P1) (%)	Mean (P2) (%)	Hiệu	t	p
Macro F1	40.1	60.9	0.2076	17.9804	4.53×10^{-11}
Macro Recall	69.3	82.1	0.1283	7.6602	2.26×10^{-6}
Macro AUC	75.3	88.5	0.1321	10.5035	5.07×10^{-8}

3.6.3. Kết quả phân tích theo mức độ tin cậy

Bên cạnh các chỉ số đánh giá toàn cục, mô hình được phân tích theo mức độ tin cậy của dự đoán nhằm xem xét mối liên hệ giữa entropy đầu ra và lỗi phân loại. Trong khung phương pháp, mỗi dự đoán được chia thành ba nhóm: *High Trust*, *Medium Trust* và *Low Trust*. Cách phân tầng này cho phép đánh giá kết quả theo từng nhóm mức độ chắc chắn của dự đoán.

Phân tích theo mức độ tin cậy được thực hiện trên toàn bộ tập dữ liệu chưa gán nhãn D_U (ISIC 2020 Test) gồm 10,982 mẫu, đúng như phân bố dữ liệu đã trình bày trong Bảng 3.1. Việc phân tầng sử dụng hai ngưỡng entropy α_1 và α_2 theo ánh xạ ở Công thức 3.26, do đó mỗi mẫu trong D_U được xếp vào đúng một nhóm và tổng số mẫu của ba nhóm bằng đúng kích thước của D_U .

Kết quả phân bố số lượng mẫu theo ba mức độ tin cậy được thể hiện trong Hình 3.28 và Bảng 3.13. Số lượng mẫu trong các nhóm *High Trust*, *Medium Trust* và *Low Trust* lần lượt là 3,623, 3,621 và 3,738 mẫu, tương ứng với tỷ lệ 33.0%, 33.0% và 34.0% của tập D_U . Phân bố này cho thấy hai ngưỡng entropy được lựa chọn tạo ra ba nhóm mẫu có quy mô gần nhau trong thiết lập thực nghiệm hiện tại; nói cách khác, entropy dự đoán trên D_U trải đều trên miền giá trị thay vì tập trung về một phía.



Hình 3.28: Phân bố số lượng mẫu theo ba mức độ tin cậy trên tập chưa gán nhãn D_U (ISIC 2020 Test, 10,982 mẫu).

Nhóm *High Trust* được sử dụng để tạo tập nhãn giả trong pha bán giám sát, còn hai nhóm *Medium Trust* và *Low Trust* không được sử dụng trong bước tái huấn luyện. Cách xử lý này nhằm giảm khả năng đưa các mẫu có dự đoán kém chắc chắn vào tập huấn

Bảng 3.13: Phân bố số lượng mẫu theo mức độ tin cậy trên tập chưa gán nhãn D_U

Mức độ tin cậy	Số lượng mẫu	Tỷ lệ (%)	Sử dụng cho tái huấn luyện
<i>High Trust</i>	3,623	33.0	Có
<i>Medium Trust</i>	3,621	33.0	Không
<i>Low Trust</i>	3,738	34.0	Không
Tổng cộng	10,982	100.0	–

luyện mở rộng.

Theo phân bố trong Bảng 3.13, chỉ 3,623 mẫu thuộc nhóm *High Trust* được gán nhãn giả và bổ sung vào tập huấn luyện mở rộng $D_L \cup D_{\text{pseudo}}^{\text{High}}$ theo Công thức 3.29, tương ứng khoảng 33.0% dữ liệu chưa gán nhãn. Phần còn lại gồm 7,359 mẫu thuộc hai nhóm *Medium Trust* và *Low Trust* không tham gia bước tái huấn luyện. Tỷ lệ chọn lọc này cho thấy cơ chế lọc theo entropy loại bỏ khoảng hai phần ba dữ liệu chưa gán nhãn, đổi mức tăng kích thước tập huấn luyện lấy độ tin cậy của nhãn giả.

Phân tích theo mức độ tin cậy cho thấy nhóm *High Trust* có tỷ lệ dự đoán đúng cao hơn nhóm *Low Trust*, trong khi nhiều trường hợp nhầm lẫn tập trung ở nhóm có entropy cao. Kết quả này cho thấy entropy có thể dùng làm tiêu chí lọc mẫu trong phạm vi thiết lập hiện tại. Tuy nhiên, các dự đoán thuộc nhóm *High Trust* chỉ nên được hiểu là dự đoán mà mô hình đưa ra với mức độ chắc chắn cao hơn, không phải kết luận chắc chắn.

Cần lưu ý thêm rằng tập D_U không có nhãn công khai, nên phân bố trong Bảng 3.13 chỉ phản ánh mức độ chắc chắn trong dự đoán của mô hình, không phải phân bố lớp thật của dữ liệu.

3.7. Phân tích và thảo luận kết quả thực nghiệm

Các kết quả thực nghiệm được phân tích theo bốn khía cạnh: vai trò của biểu diễn đa nguồn, ảnh hưởng của học trên đồ thị, tác động của gán nhãn giả có chọn lọc và phạm vi diễn giải của mô hình.

Thứ nhất, các mô hình học máy truyền thống đạt Specificity cao nhưng Sensitivity thấp ở một số cấu hình. Kết quả này phản ánh ảnh hưởng của mất cân bằng lớp trong ISIC 2020: mô hình có thể đạt kết quả tốt trên lớp lành tính nhưng vẫn bỏ sót các mẫu ác tính. Vì vậy, việc đánh giá cần dựa trên các chỉ số cân bằng lớp như Macro F1, Macro Recall, Macro AUC, Sensitivity và Specificity, thay vì chỉ dựa trên Accuracy.

Thứ hai, các mô hình GNN có tích hợp metadata làm thay đổi cân bằng giữa Sensitivity và Specificity. Một số kiến trúc nền GNN tăng Sensitivity cho lớp ác tính, nhưng Macro F1 và Macro AUC ở pha có giám sát chưa vượt nhóm học máy truyền thống. Kết quả này cho thấy metadata và quan hệ đồ thị có vai trò hỗ trợ, nhưng hiệu quả phụ thuộc vào chất lượng metadata, cách xây dựng đồ thị và mức độ phù hợp của cấu trúc lân cận. Metadata đơn lẻ hoặc metadata có thiếu dữ liệu không đủ thay thế đặc trưng ảnh. Do đó, kết quả của nhóm GNN có metadata trong Bảng 3.10 cần được hiểu là đánh giá vai trò hỗ trợ của metadata và quan hệ đồ thị, không phải bằng chứng rằng metadata riêng lẻ luôn cải thiện mọi chỉ số so với mô hình học máy truyền thống.

Thứ ba, pha gán nhãn giả dựa trên entropy làm tăng một số chỉ số sau khi tái huấn luyện, đặc biệt là Macro F1, Macro Recall và Macro AUC. Tuy nhiên, tập ISIC 2020

Test không có nhãn công khai nên các nhãn giả trên tập này không được xem là nhãn thật. Hiệu quả của cơ chế gán nhãn giả được diễn giải thông qua kết quả kiểm chứng trên các phần dữ liệu có nhãn và qua sự thay đổi của các chỉ số trước–sau tái huấn luyện.

Thứ tư, entropy được sử dụng như tiêu chí ước lượng độ bất định của dự đoán. Nhóm mẫu có entropy thấp được chọn cho gán nhãn giả, còn nhóm entropy cao được loại khỏi bước tái huấn luyện. Cơ chế này giúp giảm nguy cơ đưa các mẫu có dự đoán không ổn định vào tập huấn luyện mở rộng. Tuy vậy, entropy chưa thay thế các đánh giá calibration đầy đủ như ECE, Brier score hoặc calibration curve; đây là hướng cần kiểm chứng thêm.

Kết quả so sánh với một số nghiên cứu liên quan được trình bày trong Bảng 3.14.

Bảng 3.14: So sánh mô hình đề xuất với một số phương pháp liên quan trên ISIC 2020 theo chỉ số Accuracy

Phương pháp	Hướng tiếp cận	Accuracy (%)
Gupta và cộng sự [31]	Knowledge distillation	87.3
Alhudhaif và cộng sự [130]	Mạng CNN sâu cho phát hiện và phân đoạn tổn thương da	85.5
Choudhary và cộng sự [131]	Mask-RCNN kết hợp GRU	88.0
Rana và cộng sự [132]	Mô hình CNN xếp chồng	86.7
Mô hình đề xuất	Hợp nhất đa nguồn + học trên đồ thị + gán nhãn giả dựa trên độ tin cậy	89.2

Bảng 3.14 cho thấy mô hình đề xuất đạt Accuracy cạnh tranh với một số phương pháp đã công bố trên ISIC 2020. Tuy nhiên, so sánh này cần được diễn giải thận trọng do khác biệt về cách chia dữ liệu, tiền xử lý, kiến trúc nền, thiết lập huấn luyện và chỉ số báo cáo. Đóng góp chính của chương nằm ở việc kết hợp biểu diễn đa nguồn, đồ thị k-NN và gán nhãn giả có kiểm soát bằng entropy, thay vì khẳng định ưu thế tuyệt đối so với mọi kiến trúc hiện đại.

Phương pháp còn một số hạn chế. Thứ nhất, hiệu quả của mô hình phụ thuộc vào chất lượng đồ thị k-NN; nếu cạnh nổi các mẫu không cùng ngữ nghĩa, quá trình lan truyền thông tin có thể làm giảm hiệu quả phân loại. Thứ hai, metadata trong bộ dữ liệu công khai có thể thiếu hoặc không đồng nhất, nên vai trò của metadata cần được xem là bổ trợ. Thứ ba, luận án chưa thực hiện đánh giá riêng theo từng mức nhiễu ảnh, ví dụ nhiễu do ánh sáng, nền ảnh, độ nét hoặc sai khác thiết bị thu nhận. Thứ tư, các kiến trúc nền hiện đại như EfficientNet, EfficientNetV2, ViT hoặc các mô hình thị giác nền tảng chưa được khảo sát đầy đủ trong chương này. Thứ năm, mô hình mới được đánh giá trong phạm vi dữ liệu ISIC 2020 và chưa có kiểm chứng lâm sàng đa trung tâm.

Do đó, kết quả của chương này nên được hiểu trong phạm vi nghiên cứu mô hình và đánh giá thực nghiệm trên bộ dữ liệu công khai. Mô hình có thể hỗ trợ nghiên cứu và sàng lọc, nhưng chưa phải công cụ chẩn đoán độc lập nếu chưa có đánh giá lâm sàng, ngoại kiểm trên dữ liệu độc lập và xem xét các yêu cầu đạo đức liên quan đến sử dụng AI trong y tế.

Kết luận chương

Chương này mở rộng hướng nghiên cứu của Chương 2 từ học biểu diễn đặc trưng sang học quan hệ dữ liệu bằng đồ thị cho bài toán phân loại tổn thương da. Sự liên hệ giữa hai chương nằm ở định hướng khai thác thông tin thị giác bổ trợ và entropy, không phải ở việc Chương 3 tái sử dụng biểu diễn BlockNet. Cụ thể, Chương 2 chọn lọc bản đồ đặc trưng theo entropy, còn Chương 3 sử dụng các vector đầu ra sau pooling của VGG19, DenseNet121 và MobileNet được lựa chọn bằng Chi-square, kết hợp với histogram RGB, histogram HSV, đặc trưng fractal và metadata. Entropy trong Chương 3 chỉ được tính trên phân phối xác suất dự đoán để phân tầng độ tin cậy và chọn mẫu gán nhãn giả. Trên biểu diễn riêng này, Chương 3 khai thác quan hệ tương đồng giữa các mẫu thông qua mô hình mạng nơ-ron đồ thị.

Đóng góp chính của chương này tương ứng với công trình [CT4], trong đó mỗi mẫu dữ liệu được biểu diễn đồng thời bởi đặc trưng học sâu, đặc trưng thủ công và metadata lâm sàng. Các biểu diễn đa nguồn được hợp nhất để xây dựng biểu diễn nút cho đồ thị tương đồng k-NN, sau đó được đưa vào mô hình mạng nơ-ron đồ thị để học quan hệ giữa các mẫu dữ liệu. Bên cạnh đó, chương cũng đề xuất cơ chế gán nhãn giả có chọn lọc dựa trên entropy, trong đó entropy dự đoán được sử dụng để phân tầng độ tin cậy và lựa chọn các mẫu có độ tin cậy cao trước khi đưa vào quá trình tái huấn luyện. Cách tiếp cận này cho phép khai thác dữ liệu chưa gán nhãn theo hướng có kiểm soát độ tin cậy và hạn chế nguy cơ lan truyền nhãn sai trong học bán giám sát.

Các kết quả thực nghiệm cho thấy việc kết hợp biểu diễn đa nguồn, học quan hệ trên đồ thị và chọn lọc nhãn giả theo entropy làm thay đổi các chỉ số cân bằng lớp như Macro Recall, Macro F1-score, Macro AUC, Sensitivity và Specificity trong thiết lập ISIC 2020. Phân tích theo mức độ tin cậy và ma trận nhầm lẫn cho thấy entropy có thể dùng làm tiêu chí lọc mẫu trước khi tái huấn luyện. Kết quả của chương bổ sung cho hướng nghiên cứu của luận án từ tối ưu hóa biểu diễn đặc trưng sang khai thác quan hệ dữ liệu và học bán giám sát có kiểm soát độ tin cậy cho bài toán phân loại ảnh tổn thương da.

KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TƯƠNG LAI

Luận án nghiên cứu bài toán phân loại ảnh tổn thương da trong điều kiện dữ liệu mất cân bằng, khác biệt về nguồn thu nhận, có nhiều nguồn thông tin và hạn chế dữ liệu gán nhãn. Trên cơ sở tổng quan các nghiên cứu liên quan, luận án tập trung giải quyết hai vấn đề. Vấn đề thứ nhất là tổ chức biểu diễn ảnh thông qua chọn lọc bản đồ đặc trưng và hợp nhất với đặc trưng thủ công. Vấn đề thứ hai là khai thác đồng thời thông tin đa nguồn, quan hệ giữa các mẫu và dữ liệu chưa gán nhãn. Hai vấn đề này tương ứng với hai đóng góp chính được trình bày trong Chương 2 và Chương 3.

Đóng góp thứ nhất là mô hình BlockNet cho phân loại ảnh tổn thương da [CT3]. Mô hình sử dụng VGG16 để trích xuất bản đồ đặc trưng tại nhiều block tích chập. Entropy kích hoạt được tính từ histogram giá trị của từng bản đồ và dùng làm tiêu chí sàng lọc không giám sát theo khoảng entropy riêng của từng block. Entropy trong trường hợp này chỉ mô tả mức phân tán của đáp ứng, không trực tiếp đo thông tin liên quan nhãn hoặc khả năng phân biệt lớp. Vì vậy, giá trị của biểu diễn sau sàng lọc được kiểm chứng bằng hiệu năng phân loại, không được suy ra chỉ từ entropy của từng bản đồ.

Các bản đồ được giữ lại được chuyển thành vector đặc trưng học sâu và hợp nhất với đặc trưng thủ công gồm histogram HSV, GLCM và fractal. Dữ liệu sau hợp nhất được cân bằng trong không gian đặc trưng trước khi huấn luyện bộ phân loại. Cách tổ chức này tạo thành một quy trình gồm chọn lọc, hợp nhất và cân bằng biểu diễn, trong khi vẫn giữ nguyên backbone CNN. Do đó, đóng góp của BlockNet không nằm ở việc xây dựng một backbone mới, sử dụng riêng entropy hoặc thực hiện phép ghép nối, mà nằm ở cách kết hợp các thành phần trên trong một quy trình thống nhất cho phân loại ảnh tổn thương da.

Kết quả thực nghiệm trên VNCancer và HAM10000 ghi nhận sự khác biệt giữa cấu hình sử dụng toàn bộ bản đồ đặc trưng và cấu hình áp dụng sàng lọc entropy. Cấu hình BlockNet đầy đủ đạt kết quả cạnh tranh so với các backbone được khảo sát. Các kết quả này cho thấy biểu diễn được tổ chức lại có thể hỗ trợ phân loại trong phạm vi dữ liệu và giao thức thực nghiệm của luận án. Tuy nhiên, các thí nghiệm hiện có chưa tách độc lập hoàn toàn ảnh hưởng của entropy, đặc trưng thủ công và bước cân bằng dữ liệu trong cùng một giao thức loại trừ. Vì vậy, luận án không quy toàn bộ mức cải thiện của cấu hình đầy đủ cho riêng một thành phần.

Đóng góp thứ hai là khung học bán giám sát đa nguồn dựa trên mạng nơ ron đồ thị [CT4]. Hướng này không sử dụng vector BlockNet và không thực hiện chọn lọc bản đồ đặc trưng theo entropy kích hoạt của Chương 2. Đặc trưng học sâu được tạo từ các vector đầu ra sau pooling của VGG19, DenseNet121 và MobileNet. Đặc trưng thủ công chỉ gồm histogram RGB, histogram HSV và fractal. Các nhóm đặc trưng thị giác được lựa chọn bằng phân tích Chi square trong thiết lập của Chương 3.

Biểu diễn đa nguồn được xây dựng qua hai tầng. Đặc trưng học sâu và đặc trưng thủ công được ghép nối thành biểu diễn ảnh. Biểu diễn này sau đó được hợp nhất với metadata lâm sàng bằng cơ chế chú ý có xét metadata để tạo biểu diễn nút. Cơ chế chú ý sinh trọng số cho đường ảnh và đường metadata, không sinh ba trọng số độc lập cho đặc trưng học sâu, đặc trưng thủ công và metadata. Trên cơ sở biểu diễn nút, đồ thị k láng

giềng gần nhất được xây dựng để mô hình hóa quan hệ tương đồng giữa các mẫu.

Các kiến trúc GNN gồm GCN, GAT, GraphSAGE, GINE và mô hình kết hợp GraphSAGE với GAT được khảo sát trên cùng thiết lập dữ liệu. Quy trình huấn luyện gồm pha có giám sát trên tập dữ liệu có nhãn và pha gán nhãn giả có chọn lọc để tái huấn luyện. Trong Chương 3, entropy được tính trên phân phối xác suất dự đoán và được dùng để phân tầng độ tin cậy. Đại lượng này khác với entropy kích hoạt của Chương 2 cả về đối tượng tính toán và mục đích sử dụng.

Kết quả thực nghiệm cho thấy việc kết hợp biểu diễn đa nguồn, quan hệ đồ thị và dữ liệu chưa gán nhãn làm thay đổi cân bằng giữa các chỉ số phân loại. Sau pha gán nhãn giả, một số chỉ số như Macro F1, Macro Recall và Macro AUC được cải thiện trong giao thức kiểm chứng. Toàn bộ 10,982 mẫu của tập ISIC 2020 Test được phân vào ba nhóm tin cậy gồm 3,623 mẫu High Trust, 3,621 mẫu Medium Trust và 3,738 mẫu Low Trust. Chỉ nhóm High Trust được dùng để tạo nhãn giả và mở rộng tập huấn luyện.

Việc phân tầng trên chỉ phản ánh mức độ chắc chắn tương đối của dự đoán theo entropy. Do ISIC 2020 Test không có nhãn phân loại công khai, các nhãn giả không phải nhãn tham chiếu, còn ba nhóm tin cậy không biểu thị độ chính xác hoặc phân bố lớp thật. Hiệu quả của bước tái huấn luyện được đánh giá thông qua dữ liệu có nhãn và sự thay đổi của các chỉ số trong giao thức thực nghiệm, không dựa trên giả định rằng nhãn giả là đúng.

Hai đóng góp sử dụng entropy ở hai cấp độ khác nhau. Chương 2 dùng entropy kích hoạt để sàng lọc bản đồ đặc trưng trước khi hợp nhất. Chương 3 dùng entropy dự đoán để lựa chọn mẫu chưa gán nhãn trước khi tái huấn luyện. Hai hướng không dùng chung cấu hình đặc trưng, nhưng cùng hướng tới tổ chức và khai thác biểu diễn dữ liệu phù hợp hơn cho phân loại ảnh tổn thương da.

Bên cạnh các kết quả đạt được, luận án còn một số hạn chế. Thứ nhất, ảnh hưởng riêng của từng thành phần trong BlockNet chưa được đánh giá đầy đủ bằng thí nghiệm loại trừ trên cùng các lần chia dữ liệu. Thứ hai, việc cân bằng trong không gian đặc trưng chưa được đối chứng độc lập với class weighted loss, focal loss, SMOTE và các phương pháp lấy mẫu khác. Kết quả riêng của lớp melanoma trước và sau khi chỉ thay đổi cơ chế cân bằng cũng chưa được báo cáo đầy đủ. Thứ ba, metadata lâm sàng trong các bộ dữ liệu công khai còn thiếu và chưa đồng nhất. Thứ tư, chất lượng đồ thị phụ thuộc vào biểu diễn nút, độ đo tương đồng và số lượng nút lân cận. Thứ năm, entropy dự đoán chưa thay thế đánh giá hiệu chỉnh xác suất bằng ECE, Brier score hoặc đường cong calibration. Cuối cùng, các phương pháp mới được kiểm chứng trên dữ liệu công khai và chưa được ngoại kiểm đa trung tâm hoặc đánh giá lâm sàng trên người bệnh.

Từ các giới hạn trên, các hướng nghiên cứu tiếp theo gồm:

- Thực hiện thí nghiệm loại trừ đầy đủ cho chọn lọc entropy, đặc trưng thủ công và cân bằng dữ liệu của BlockNet trên cùng giao thức.
- So sánh các cơ chế xử lý mất cân bằng và phân tích riêng hiệu quả trên các lớp thiếu số có ý nghĩa lâm sàng.
- Nghiên cứu cơ chế hợp nhất linh hoạt hơn cho từng nguồn đặc trưng, bao gồm chú ý chéo và Transformer đa nguồn.
- Đánh giá độ nhạy theo cách xây dựng đồ thị, số lượng láng giềng và độ đo tương đồng, đồng thời mở rộng sang graph transformer hoặc đồ thị không đồng nhất.

- Kết hợp entropy với các phương pháp hiệu chỉnh xác suất và ước lượng độ không chắc chắn để kiểm soát nhãn giả chặt chẽ hơn.
- Tích hợp phương pháp giải thích mô hình, ngoại kiểm trên dữ liệu độc lập và đánh giá lâm sàng trước khi xem xét ứng dụng trong hệ thống hỗ trợ sàng lọc.

Nhìn chung, luận án đã hoàn thành hai mục tiêu nghiên cứu thông qua BlockNet [CT3] và khung học bán giám sát đa nguồn trên đồ thị [CT4]. Các kết quả khẳng định khả năng khai thác biểu diễn đa nguồn và dữ liệu chưa gán nhãn trong phạm vi thực nghiệm đã xác định. Mô hình chưa được xem là công cụ chẩn đoán độc lập nếu chưa có ngoại kiểm và đánh giá lâm sàng đầy đủ.

DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN

- [CT1]. Nguyen, Thi Trang; Nguyen, Ha Phuong; Vu, Van Hieu. "Hybrid Feature Fusion: Enhancing Scene Image Classification with Handcrafted and Deep Learning Approaches". In: Kỷ yếu Hội nghị Khoa học Công nghệ Quốc gia lần thứ XVII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), Hà Nội, Việt Nam, 08--09/08/2024. 2024. DOI: 10.15625/vap.2024.0226.
- [CT2]. Vu, Van Hieu; Nguyen, Thi Trang; Nguyen-Thi, Thao. "Ensemble Deep Learning Models for Enhanced Skin Lesion Diagnosis Using the HAM10000 Dataset". In: Hội thảo quốc gia lần thứ XXVII: Một số vấn đề chọn lọc của Công nghệ thông tin và truyền thông, Nha Trang, Việt Nam, 11--12/10/2024. 2024.
- [CT3] Nguyen, T.T., Vu, V.H. "Entropy-based feature fusion via BlockNet for robust skin cancer classification". Iran Journal of Computer Science. Volume 8, pages 2871–2891 (2025). DOI:10.1007/s42044-025-00351-8 (Scopus, Q2).
- [CT4] Nguyen, T.T., Vu, V.H. \& Nguyen, V.A. Semi-supervised GNN for Multi-modal Skin Lesion Classification. International Journal of Computational Intelligence Systems. Volume 19, article number 208 (2026). <https://doi.org/10.1007/s44196-026-01296-x> (SCI-E/Q1).

TÀI LIỆU THAM KHẢO

- [1] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. “Dermatologist-level classification of skin cancer with deep neural networks”. In: *nature* 542.7639 (2017), pp. 115–118.
- [2] Geert Litjens, Thijs Kooi, Babak E. Bejnordi, Arnaud A. A. Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sanchez. “A Survey on Deep Learning in Medical Image Analysis”. In: *Medical Image Analysis* 42 (2017), pp. 60–88. DOI: 10.1016/j.media.2017.07.005.
- [3] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions”. In: *Scientific data* 5.1 (2018), pp. 1–9.
- [4] Mai Tiến Dũng. “Phát triển một số phương pháp phân loại ảnh với số lượng lớp lớn”. Tóm tắt luận án tiến sĩ. Luận án Tiến sĩ Khoa học Máy tính. Thành phố Hồ Chí Minh, Việt Nam: Trường Đại học Công nghệ Thông tin, Đại học Quốc gia Thành phố Hồ Chí Minh, 2017.
- [5] Nguyễn Thành Trung. “Nghiên cứu và phát triển các phương pháp nâng cao chất lượng của hình ảnh y sinh bị nhiễu”. Tóm tắt luận án tiến sĩ. Luận án Tiến sĩ Công nghệ Kỹ thuật Điện tử, Truyền thông. Hà Nội, Việt Nam: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, 2021.
- [6] Nguyễn Đức Anh. “Các phương pháp đảm bảo tính chắc chắn cho một số mô hình học sâu”. Tóm tắt luận án tiến sĩ. Luận án Tiến sĩ Kỹ thuật Phần mềm. Hà Nội, Việt Nam: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, 2023.
- [7] Trần Văn Tiến. “Ứng dụng kỹ thuật quang học đa bước sóng trong chẩn đoán hình ảnh da liễu, nội soi và phụ khoa”. Tóm tắt luận án tiến sĩ. Luận án Tiến sĩ. Thành phố Hồ Chí Minh, Việt Nam: Trường Đại học Bách khoa, Đại học Quốc gia Thành phố Hồ Chí Minh, 2022.
- [8] Catarina Barata, M. Emre Celebi, and Jorge S. Marques. “A System for the Detection of Melanomas in Dermoscopy Images Using Shape and Color Features”. In: *IEEE Journal of Biomedical and Health Informatics* 18.3 (2014), pp. 949–959.
- [9] Jose-Agustin Almaraz-Damian, Volodymyr I. Ponomaryov, Sergiy Sadovnychiy, and Heydy Castillejos-Fernandez. “Melanoma and Nevus Skin Lesion Classification Using Handcraft and Deep Learning Feature Fusion via Mutual Information Measures”. In: *Entropy* 22 (2020).
- [10] Misaj Sharafudeen and Vinod Chandra S S. “Detecting skin lesions fusing handcrafted features in image network ensembles”. In: *Multimedia Tools and Applications* 82.2 (2023), pp. 3155–3175.
- [11] Amirreza Mahbod, Gerald Schaefer, Isabella Ellinger, Rupert Ecker, Alain Pitiot, and Chunliang Wang. “Fusing fine-tuned deep features for skin lesion classification”. In: *Computerized Medical Imaging and Graphics* 71 (2019), pp. 19–29.

- [12] Navpreet Kaur et al. “Skin lesion classification using transfer learning”. In: *Soft Computing* 28 (2024), pp. 3935–3947. DOI: 10.1007/s00500-024-09949-9.
- [13] Abida Noaman, Reyaz Ahmad, Muhammad Farhan Khan, Abdul Salam Mohammed, Muhammad Farooq, and Khan Muhammad Adnan. “Beyond binary: multi-class skin lesion classification with AlexNet transfer learning-towards enhanced dermatological diagnosis”. In: *Discover Applied Sciences* 7.1 (2024), p. 35.
- [14] W. Liu, Y. Wang, et al. “Metadata-P2T-Net: A Skin Lesion Classification Network Integrating Dermoscopic Metadata and Image Data”. In: *Proceedings of SPIE*. Vol. 13250. 2024, p. 132500L. DOI: 10.1117/12.3038813.
- [15] V. Chakkarapani, S. Poornapushpakala, and S. Suresh. “Enhancing Skin Cancer Detection with Multimodal Data Integration: A Combined Approach Using Images and Clinical Notes”. In: *SN Computer Science* 6.1 (2025), p. 72. DOI: 10.1007/s42979-024-03601-x.
- [16] Alma Alhelí Pedro Pérez, Pamela Ignacia Romero Jofré, Soledad Vidaurre, Cristian Navarrete-Dechent, and Domingo Mery. “A Review of Metadata and Deep Learning Strategies for Skin Lesion Classification”. In: *Journal of Investigative Dermatology* 146.1 (2026), 249–251.e8. DOI: 10.1016/j.jid.2025.06.1593.
- [17] N. Y. Tran-Van et al. “A multimodal skin lesion classification through cross-attention fusion and collaborative inference”. In: *Computer Methods and Programs in Biomedicine* (2025).
- [18] M. Ahmad et al. “Fusion of metadata and dermoscopic images for skin disease classification using discounted evidential fusion”. In: *Information Fusion* (2025).
- [19] Pampana Murali and Dilwar Hussain Mazumder. “DermaScanAI: An Explainable Hybrid Deep Learning Framework for Automated Skin Lesion Classification Using Dual Attention and Metadata Fusion”. In: *Scientific Reports* (2026). DOI: 10.1038/s41598-026-46011-0.
- [20] Thomas N. Kipf and Max Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In: *International Conference on Learning Representations (ICLR)*. arXiv:1609.02907. 2017.
- [21] William L. Hamilton, Rex Ying, and Jure Leskovec. “Inductive Representation Learning on Large Graphs”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:1706.02216. 2017.
- [22] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. “Graph Attention Networks”. In: *International Conference on Learning Representations (ICLR)*. arXiv:1710.10903. 2018.
- [23] Dong-Hyun Lee. “Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks”. In: *ICML Workshop on Challenges in Representation Learning*. 2013.

- [24] Kihyuk Sohn, David Berthelot, Nicholas Carlini, et al. “FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence”. In: *NeurIPS*. 2020.
- [25] Eric Arazo, Diego Ortego, Paul Albert, Noel E. O’Connor, and Kevin McGuinness. “Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning”. In: *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207302.
- [26] Noel C. Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen W. Dusza, David A. Gutman, Brian Helba, Aditi Kalloo, Konstantinos Liopyris, Michael A. Marchetti, et al. “Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)”. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. 2018, pp. 168–172. DOI: 10.1109/ISBI.2018.8363547.
- [27] Nils Gessert, Maximilian Nielsen, et al. “Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data”. In: *MethodsX 7* (2020), p. 100864. DOI: 10.1016/j.mex.2020.100864.
- [28] Sakib Ahammed, Xia Cui, Wenqi Lu, and Moi Hoon Yap. “Skin Lesion Classification Using Dermoscopic Images and Clinical Metadata: Insights from Multimodal Models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2025, pp. 222–230. DOI: 10.1109/CVPRW67362.2025.00027.
- [29] G. Cai et al. “A multimodal transformer to fuse images and metadata for skin disease classification”. In: *The Visual Computer* 39 (2023), pp. 2781–2793. DOI: 10.1007/s00371-022-02492-4.
- [30] Chirag Patel et al. “Classification of skin cancer from dermoscopic images using deep neural network architectures”. In: *Multimedia Tools and Applications* 81 (2022), pp. 27467–27485. DOI: 10.1007/s11042-022-13847-3.
- [31] Rahul Gupta et al. “Melanoma classification from dermatoscopy images using knowledge distillation for highly imbalanced data”. In: *Computers in Biology and Medicine* 158 (2023), p. 106571. DOI: 10.1016/j.combiomed.2023.106571.
- [32] Aman Vashisth et al. “Fusion of transformer attention and CNN features for skin cancer detection”. In: *Applied Soft Computing* 153 (2024), p. 112013. DOI: 10.1016/j.asoc.2024.112013.
- [33] D. P. Yadav, S. Samiappan, E. O. Aboagye, S. Mishra, S. Plein, and M. R. Dweck. “DSCATNet: Deep skin cancer attention transformer network for classification of skin cancer from dermoscopic images”. In: *PLOS ONE* 19.9 (2024), e0312598. DOI: 10.1371/journal.pone.0312598. URL: <https://doi.org/10.1371/journal.pone.0312598>.
- [34] Getenet Hailu Dagnaw. “Skin cancer classification using vision transformers and explainable artificial intelligence”. In: *Journal of Medical Artificial Intelligence* 7 (2024), p. 14. DOI: 10.21037/jmai-24-6. URL: <https://doi.org/10.21037/jmai-24-6>.

- [35] Huan Liu, Rudy Setiono, and Lei Yu. “Chi2: Feature selection and discretization of numeric attributes”. In: *Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence*. IEEE. 1995, pp. 388–391.
- [36] Robert M. Haralick, K. Shanmugam, and I. Dinstein. “Textural Features for Image Classification”. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6 (1973), pp. 610–621. DOI: 10.1109/TSMC.1973.4309314.
- [37] Michael J Swain and Dana H Ballard. “Color indexing”. In: *International Journal of Computer Vision*. Vol. 7. Springer, 1991, pp. 11–32.
- [38] Markus Stricker and Markus Orengo. “Similarity of Color Images”. In: *Storage and Retrieval for Image and Video Databases III (SPIE)*. Vol. 2420. SPIE, 1995, pp. 381–392.
- [39] Navneet Dalal and Bill Triggs. “Histograms of Oriented Gradients for Human Detection”. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. 2005, pp. 886–893. DOI: 10.1109/CVPR.2005.177.
- [40] John R. Smith. *Color Imaging Handbook*. New York, USA: Academic Press, 1996.
- [41] Benoit B. Mandelbrot. *The Fractal Geometry of Nature*. New York, USA: W. H. Freeman and Company, 1982.
- [42] Shmuel Peleg, Joseph Naor, Ralph Hartley, and David Avnir. “Multiple Resolution Texture Analysis and Classification”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.4 (1984), pp. 518–523. DOI: 10.1109/TPAMI.1984.4767557.
- [43] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.7 (2002), pp. 971–987. DOI: 10.1109/TPAMI.2002.1017623.
- [44] John G. Daugman. “Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters”. In: *Journal of the Optical Society of America A* 2.7 (1985), pp. 1160–1169. DOI: 10.1364/JOSAA.2.001160.
- [45] Stephane G. Mallat. “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.7 (1989), pp. 674–693. DOI: 10.1109/34.192463.
- [46] John Canny. “A Computational Approach to Edge Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (1986), pp. 679–698. DOI: 10.1109/TPAMI.1986.4767851.
- [47] Karl Pearson. “On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Reasonably Be Supposed to Have Arisen from Random Sampling”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50.302 (1900), pp. 157–175. DOI: 10.1080/14786440009463897.

- [48] Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S. Kankanhalli. “Multimodal Fusion for Multimedia Analysis: A Survey”. In: *Multimedia Systems* 16.6 (2010), pp. 345–379. DOI: 10.1007/s00530-010-0182-0.
- [49] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. “Multimodal Machine Learning: A Survey and Taxonomy”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.2 (2019), pp. 423–443. DOI: 10.1109/TPAMI.2018.2798607.
- [50] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *International Conference on Learning Representations (ICLR)* (2015). URL: <https://arxiv.org/abs/1409.1556>.
- [51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [52] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. “Densely Connected Convolutional Networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.
- [53] Francois Chollet. “Xception: Deep learning with depthwise separable convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. 2017, pp. 1251–1258.
- [54] Mingxing Tan and Quoc Le. “EfficientNet: Rethinking model scaling for convolutional neural networks”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6105–6114.
- [55] Mingxing Tan and Quoc V. Le. “EfficientNetV2: Smaller Models and Faster Training”. In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. Proceedings of Machine Learning Research. PMLR, 2021, pp. 10096–10106.
- [56] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *International Conference on Learning Representations (ICLR)*. 2021. arXiv: 2010.11929.
- [57] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. “Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 10012–10022.
- [58] Xuejun Zhang, Yehui Liu, Ganxin Ouyang, Wenkang Chen, Aobo Xu, Takeshi Hara, Xiangrong Zhou, and Dongbo Wu. “DermViT: Diagnosis-Guided Vision Transformer for Robust and Efficient Skin Lesion Classification”. In: *Bioengineering* 12.4 (2025), p. 421. DOI: 10.3390/bioengineering12040421.

- [59] Matthew D. Zeiler and Rob Fergus. “Visualizing and Understanding Convolutional Networks”. In: *Computer Vision – ECCV 2014*. Springer, 2014, pp. 818–833. DOI: 10.1007/978-3-319-10590-1_53.
- [60] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. “Learning Deep Features for Discriminative Localization”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2921–2929.
- [61] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 618–626.
- [62] Naveed Ahmad, Jamal Hussain Shah, Muhammad Attique Khan, Jamel Baili, Ghulam Jillani Ansari, Usman Tariq, Ye Jin Kim, and Jae-Hyuk Cha. “A novel framework of multiclass skin lesion recognition from dermoscopic images using deep learning and explainable AI”. In: *Frontiers in Oncology* 13 (2023), p. 1151257.
- [63] Syed Adil Hussain Shah, Syed Taimoor Hussain Shah, Roa’a Khaled, Andrea Buccoliero, Syed Baqir Hussain Shah, Angelo Di Terlizzi, Giacomo Di Benedetto, and Marco Agostino Deriu. “Explainable AI-based skin cancer detection using CNN, particle swarm optimization and machine learning”. In: *Journal of Imaging* 10.12 (2024), p. 332.
- [64] Pieter Van Molle, Michaël De Strooper, Davy Timmerman, Olivier Debeir, and Etienne Decenci re. “Visualizing Convolutional Neural Networks to Improve Decision Support for Skin Lesion Classification”. In: *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. Springer, 2018, pp. 115–123.
- [65] Xi Jia and Linlin Shen. “Skin Lesion Classification Using Class Activation Map”. In: *arXiv preprint arXiv:1703.01053* (2017).
- [66] Sarah Benyahia, Boudjelal Meftah, and Olivier Lezoray. “Multi-Features Extraction Based on Deep Learning for Skin Lesion Classification”. In: *Tissue and Cell* 74 (2022), p. 101701. DOI: 10.1016/j.tice.2021.101701.
- [67] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-Excitation Networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7132–7141. DOI: 10.1109/CVPR.2018.00745.
- [68] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. “CBAM: Convolutional Block Attention Module”. In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 3–19. DOI: 10.1007/978-3-030-01234-2_1.
- [69] Jielei Wang, Ting Jiang, Zongyong Cui, and Zongjie Cao. “Filter Pruning with a Feature Map Entropy Importance Criterion for Convolution Neural Networks Compressing”. In: *Neurocomputing* 461 (2021), pp. 41–54. DOI: 10.1016/j.neucom.2021.07.034.
- [70] Linsong Shao, Haorui Zuo, Jianlin Zhang, Zhan Xu, Jun Yao, Zhan Wang, and Hong Li. “Filter Pruning via Measuring Feature Map Information”. In: *Sensors* 21.19 (2021), p. 6601. DOI: 10.3390/s21196601.

- [71] Jiquan Ngiam, Aditya Khosla, Minhyung Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng. “Multimodal Deep Learning”. In: *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*. 2011, pp. 689–696. DOI: 10.5555/3104482.3104569.
- [72] Y. Li et al. “A review of deep learning-based information fusion techniques for medical image classification”. In: *Computers in Biology and Medicine* (2024).
- [73] P. Tang et al. “Fusionm4net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification”. In: *Medical Image Analysis* 76 (2022), p. 102307. DOI: 10.1016/j.media.2021.102307.
- [74] Amirreza Mahbod, Rupert Ecker, and Ramona Woitek. “Fusion of Foundation and Vision Transformer Model Features for Dermatoscopic Image Classification”. In: *arXiv preprint arXiv:2505.16338* (2025). arXiv: 2505.16338 [cs.CV].
- [75] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. “Attention is all you need”. In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.
- [76] Zhouhan Lin, Minwei Feng, Cícero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. “A Structured Self-attentive Sentence Embedding”. In: *ArXiv abs/1703.03130* (2017).
- [77] Y. Zhang et al. “A novel perspective for multi-modal multi-label skin lesion classification”. In: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2025, pp. 3549–3558.
- [78] Claude E. Shannon. “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- [79] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. 2nd ed. Wiley-Interscience, 2006.
- [80] Yarín Gal and Zoubin Ghahramani. “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”. In: *Proceedings of the 33rd International Conference on Machine Learning (ICML)*. PMLR. 2016, pp. 1050–1059.
- [81] Dan Hendrycks and Kevin Gimpel. “A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks”. In: *arXiv preprint arXiv:1610.02136* (2017).
- [82] A. Kurz et al. “Uncertainty Estimation in Medical Image Classification: Systematic Review”. In: *JMIR Medical Informatics* 10.8 (2022), e36427. DOI: 10.2196/36427.
- [83] L. Huang et al. “A Review of Uncertainty Quantification in Medical Image Analysis: Probabilistic and Non-Probabilistic Methods”. In: *Medical Image Analysis* (2024).

- [84] Zhanhong Qiu, Weiyan Gan, Zhi Yang, Ran Zhou, and Haitao Gan. “Dual Uncertainty-Guided Multi-Model Pseudo-Label Learning for Semi-Supervised Medical Image Segmentation”. In: *Mathematical Biosciences and Engineering* 21.2 (2024), pp. 2212–2232. DOI: 10.3934/mbe.2024097.
- [85] Hao He et al. “TrustMatch: Mitigating Pseudo-Label Bias in Semi-Supervised Learning with Trust-Aware Refinement”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2025.
- [86] OK Sikha, Alaysia Leilani B Stone, and Miguel A González Ballester. “Meta-UNet: enhancing skin-lesion segmentation with multimodal feature integration and uncertainty estimation”. In: *International Journal of Computer Assisted Radiology and Surgery* (2025), pp. 1–12.
- [87] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. “On calibration of modern neural networks”. In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1321–1330.
- [88] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. “Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles”. In: *Neural Information Processing Systems*. 2016.
- [89] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. “Graph neural networks: A review of methods and applications”. In: *AI open* 1 (2020), pp. 57–81.
- [90] William L. Hamilton. “Graph Representation Learning”. In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* (2020).
- [91] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. “How Powerful Are Graph Neural Networks?” In: *International Conference on Learning Representations (ICLR)*. 2019.
- [92] Ziheng Huang et al. “Systematic Comparison of Semi-supervised and Self-supervised Learning for Medical Image Classification”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
- [93] Abel Díaz Berenguer, Maryna Kvasnytsia, Matías Nicolás Bossa, Tanmoy Mukherjee, Nikos Deligiannis, and Hichem Sahli. “Semi-supervised medical image classification via distance correlation minimization and graph attention regularization”. In: *Medical Image Analysis* 94 (2024), p. 103107. DOI: 10.1016/j.media.2024.103107.
- [94] Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. “Disease Prediction using Graph Convolutional Networks: Application to Autism Spectrum Disorder and Alzheimer’s Disease”. In: *Medical Image Analysis* 48 (2018), pp. 117–130.
- [95] Kun Liu, Shuyi Ling, and Sidong Liu. “Semi-Supervised Medical Image Classification with Pseudo Labels Using Coalition Similarity Training”. In: *Mathematics* 12.10 (2024), p. 1537. DOI: 10.3390/math12101537.
- [96] Xiaojin Zhu. “Semi-Supervised Learning Literature Survey”. In: *Computer Sciences Technical Report 1530, University of Wisconsin-Madison* (2005).

- [97] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-Supervised Learning*. MIT Press, 2009.
- [98] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Robert Bamler, and Xiao Xiang Zhu. “A Survey of Uncertainty in Deep Neural Networks”. In: *Artificial Intelligence Review* 56 (2023), pp. 1513–1589. DOI: 10.1007/s10462-023-10562-9.
- [99] Qizhe Xie, Eduard H. Hovy, Minh-Thang Luong, and Quoc V. Le. “Self-Training With Noisy Student Improves ImageNet Classification”. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 10684–10695. URL: <https://doi.org/10.48550/arXiv.1911.04252>.
- [100] David M. W. Powers. “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation”. In: *Journal of Machine Learning Technologies* 2.1 (2011), pp. 37–63.
- [101] Tom Fawcett. “An Introduction to ROC Analysis”. In: *Pattern Recognition Letters* 27 (2006), pp. 861–874. DOI: 10.1016/j.patrec.2005.10.010.
- [102] Marina Sokolova and Guy Lapalme. “A systematic analysis of performance measures for classification tasks”. In: *Information Processing & Management* 45.4 (2009), pp. 427–437.
- [103] Van-Hieu Vu. “Hybrid feature integration for enhanced atopic dermatitis diagnosis using multi-ResNet transformer models”. In: *Journal of Electronic Imaging* 33.6 (2024), pp. 063039–063039.
- [104] Van-Hieu Vu. “A deep learning method combined with clustering to support the diagnosis of atopic dermatitis and contact dermatitis in Vietnam”. In: *Neural Computing and Applications* 37.16 (2025), pp. 9911–9935.
- [105] Andrew P. Bradley. “The use of the area under the ROC curve in the evaluation of machine learning algorithms”. In: *Pattern Recognition* 30.7 (1997), pp. 1145–1159.
- [106] Ron Kohavi. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI 1995)*. 1995, pp. 1137–1143. DOI: 10.5555/1643031.1643047.
- [107] Thomas G. Dietterich. “Approximate statistical tests for comparing supervised classification learning algorithms”. In: *Neural Computation* 10.7 (1998), pp. 1895–1923.
- [108] Janez Demšar. “Statistical comparisons of classifiers over multiple data sets”. In: *Journal of Machine Learning Research* 7 (2006), pp. 1–30.
- [109] Abhijith Das et al. “Comparative Analysis of Multimodal Architectures for Effective Skin Lesion Detection Using Clinical and Image Data”. In: *Frontiers in Artificial Intelligence* (2025). DOI: 10.3389/frai.2025.1608837.

- [110] Ramya Panneerselvam, Sathiyabhama Balasubramaniam, Vidhushavarshini Sureshku-mar, Vinayakumar Ravi, and Siti Sarah Maidin. “Multimodal Skin Cancer Pre-diction: Integrating Dermoscopic Images and Clinical Metadata with Transfer Learning”. In: *Biological Sciences* (2025). DOI: 10.2174/011875036235844425012007
- [111] Xinyu Zhao, Lei Zhang, Xiaohong Li, and Yanning Zhang. “Multi-feature fu-sion for skin lesion classification based on deep learning”. In: *IEEE Access* 9 (2021), pp. 123456–123468.
- [112] Zhanlin Ji, Xuan Wang, Chunling Liu, Zhiwu Wang, Na Yuan, and Ivan Ganchev. “EFAM-Net: A Multi-Class Skin Lesion Classification Model Utilizing En-hanced Feature Fusion and Attention Mechanisms”. In: *IEEE Access* 12 (2024), pp. 143029–143041. DOI: 10.1109/ACCESS.2024.3468612.
- [113] Ying Weng, Yiming Zhang, Wenxin Wang, and Tom Dening. “Semi-supervised Information Fusion for Medical Image Analysis: Recent Progress and Future Perspectives”. In: *Information Fusion* (2024). DOI: 10.1016/j.inffus.2024.102263.
- [114] Justin M Johnson and Taghi M Khoshgoftaar. “Survey on deep learning with class imbalance”. In: *Journal of Big Data* 6.1 (2019), pp. 1–54.
- [115] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. “Fo-cal Loss for Dense Object Detection”. In: *Proceedings of the IEEE Interna-tional Conference on Computer Vision (ICCV)*. 2017. arXiv: 1708.02002 [cs.CV].
- [116] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. “SMOTE: Synthetic Minority Over-sampling Technique”. In: *Journal of Arti-ficial Intelligence Research* 16 (2002), pp. 321–357. DOI: 10.1613/jair.953.
- [117] Haibo He and Edwardo A. Garcia. “Learning from Imbalanced Data”. In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), pp. 1263–1284. DOI: 10.1109/TKDE.2008.239.
- [118] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. “A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks”. In: *Neu-ral Networks* 106 (2018), pp. 249–259. DOI: 10.1016/j.neunet.2018.07.011.
- [119] Dermatology Experts at Vietnam National Hospital of Dermatology and Venere-ology. *VNCancer Dataset: A Dermatological Dataset for Skin Cancer Diag-nosis in Vietnam*. <https://dalieu.vn>. 2024.
- [120] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. “Rethinking the Inception Architecture for Computer Vi-sion”. In: *Proceedings of the IEEE Conference on Computer Vision and Pat-tern Recognition*. 2016, pp. 2818–2826.
- [121] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. “MobileNetV2: Inverted Residuals and Linear Bottlenecks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recog-nition (CVPR)*. 2018, pp. 4510–4520.