

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Đỗ Minh Tuấn

**NGHIÊN CỨU KỸ THUẬT KẾT HỢP CÁC MẠNG NƠ-RON TÍCH CHẬP
VÀ ỨNG DỤNG CHẨN ĐOÁN MỘT SỐ BỆNH UNG THƯ DA Ở
VIỆT NAM**

LUẬN VĂN THẠC SĨ MÁY TÍNH

Hà Nội - 2024

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Đỗ Minh Tuấn

**NGHIÊN CỨU KỸ THUẬT KẾT HỢP CÁC MẠNG NƠ-RON TÍCH CHẬP
VÀ ỨNG DỤNG CHẨN ĐOÁN MỘT SỐ BỆNH UNG THƯ DA Ở
VIỆT NAM**

LUẬN VĂN THẠC SĨ MÁY TÍNH

**Ngành: Hệ thống thông tin
Mã số: 8480104**

NGƯỜI HƯỚNG DẪN KHOA HỌC:

A handwritten signature in blue ink, appearing to be 'Vũ Văn Hiệu', written over a horizontal line.

TS. Vũ Văn Hiệu

Hà Nội - 2024

LỜI CAM ĐOAN

Tôi xin cam đoan rằng luận văn này là công trình nghiên cứu của riêng tôi, được thực hiện dưới sự hướng dẫn khoa học của TS. Vũ Văn Hiệu. Các số liệu và kết quả trong luận văn là trung thực và chưa từng được công bố trong bất kỳ công trình nghiên cứu nào khác.

Ngoài ra, các trích dẫn từ tài liệu tham khảo đều được ghi rõ nguồn gốc theo quy định. Nếu có điều gì không đúng, tôi xin hoàn toàn chịu trách nhiệm về nội dung luận văn này.

Hà Nội, ngày 18 tháng 11 năm 2024

Học viên

Đỗ Minh Tuấn

LỜI CẢM ƠN

Trước hết, tôi xin gửi lời cảm ơn chân thành và sâu sắc nhất tới TS Vũ Văn Hiệu, người đã tận tình hướng dẫn, chỉ bảo và hỗ trợ tôi trong suốt quá trình thực hiện luận văn này. Thầy đã luôn sát cánh, đưa ra những nhận xét quý báu, những định hướng đúng đắn và kịp thời, giúp tôi vượt qua những khó khăn và thử thách trong quá trình nghiên cứu.

Tôi cũng xin gửi lời cảm ơn tới ban Lãnh đạo, phòng Đào tạo, các thầy cô trong Khoa Công nghệ Thông tin và Viễn thông, Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã trang bị cho tôi những kiến thức nền tảng, tạo điều kiện tốt nhất cho tôi trong suốt quá trình học tập và nghiên cứu.

Bên cạnh đó, tôi xin gửi lời cảm ơn chân thành tới gia đình và bạn bè, những người đã luôn động viên, khích lệ và chia sẻ những khó khăn, thử thách trong suốt quá trình tôi thực hiện luận văn.

Cuối cùng, tôi xin cảm ơn tất cả những cá nhân và tổ chức đã trực tiếp và gián tiếp giúp đỡ tôi hoàn thành công trình này. Sự hỗ trợ, góp ý từ quý vị đã góp phần quan trọng giúp tôi hoàn thành luận văn này.

Hà Nội, ngày 18. tháng 11. năm 2024

Học viên

Đỗ Minh Tuấn

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN.....	ii
MỤC LỤC	v
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT	vi
DANH MỤC CÁC HÌNH VẼ.....	viii
DANH MỤC CÁC BẢNG BIỂU	x
MỞ ĐẦU.....	1
CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU CHẨN ĐOÁN UNG THƯ DA SỬ DỤNG HỌC SÂU VÀ ĐẶC TRUNG HỌC SÂU	7
1.1 Giới thiệu.....	7
1.1.1 Ung thư da ở Việt Nam.....	7
1.1.2 Một số nguyên nhân gây ung thư da.....	8
1.1.3 Một số kỹ thuật gần đây trong việc chẩn đoán ung thư da	12
1.2 Nghiên cứu liên quan	14
1.2.1 Phát hiện Ung thư da Melanoma sử dụng đặc trưng bằng kỹ thuật Xử lý ảnh	15
1.2.2 Phát hiện ung thư Melanoma bằng học sâu và đặc trưng sâu	15
1.2.3 Phát hiện ung thư Melanoma bằng học sâu và tăng cường dữ liệu.....	17
1.2.4 Tổ hợp các mô hình học sâu để cải thiện việc phát hiện ung thư da.....	18
1.2.5 Tổ hợp theo tầng (Stacked Ensemble)	19
Kết chương 1.....	21
CHƯƠNG 2. MÔ HÌNH HỌC SÂU KẾT HỢP HỖ TRỢ CHẨN ĐOÁN UNG THƯ DA	22
2.1 Kiến trúc hệ thống chẩn đoán ung thư da sử dụng học sâu.....	22

2.2	Trích rút đặc trưng học sâu trong bài toán phân lớp ảnh.....	24
2.2.1	Mạng Nơ-ron Tích chập (CNN) trong trích rút đặc trưng.....	24
2.2.2	Sử dụng Đặc trưng Học Sâu cho Phân Lớp Ảnh.....	25
2.2.3	Phân lớp ảnh sử dụng đặc trưng học sâu.....	26
2.2.4	Ưu điểm của đặc trưng học sâu trong phân lớp ảnh.....	26
2.3	Các kiến trúc mạng nơ-ron tích chập truyền thống.....	27
2.3.1	GoogleNet.....	28
2.3.2	InceptionV3.....	28
2.3.3	ViT_B16.....	28
2.3.4	VGG16 và VGG19.....	28
2.3.5	ResNet152.....	28
2.3.6	EfficientNetB0 và EfficientNetB7.....	29
2.3.7	MobileNetV2.....	29
2.3.8	Xception.....	29
2.3.9	DenseNet121.....	29
2.4	Học chuyển giao (Transfer learning).....	29
2.5	Trích xuất đặc trưng (Features extraction).....	31
2.6	Phân loại có giám sát và các tập hợp bộ phân loại.....	32
2.6.1	Cây quyết định (Decision Tree).....	33
2.6.2	Naive Bayes.....	33
2.6.3	Máy vector hỗ trợ (SVM).....	33
2.6.4	k-Nearest Neighbor (KNN).....	34
2.6.5	Phương pháp kết hợp (Ensemble Methods).....	34
2.6.6	XGBoost.....	35
2.7	Mô hình đề xuất.....	35
2.7.1	Phát biểu bài toán và mô hình.....	35
2.7.2	Các thành phần của mô hình CNN.....	37
2.7.3	Thuật toán Stacked Ensemble.....	42
	Kết chương 2.....	46

**CHƯƠNG 3. THỰC NGHIỆM MÔ HÌNH KẾT HỢP CÁC MẠNG NƠ-
RON TÍCH CHẬP CHẨN ĐOÁN BỆNH UNG THƯ DA 47**

3.1	Các chỉ số đánh giá hiệu năng cho phân loại đa lớp	47
3.1.1	Ma trận nhầm lẫn (Confusion Matrix).....	47
3.1.2	Precision và Recall	48
3.1.3	Độ chính xác (Accuracy).....	48
3.1.4	F1-Score	49
3.1.5	Macro F1-Score.....	49
3.1.6	Micro F1-Score	49
3.1.7	Độ nhạy (Sensitivity).....	50
3.1.8	Độ đặc hiệu (Specificity).....	50
3.1.9	AUC - Diện tích dưới đường cong (Area Under the Curve)	50
3.2	Cơ sở dữ liệu thực nghiệm	52
3.2.1	Tập dữ liệu VNCancer.....	52
3.2.2	Tiền xử lý	54
3.2.3	Cân bằng dữ liệu.....	56
3.3	Các kết quả thực nghiệm và luận giải	57
3.3.1	Cấu hình máy tính và các thư viện sử dụng.....	57
3.3.2	Các mô hình học sâu cơ sở	58
3.3.3	Kết quả thực nghiệm của mô hình đề xuất	73
3.3.4	Phân tích kết quả mô hình đề xuất và các mô hình cơ sở được đánh giá tốt nhất	79
	Kết chương 3	81
	KẾT LUẬN VÀ KIẾN NGHỊ	82
	TÀI LIỆU THAM KHẢO	83

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Kí hiệu	Tiếng Anh	Tiếng Việt
AUC	Area under the ROC curve	Chỉ số diện tích dưới đường cong ROC
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
RNN	Recurrent Neural Network	Mạng nơ-ron hồi quy
SVM	Support Vector Machine	Máy vectơ hỗ trợ
BLSTM	Bidirectional Long Short-Term Memory	Mạng nơ-ron ghi nhớ ngắn hạn hai chiều
FN	False Negative	Âm tính giả
FP	False Positive	Dương tính giả
TN	True Negative	Âm tính thực
TP	True Positive	Dương tính thực
ROC	Receiver Operating Characteristic	Đặc tính vận hành nhận
XGBoost	Extreme Gradient Boosting	Tăng cường gradient cực đại
EfficientNet	Efficient Neural Network	Mạng nơ-ron hiệu quả
F1-Score	F1 Score	Chỉ số F1
IoT	Internet of Things	Internet vạn vật
HOG	Histogram of Oriented Gradients	Biểu đồ độ dốc định hướng

DANH MỤC CÁC HÌNH VẼ

Hình 1.1	Phân loại tổn thương trên da	9
Hình 2.1	Mô hình chẩn đoán ung thư da sử dụng học sâu	22
Hình 2.2	Mô hình học chuyển giao.	31
Hình 2.3	Mô hình trích xuất đặc trưng từ ảnh tổn thương da.	32
Hình 2.4	Kiến trúc mô hình mạng học sâu kết hợp	37
Hình 2.5	Kiến trúc mô hình mạng CNN 1	39
Hình 2.6	Kiến trúc mô hình mạng CNN 2	40
Hình 2.7	Kiến trúc mô hình mạng CNN 3	43
Hình 3.1	Biểu đồ đường cong ROC và AUC	52
Hình 3.2	Một số ảnh chụp trong tập VNCancer.	54
Hình 3.3	Tiền xử lý ảnh da liễu.	55
Hình 3.4	So sánh biểu đồ Accuracy và Loss cho mô hình Xception với batch size = 8 và epoch = 100.	64
Hình 3.5	Hiệu năng của mô hình Xception với batch size 8 và 100 epochs: (a) Ma trận nhầm lẫn, (b) Biểu đồ Precision-Recall, và (c) Biểu đồ ROC cho từng lớp.	65
Hình 3.6	Hai đồ thị biểu diễn quá trình huấn luyện và đánh giá mô hình MobileNetV2. (a) Đồ thị Loss giữa Train và Validation, (b) Đồ thị Accuracy giữa Train và Validation.	66
Hình 3.7	Hiệu năng của mô hình MobileNetV2 trong việc chẩn đoán ung thư da.	66
Hình 3.8	Quá trình huấn luyện mô hình DenseNet121 với batch size 32 và 10 epochs. Đồ thị a) thể hiện độ chính xác và đồ thị b) thể hiện hàm mất mát trên cả tập huấn luyện và tập kiểm tra.	67
Hình 3.9	Các chỉ số hiệu năng của mô hình DenseNet121 với batch size = 32 và 10 epoch.	68

Hình 3.10 Các chỉ số hiệu năng của mô hình DenseNet121 với batch size = 32 và 10 epoch.	69
Hình 3.11 Kết quả đánh giá hiệu suất của mô hình InceptionV3 với batch size 16 và 200 epochs.	70
Hình 3.12 Biểu đồ độ chính xác và độ mất mát của mô hình VGG16 qua 13 epochs với batch size 16.	71
Hình 3.13 Hiệu năng của mô hình VGG16 với batch size 16 và 200 epoch trên tập dữ liệu da liễu. Hình (a) là ma trận nhầm lẫn chuẩn hóa, cho thấy độ chính xác của từng nhãn bệnh. Hình (b) là đường cong Precision-Recall, biểu diễn mối quan hệ giữa độ chính xác và độ nhạy. Hình (c) là đường cong ROC với diện tích AUC, đánh giá khả năng phân loại của mô hình dựa trên các ngưỡng khác nhau.	72
Hình 3.14 So sánh giữa độ chính xác và tổn thất trong quá trình huấn luyện và đánh giá mô hình.	73
Hình 3.15 Biểu diễn kết quả phân loại với mô hình EfficientNetB0.	74
Hình 3.16 Biểu đồ thể hiện quá trình huấn luyện và đánh giá của các mô hình Bộ học cơ sở với batch size 16 và 10 epochs.	78
Hình 3.17 Ma trận nhầm lẫn của các mô hình đề xuất với batch size 16 và 10 epochs.	79
Hình 3.18 Biểu đồ ROC cho từng lớp của các mô hình với batch size 16 và 10 epochs.	80

DANH MỤC CÁC BẢNG BIỂU

Bảng 2.1	Tổng quan về các mô hình CNN được sử dụng.....	30
Bảng 2.2	Kiến trúc mô hình CNN 1	38
Bảng 2.3	Kiến trúc mô hình CNN 2	41
Bảng 2.4	Cấu trúc mô hình CNN 3.....	44
Bảng 3.1	Số lượng mẫu cho mỗi lớp trong tập dữ liệu VNCancer.....	53
Bảng 3.2	Kết quả thực nghiệm các mô hình cơ sở với batch size = 8. Trong đó: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Acc = Train Accuracy, Val Acc = Val Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T= Time (in seconds).	59
Bảng 3.3	Bảng so sánh chỉ số hiệu năng của các mô hình với batch size = 16. Trong đó: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Loss = Train Loss, Val Loss = Validation Loss, Tr Acc = Train Accuracy, Val Acc = Validation Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T = Time Taken (in seconds).....	61
Bảng 3.4	Kết quả thực nghiệm các mô hình cơ sở với batch size = 32. Trong đó: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Loss = Train Loss, Val Loss = Validation Loss, Tr Acc = Train Accuracy, Val Acc = Validation Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T = Time (in seconds).	62
Bảng 3.5	Kết quả so sánh hiệu năng các mô hình học sâu. Các tiêu chí viết tắt: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Loss = Train Loss, Val Loss = Validation Loss, Tr Acc = Train Accuracy, Val Acc = Validation Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T = Time Taken (in seconds). Mô hình có hiệu năng tốt nhất được in đậm.	64

Bảng 3.6 Kết quả thực nghiệm của mô hình Stacked Ensemble. Trong đó: Bs là Batch Size, Ep là Epoch, Acc là độ chính xác (Accuracy), Re là Recall, Pre là độ chính xác của mô hình (Precision), Sens là độ nhạy (Sensitivity), Spec là độ đặc hiệu (Specificity), F1 là điểm F1 (F1 Score), AUC là diện tích dưới đường cong ROC (Area Under Curve), F1 Macro, F1 Micro, và F1 Weighted đại diện cho các chỉ số F1 tương ứng trên từng tập lớp. 75

Bảng 3.7 So sánh hiệu năng giữa các mô hình học sâu cơ sở và mô hình kết hợp Stacked Ensemble. Các tiêu chí viết tắt: BS = Batch Size, Ep = Epoch, Acc = Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, AUC = Area Under Curve, Sen = Sensitivity, Spec = Specificity, T = Time Taken (in seconds). Mô hình có hiệu năng tốt nhất được in đậm. 81

MỞ ĐẦU

1. Lý do chọn đề tài

Việc chẩn đoán bệnh ung thư da là một thách thức lớn trong y học hiện đại, đặc biệt với các loại ung thư như Melanoma, BCC, và SCC, vốn là những loại ung thư nguy hiểm và phổ biến nhất. Trong khi phương pháp chẩn đoán truyền thống dựa vào chuyên môn của các bác sĩ da liễu, thì hiện nay với sự phát triển của công nghệ học sâu, đặc biệt là các mô hình mạng nơ-ron tích chập (CNN), đã mở ra nhiều cơ hội mới cho việc phát hiện và phân loại các loại tổn thương da một cách chính xác và nhanh chóng. Tuy nhiên, các nghiên cứu hiện tại vẫn còn nhiều hạn chế về kích thước dữ liệu, sự mất cân bằng trong các lớp dữ liệu, cũng như hiệu suất của các mô hình chẩn đoán tự động chưa thực sự cao. Do đó, việc áp dụng và cải tiến các mô hình học sâu để nâng cao độ chính xác và hiệu suất của hệ thống chẩn đoán tự động là cần thiết.

Ngoài ra, tại Việt Nam, ung thư da vẫn chưa được nghiên cứu sâu rộng và chưa có nhiều giải pháp chẩn đoán tự động áp dụng cho thực tiễn. Điều này đòi hỏi phải có thêm các nghiên cứu để phát triển các hệ thống hỗ trợ chẩn đoán ung thư da nhằm hỗ trợ bác sĩ trong việc phát hiện và điều trị kịp thời cho bệnh nhân. Với mong muốn đóng góp vào lĩnh vực này, đề tài nghiên cứu này đã được lựa chọn nhằm phát triển một mô hình học sâu kết hợp, có khả năng cải thiện hiệu quả chẩn đoán và phân loại ung thư da.

2. Mục tiêu nghiên cứu của luận văn

Mục tiêu của luận văn này là nghiên cứu và phát triển các mô hình học sâu hiện đại nhằm hỗ trợ quá trình chẩn đoán bệnh ung thư da. Cụ thể, các mục tiêu chính bao gồm:

- Xây dựng và triển khai một hệ thống chẩn đoán ung thư da sử dụng các mạng nơ-ron tích chập (CNN) và các phương pháp học sâu tiên tiến.
- Nghiên cứu các kiến trúc mạng CNN phổ biến như GoogleNet, VGG16, ResNet,

và EfficientNet để phân loại các loại tổn thương da dựa trên hình ảnh.

- Áp dụng kỹ thuật học chuyển giao (transfer learning) và trích xuất đặc trưng để cải thiện hiệu suất mô hình trên các tập dữ liệu có quy mô nhỏ.
- Kết hợp các mô hình học sâu theo phương pháp tổ hợp (Stacked Ensemble) để nâng cao độ chính xác và khả năng tổng quát hóa của hệ thống chẩn đoán.
- Thực nghiệm và đánh giá hiệu năng của hệ thống trên tập dữ liệu thực tế, đồng thời phân tích và tối ưu hóa các tham số của mô hình nhằm đạt được kết quả tốt nhất.

Những mục tiêu này nhằm hướng đến việc xây dựng một giải pháp chẩn đoán tự động với hiệu suất cao, góp phần hỗ trợ các bác sĩ trong việc phát hiện sớm và chính xác các loại ung thư da, đặc biệt là melanoma, loại ung thư da nguy hiểm nhất.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: Đối tượng nghiên cứu của luận văn là các phương pháp học sâu và ứng dụng của chúng trong chẩn đoán bệnh da liễu. Đặc biệt, nghiên cứu tập trung vào việc sử dụng mô hình học sâu kết hợp với các phương pháp xử lý ảnh nhằm hỗ trợ quá trình phát hiện và chẩn đoán bệnh viêm da cơ địa và các bệnh da khác thông qua các hình ảnh y tế.

Phạm vi nghiên cứu: Phạm vi nghiên cứu của luận văn bao gồm các vấn đề chính sau:

- Tập trung vào các kỹ thuật học sâu, bao gồm các mô hình CNN, ResNet và các kiến trúc mạng khác, nhằm phân tích hình ảnh y tế da liễu.
- Áp dụng các phương pháp xử lý và phân tích hình ảnh để tối ưu hóa khả năng phát hiện bệnh từ các dữ liệu hình ảnh da.
- Nghiên cứu được thực hiện trên tập dữ liệu hình ảnh da được thu thập từ các bệnh viện lớn và các tổ chức y tế uy tín, nhằm đảm bảo tính xác thực và độ tin cậy của dữ liệu.

- Phạm vi nghiên cứu tập trung vào việc chẩn đoán bệnh ung thư da.

Mục tiêu của phạm vi nghiên cứu là đề xuất và phát triển mô hình hỗ trợ hiệu quả trong quá trình chẩn đoán bệnh da thông qua phân tích hình ảnh y tế, hướng đến việc nâng cao độ chính xác và khả năng ứng dụng trong thực tiễn.

4. Phương pháp nghiên cứu

Phương pháp nghiên cứu được xây dựng với các bước chính như sau:

- **Nghiên cứu tổng quan:**

- Tiến hành thu thập và tổng hợp các nghiên cứu trước đây liên quan đến chẩn đoán ung thư da và các kỹ thuật học sâu phổ biến, bao gồm mạng nơ-ron tích chập (CNN) và các phương pháp kết hợp (Ensemble Methods) như Stacked Ensemble.
- Đánh giá các phương pháp đã có và xác định khoảng trống nghiên cứu, từ đó đề xuất phương pháp mới.

- **Phân tích và xây dựng mô hình đề xuất:**

- **Mô hình Stacked Ensemble:** Dựa trên kết quả thực nghiệm từ các mô hình CNN cơ sở luận văn tiến hành kết hợp các mô hình này để xây dựng một mô hình Stacked Ensemble nhằm tối ưu hóa độ chính xác và khả năng dự đoán.
- **Các thành phần của mô hình:** Mỗi thành phần của mô hình bao gồm các mạng CNN được huấn luyện trên các đặc trưng hình ảnh ung thư da đã qua tiền xử lý và cân bằng dữ liệu. Các kết quả dự đoán của mỗi mô hình sau đó được tổng hợp để đưa ra kết quả cuối cùng.

- **Thực nghiệm và phân tích kết quả:**

- Tiến hành các thí nghiệm trên tập dữ liệu VNCancer đã qua tiền xử lý, đảm bảo rằng dữ liệu được chuẩn hóa và cân bằng trước khi đưa vào huấn luyện.
- So sánh hiệu năng giữa mô hình đề xuất với các mô hình cơ sở, dựa trên các chỉ số đánh giá như Accuracy, Precision, Recall, F1 Score, AUC, và thời gian tính toán.

- **Kết luận:**

- Phân tích, tổng hợp kết quả thực nghiệm, và đánh giá khả năng áp dụng mô hình vào thực tế. Mô hình Stacked Ensemble được chứng minh có hiệu suất cao trong phân loại ung thư da và có tiềm năng ứng dụng trong hệ thống chẩn đoán tự động.

5. Nội dung nghiên cứu

Nghiên cứu trong luận văn này tập trung vào việc phát triển và ứng dụng các kỹ thuật học sâu nhằm hỗ trợ chẩn đoán ung thư da. Dựa trên sự phát triển mạnh mẽ của các mô hình học sâu, cụ thể là các mạng nơ-ron tích chập (CNN) và các kỹ thuật kết hợp, nội dung nghiên cứu của luận văn bao gồm các phần chính như sau:

- **Chương 1:** Tổng quan các phương pháp chẩn đoán ung thư da hiện nay, bao gồm các kỹ thuật từ học máy đến học sâu, cùng với việc đánh giá các phương pháp trích xuất đặc trưng ảnh phục vụ chẩn đoán.
- **Chương 2:** Trình bày chi tiết về mô hình học sâu kết hợp, phương pháp trích xuất đặc trưng, và áp dụng mô hình học chuyên giao để cải thiện độ chính xác của việc phát hiện các loại ung thư da.
- **Chương 3:** Thực nghiệm và phân tích kết quả trên tập dữ liệu VNCancer, được thu thập từ Bệnh viện Da liễu Trung ương. Các kết quả bao gồm đánh giá mô hình thông qua các chỉ số như độ chính xác, độ nhạy, và AUC (diện tích dưới đường cong ROC).

Luận văn đi sâu vào việc thử nghiệm và so sánh giữa các mô hình học sâu cơ sở và phương pháp đề xuất nhằm nâng cao hiệu quả chẩn đoán ung thư da trong thực tiễn.

6. Những đóng góp chính của luận văn

Luận văn đã có những đóng góp quan trọng sau đây trong lĩnh vực chẩn đoán ung thư da sử dụng các mô hình học sâu:

- **Phát triển mô hình Stacked Ensemble:** Luận văn đã đề xuất và triển khai mô hình Stacked Ensemble, một phương pháp kết hợp các mô hình mạng nơ-ron tích chập (CNN) khác nhau như DenseNet, MobileNet, và Xception. Mô hình này đã cải thiện đáng kể độ chính xác trong việc chẩn đoán ung thư da, giúp tối ưu hóa khả năng dự đoán và giảm thiểu hiện tượng quá khớp.
- **Áp dụng kỹ thuật học sâu:** Luận văn đã áp dụng thành công các kỹ thuật học sâu tiên tiến như học chuyển giao (Transfer Learning) và trích xuất đặc trưng để cải thiện hiệu suất của hệ thống chẩn đoán ung thư da. Việc kết hợp giữa các mô hình học sâu với phương pháp tăng cường dữ liệu đã nâng cao khả năng phân loại tổn thương da, đặc biệt đối với các loại ung thư như melanoma, BCC, và SCC.
- **Cải thiện tốc độ tính toán:** Mô hình đề xuất không chỉ mang lại độ chính xác cao hơn mà còn giảm thiểu thời gian huấn luyện, góp phần quan trọng trong việc ứng dụng các mô hình chẩn đoán tự động vào thực tiễn, nâng cao khả năng chăm sóc sức khỏe nhanh chóng và hiệu quả.
- **Mở rộng tiềm năng ứng dụng:** Nghiên cứu đã mở ra những tiềm năng lớn trong việc ứng dụng các mô hình học sâu vào hệ thống chẩn đoán tự động cho bệnh ung thư da, đồng thời đưa ra hướng đi mới cho việc phát triển các giải pháp tương tự trong các lĩnh vực khác của y học.
- **Sử dụng tập dữ liệu thực nghiệm phong phú:** Luận văn đã sử dụng tập dữ liệu VNCancer với các kỹ thuật tiền xử lý và cân bằng dữ liệu, đảm bảo tính khách quan và độ tin cậy cao trong các kết quả thực nghiệm. Điều này tạo ra cơ sở dữ liệu chuẩn để so sánh và phát triển thêm các nghiên cứu trong tương lai.

Những đóng góp này không chỉ có ý nghĩa về mặt lý thuyết mà còn mở ra các hướng nghiên cứu mới trong việc kết hợp các mô hình học sâu và y học, góp phần quan trọng trong lĩnh vực chăm sóc sức khỏe tự động.

7. Bố cục luận văn

Luận văn được trình bày thành các chương với nội dung như sau:

- **Chương 1: Tổng quan nghiên cứu chẩn đoán ung thư da sử dụng học sâu và đặc trưng sâu.** Chương này giới thiệu tổng quan về bệnh ung thư da, những thách thức trong chẩn đoán và phương pháp tiếp cận sử dụng các kỹ thuật học sâu trong y học, cụ thể là chẩn đoán ung thư da.
- **Chương 2: Mô hình học kết hợp hỗ trợ chẩn đoán ung thư da.** Nội dung chương này tập trung vào việc trình bày các mô hình học sâu như CNN, học chuyển giao (Transfer Learning), và các kiến trúc mạng nơ-ron tích chập phổ biến như GoogleNet, VGG, ResNet. Đặc biệt, chương này đề xuất mô hình kết hợp theo tầng (Stacked Ensemble) trong chẩn đoán bệnh ung thư da.
- **Chương 3: Thử nghiệm mô hình kết hợp các mạng nơ-ron tích chập chẩn đoán bệnh ung thư da.** Chương này tiến hành thực nghiệm các mô hình được đề xuất trên tập dữ liệu VNCancer, trình bày các bước tiền xử lý dữ liệu, đánh giá hiệu năng của các mô hình cơ sở, cũng như phân tích kết quả của mô hình Stacked Ensemble trong phân loại ung thư da.

Mỗi chương đều cung cấp các nội dung và các nhận xét quan trọng, giúp luận văn hoàn thiện quá trình chẩn đoán ung thư da tự động thông qua mô hình học sâu.

Phần kết luận và hướng phát triển: Phần Kết luận tóm tắt nghiên cứu, nêu bật những đóng góp của luận văn và gợi ý những hướng đi cho công việc trong tương lai.

CHƯƠNG 1: TỔNG QUAN NGHIÊN CỨU CHẨN ĐOÁN UNG THƯ DA SỬ DỤNG HỌC SÂU VÀ ĐẶC TRƯNG HỌC SÂU

Chương này sẽ giới thiệu tổng quan về tình trạng bệnh ung thư da và các kỹ thuật chẩn đoán hiện đại đang được sử dụng, đặc biệt là những phương pháp học sâu và trích xuất đặc trưng học sâu. Chương cũng sẽ đề cập đến các loại ung thư da phổ biến, các yếu tố nguy cơ, và tình hình nghiên cứu về ung thư da trên thế giới và tại Việt Nam. Ngoài ra, chương sẽ trình bày các phương pháp xử lý ảnh, học sâu, cùng các kỹ thuật tăng cường dữ liệu để phát hiện và phân loại ung thư da, đặc biệt là Melanoma, loại ung thư da nguy hiểm nhất.

1.1. Giới thiệu

Da là cơ quan lớn nhất trong cơ thể, nó bảo vệ cơ thể khỏi nhiễm trùng, nhiệt độ và tia tử ngoại, nhưng đối với sự sống của con người, căn bệnh nguy hiểm nhất là ung thư. Cơ thể con người có thể bị nhiều loại ung thư khác nhau, và ung thư da là một trong những khối u nguy hiểm và phát triển nhanh nhất. Một trong ba trường hợp ung thư được chẩn đoán là ung thư da và, theo Thống kê Quỹ Ung thư Da, một trong năm người Mỹ sẽ mắc ung thư da trong suốt cuộc đời của họ. Ở Hoa Kỳ, có hơn 3,5 triệu ca mới xuất hiện mỗi năm và con số này đang liên tục tăng lên [1].

1.1.1. Ung thư da ở Việt Nam

Ở Việt Nam, nhiều năm gần đây, chưa có công bố cụ thể về tỷ lệ mắc và tử vong Ung thư hắc tố da. Theo Phạm và cộng sự [2], tỷ lệ mắc là 0,3-0,4/100.000 dân. Ung thư hắc tố da là bệnh ít gặp, nhưng đến khám và điều trị tại bệnh viện K¹ tăng lên hàng năm. Do hiểu biết về bệnh này còn hạn chế nên trên 80% bệnh nhân đến khám ở giai đoạn muộn. Điều trị hiện nay, phẫu thuật là phương pháp điều trị chủ yếu khi bệnh ở giai đoạn tại chỗ, tại vùng, phương pháp điều trị bằng hóa chất kém hiệu quả,

¹<https://benhvienk.vn/>

mang tính chất điều trị triệu chứng khi bệnh ở giai đoạn muộn.

Trên Thế giới và Việt Nam, ba loại ung thư da này chiếm một tỷ lệ cao trong các ca ung thư da được chẩn đoán hàng năm. BCC đứng đầu với tỷ lệ chiếm gần 80% trong tổng số các trường hợp ung thư da được phát hiện, trong khi SCC và MM chiếm tỷ lệ cao hơn so với các loại ung thư da khác. Tình trạng ung thư da đang ngày càng diễn biến phức tạp tại Việt Nam, đặc biệt ở nhóm người tiếp xúc với ánh nắng mặt trời nhiều hoặc có tiền sử tiếp xúc với tia tử ngoại [3].

Nguyên nhân mắc phải các loại ung thư da này chủ yếu liên quan đến tiếp xúc với ánh nắng mặt trời, đặc biệt là tia tử ngoại (UV). Tia UV có khả năng gây tổn hại tế bào da, gây đột biến trong DNA và kích thích phát triển bất thường của các tế bào, dẫn đến hình thành khối u ung thư. Ngoài ra, các yếu tố di truyền, tuổi tác, hút thuốc lá và tiếp xúc với các chất gây kích thích khác cũng đóng vai trò trong nguy cơ mắc phải ung thư da.

Mặc dù tỷ lệ tử vong của ung thư da thấp hơn so với nhiều loại ung thư khác, việc nhận thức và tìm kiếm sớm các triệu chứng bất thường vẫn là yếu tố quan trọng để giảm thiểu nguy cơ mắc phải và nâng cao hiệu quả điều trị cho những người bị ung thư da. Sự chăm sóc sức khỏe da và tìm kiếm tư vấn y tế từ các chuyên gia đáng tin cậy là điều cần thiết để đảm bảo phòng ngừa và chăm sóc sức khỏe hiệu quả trong việc đối phó với tình trạng ung thư da ngày càng lan rộng.

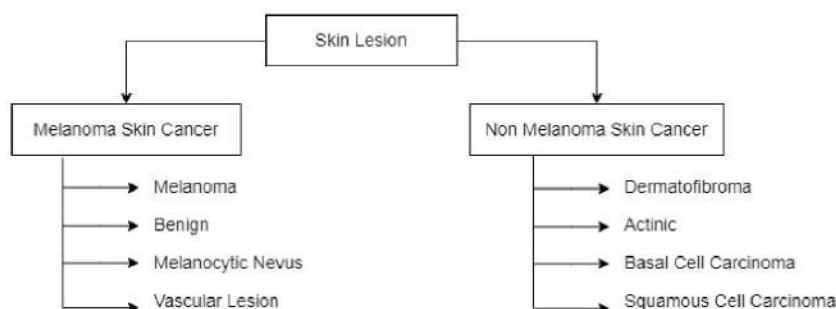
Đối diện với tình trạng phổ biến và nguy hiểm này, việc nghiên cứu về phòng ngừa, phát hiện sớm và điều trị hiệu quả ba loại ung thư da là vô cùng cấp thiết. Các công trình nghiên cứu và ứng dụng công nghệ hiện đại trong việc phân loại và chẩn đoán ung thư da đang đóng góp quan trọng vào việc nâng cao chất lượng chăm sóc y tế và cung cấp cơ hội tốt hơn để đối phó với tình trạng ung thư da ngày càng gia tăng.

1.1.2. Một số nguyên nhân gây ung thư da

Nhiều loại ung thư da bắt đầu từ tầng trên cùng của da. Ung thư da xảy ra khi các tế bào da chia tách và mở rộng một cách không kiểm soát. Thường thì, các tế bào da mới phát triển khi các tế bào cũ chết hoặc bị tổn thương. Khi quá trình này không hoạt động đúng cách, tế bào sẽ phát triển nhanh chóng theo một cách không gọn gàng. Đây là lý do tại sao những tế bào này được gọi là khối u. Ung thư da được gây ra bởi

một số yếu tố như uống rượu, hút thuốc, dị ứng, virus, thay đổi môi trường và tiếp xúc với tia cực tím (UV). Ngoài ra, ung thư da cũng có thể xuất hiện do các khối u bất thường trên cơ thể.

Ung thư da được chia thành hai nhóm là ung thư biểu mô melanoma (MM) và ung thư da không phải là melanoma được minh họa trong Hình 1.1 [4]. Ung thư da không phải melanoma là ung thư biểu mô tế bào gốc (Squamous Cell Carcinoma - SCC), và ung thư tế bào gốc trung tâm (Basal Cell Carcinoma - BCC). Ung thư biểu mô melanoma có thể nhanh chóng lan ra từ một cơ quan sang cơ quan khác [5].



Hình 1.1: Phân loại tổn thương trên da

Theo phân loại trên, ung thư da phổ biến nhất bao gồm BCC, SCC và MM, đang thu hút sự quan tâm và nghiên cứu đặc biệt từ cộng đồng y tế. Trên Thế giới và Việt Nam, ba loại ung thư da này chiếm một tỷ lệ cao trong các ca ung thư da được chẩn đoán hàng năm. BCC đứng đầu với tỷ lệ chiếm gần 80% trong tổng số các trường hợp ung thư da được phát hiện, trong khi SCC và MM chiếm tỷ lệ cao hơn so với các loại ung thư da khác. Tình trạng ung thư da đang ngày càng diễn biến phức tạp tại Việt Nam, đặc biệt ở nhóm người tiếp xúc với ánh nắng mặt trời nhiều hoặc có tiền sử tiếp xúc với tia tử ngoại. BCC và SCC là hai loại ung thư da không phải melanoma phổ biến nhất [6].

Ung thư da không phải là melanoma (NMSC) bao gồm ung thư tế bào gốc trung tâm (BCC) và ung thư biểu mô tế bào gốc (SCC), đại diện cho những loại ung thư phổ biến nhất trong dân số người da màu da trắng. Ung thư biểu mô melanoma có thể nhanh chóng lan ra từ một cơ quan sang cơ quan khác.

Rất nhiều người trên khắp thế giới đã bị ảnh hưởng bởi bệnh ung thư da. Bệnh này xảy ra khi các tế bào melanocyte trong cơ thể phát triển không bình thường và gây tổn thương cho các mô xung quanh da.

Có ba loại ung thư da phổ biến nhất là ung thư biểu mô melanoma, ung thư biểu mô tế bào gốc (Squamous Cell Carcinoma - SCC), và ung thư tế bào gốc trung tâm (Basal Cell Carcinoma - BCC).

Ung thư tế bào gốc trung tâm (BCC) được miêu tả lần đầu vào năm 1827 [7] và là loại ung thư da phổ biến nhất trên toàn thế giới, chiếm khoảng 80% trong số các ung thư da không phải là melanoma. Tỷ lệ mắc BCC đang gia tăng trên toàn cầu, với tốc độ tăng khoảng 1% mỗi năm [8]. Sự gia tăng này có thể do các thay đổi về môi trường hoặc các yếu tố nguy cơ lối sống. Yếu tố nguy cơ phổ biến nhất là tiếp xúc với tia cực tím (UV) từ ánh nắng mặt trời, đặc biệt là từ khi còn nhỏ tuổi hoặc do tiếp xúc không thường xuyên.

Ung thư da là loại bệnh phổ biến nhất trong tất cả các bệnh lý ở con người. Hơn 10 triệu người trên toàn thế giới được chẩn đoán mắc một loại ung thư da vào mỗi năm. Ung thư da xảy ra khi các tế bào bình thường trải qua biến đổi và phát triển và tăng trưởng mà không có sự kiểm soát bình thường. Ung thư tế bào gốc trung tâm (BCC) và ung thư biểu mô tế bào gốc (SCC) là hai loại ung thư da phổ biến nhất. BCC xảy ra thường xuyên hơn, chiếm khoảng 80% số ca. Ung thư tế bào gốc trung tâm và ung thư biểu mô tế bào gốc là hai loại ung thư da phổ biến nhất. Chúng bắt nguồn từ lớp trên cùng của da (biểu bì) và thường liên quan đến tiếp xúc với ánh nắng mặt trời.

Ung thư tế bào gốc trung tâm thường xuất phát từ các tế bào gốc trung tâm trong da: một loại tế bào tạo ra các tế bào da mới khi các tế bào cũ mất đi. BCC thường xuất hiện dưới dạng một cục u trong suốt nhẹ trên da, nhưng cũng có thể có nhiều hình dạng khác. Đây là một loại ung thư phát triển chậm và hiếm khi lan sang các bộ phận khác của cơ thể. Các tùy chọn điều trị ung thư tế bào gốc trung tâm bao gồm phẫu thuật lấy mô Mohs, phẫu thuật cắt, phẫu thuật điện, phẫu thuật đông lạnh và phẫu thuật laser. Đối với bệnh nhân mắc ung thư tế bào gốc trung tâm, có một lựa chọn khác: Trị liệu bức xạ bề mặt (SRT), là sự lựa chọn không phẫu thuật để loại bỏ ung thư da không phải là melanoma.

Ung thư biểu mô tế bào gốc (SCC) phát triển từ các tế bào biểu mô mỏng, phẳng tạo nên lớp bên ngoài của da. Các tế bào biểu mô được tìm thấy khắp cơ thể, và loại ung thư này có thể xuất hiện bất cứ nơi nào có tế bào biểu mô. Mặc dù loại ung thư

da này thường không nguy hiểm, một điểm khác biệt quan trọng giữa ung thư tế bào gốc trung tâm và ung thư tế bào gốc biểu mô là ung thư tế bào gốc biểu mô có khả năng phát triển sâu hơn vào các lớp da và lan sang các bộ phận khác của cơ thể. Mặc dù hiếm khi xảy ra, nó được coi là một loại ung thư nguy hiểm, nếu không được điều trị. Ung thư biểu mô tế bào gốc là loại ung thư bắt nguồn từ sự hoá biểu lớp hạch tế bào biểu mô và có thể lan toả tới các vùng khác của cơ thể. Nguyên nhân xuất hiện của ung thư biểu mô tế bào gốc có thể xảy ra trên bất kỳ khu vực nào của cơ thể, nhưng thường phổ biến ở các khu vực thường xuyên tiếp xúc với tia cực tím (UV) từ ánh nắng mặt trời, chẳng hạn như tai, môi, khuôn mặt, cổ, đầu, cánh tay và chân [9].

Ung thư biểu mô tế bào gốc là kết quả của việc tiếp xúc lâu dài với tia cực tím (UV) từ ánh nắng mặt trời hoặc các máy tạo nắng. Các nguyên nhân khác có thể bao gồm da sáng, tiền sử bị cháy nắng, tiền sử các vết thương da tiền ung thư hoặc ung thư da, hoặc hệ thống miễn dịch yếu. Triệu chứng của ung thư biểu mô tế bào gốc thường xuất hiện dưới dạng da nứt nẻ, da sưng và những khối u như nốt ruồi. Nói chung, bất kỳ sự thay đổi nào về mô da hiện tại, chẳng hạn như vết thương hoặc tổn thương không lành, hoặc sự xuất hiện của những khối u mới trên da, nên được tham vấn với bác sĩ ngay lập tức. Việc điều trị ung thư biểu mô tế bào gốc phụ thuộc vào kích thước, vị trí và tính chất của khối u. Các phương pháp điều trị truyền thống - bao gồm cả phẫu thuật Mohs - thường đòi hỏi cắt xén qua mô da khỏe mạnh, dẫn đến thời gian phục hồi lâu và vết sẹo không đẹp.

Melanoma có tỷ lệ tử vong cao hơn so với các loại ung thư da khác. Đây là loại ung thư nguy hiểm nhất cần được phát hiện sớm và chủ yếu do tia tử ngoại gây ra. Theo Hiệp hội Ung thư Hoa Kỳ, sẽ có hơn 1 triệu trường hợp mới của melanoma vào năm 2020, với khoảng 6.000 trường hợp tử vong [10]. Tương tự, theo Hội nghị Ung thư da lần thứ 20, melanoma da là tế bào ác tính phổ biến nhất ở cả nam và nữ. Năm 2020, đã phát hiện có 300.000 trường hợp mới được báo cáo. Số lượng tử vong ước tính tại Hoa Kỳ đã tăng gần đây so với số trường hợp được báo cáo. Tổ chức Y tế Thế giới (WHO) báo cáo rằng tia tử ngoại từ ánh nắng mặt trời gây tử vong cho 60.000 người hàng năm. Trong số đó, melanoma gây tử vong cho 48.000 người và ung thư da gây tử vong cho 12.000 người [11]. Nếu không phát hiện sớm, melanoma có thể lan

rộng sang các cơ quan quan trọng như gan, xương, phổi và não, ảnh hưởng lớn đến cuộc sống của bệnh nhân ung thư da với sự khó khăn và đau đớn. Do đó, việc phát hiện sớm melanoma đóng vai trò quan trọng trong chẩn đoán ung thư da.

1.1.3. Một số kỹ thuật gần đây trong việc chẩn đoán ung thư da

Trong nhiều thập kỷ qua, thủ tục nội soi đã là phương pháp cơ bản nhất được sử dụng để xem xét ung thư da cho việc điều trị sớm. Mặc dù nó mang lại kết quả nhất quán, các phương pháp thay thế như bảng kiểm tra bảy điểm và quy tắc đối xứng, biên, màu sắc, đường kính (ABCD) đã được chứng minh là hiệu quả hơn. Tuy nhiên, những phương pháp này vẫn yêu cầu sự chuyên môn của bác sĩ da liễu để phát hiện ung thư một cách chính xác.

Các phương pháp đánh giá sự phát triển của khối u da để dự đoán tiên lượng ung thư biểu mô melanoma đã phát triển qua vài thập kỷ qua. Trước những năm 1980, ung thư biểu mô melanoma thường được nhận biết thông qua việc quan sát bằng mắt thường các thay đổi về đặc điểm nổi bật của nốt ruồi, như kích thước lớn, chảy máu hoặc loét. Trong trường hợp các tổn thương đáng nghi ngờ, việc chẩn đoán tiến hành bằng cách lấy mẫu tế bào của tổn thương để tiến hành phân tích tiếp. Phương pháp xâm lấn này vẫn là phương pháp chính xác nhất cho việc chẩn đoán ung thư biểu mô melanoma, nhưng đòi hỏi sử dụng nhân viên được đào tạo và thiết bị đắt tiền. Trong thời kỳ đó, việc tiên lượng sớm khó khăn do thiếu sự tiến bộ về phần cứng và phần mềm hình ảnh.

Trong những năm gần đây, bác sĩ da liễu đã tích hợp hình ảnh kính hiển vi và thông qua hình ảnh dermoscopy để chẩn đoán ung thư da. Các phương pháp không xâm lấn này dần được áp dụng, do thiết bị rẻ hơn và độ chính xác cao. Dermoscope là một công cụ quang học sử dụng nguồn sáng để loại bỏ các phản xạ từ bề mặt da [12]. Điều này cho phép tiếp cận các cấu trúc và màu sắc chi tiết của tổn thương. Thiết bị này được kết nối với máy tính và chụp hình ảnh hoặc video của tổn thương, sau đó được sử dụng để chẩn đoán. Một độ nhạy đạt 89% khi sử dụng phương pháp dermoscopy; cải thiện so với 70-85% cho phương pháp kiểm tra bằng mắt thường.

Tuy nhiên, chi phí là hạn chế cho việc sử dụng chung vì dermoscope phải được kết nối với thiết bị đắt tiền (như kính hiển vi ba chiều) để đánh giá tính ác tính của tổn thương. Bác sĩ da liễu phân tích những hình ảnh này thông qua kiểm tra thị giác. Tuy nhiên, quá trình kiểm tra thị giác và kiểm tra các vết thương trên da mất nhiều thời gian và đòi hỏi chuyên môn, khiến nó trở thành một quy trình phức tạp và kéo dài.

Với nỗ lực nhằm phổ biến việc chẩn đoán bệnh da thông qua việc sử dụng các giải pháp giá rẻ hơn so với kính hiển vi ba chiều. Một phương pháp được gọi là "teledermatology di động" [13] đã sử dụng điện thoại di động để chụp ảnh kỹ thuật số của tổn thương, nhưng cần phải kết nối với các thiết bị dermoscope cầm tay để bù đắp cho các ống kính chất lượng kém trong các thiết bị di động thế hệ đầu tiên. Những hình ảnh thu được được chuyển tới các chuyên gia tư vấn qua mạng riêng ở các địa điểm xa để phân tích và đánh giá. Hai lĩnh vực cải tiến liên quan đến: (i) phần cứng tốt hơn để chụp ảnh độ phân giải cao và (ii) hệ thống chẩn đoán hỗ trợ máy tính thông minh (CAD) để xác định chính xác ung thư biểu mô melanoma từ hình ảnh dermoscope. Hầu hết các hệ thống CAD đã được báo cáo trước đây hoạt động trên máy tính cá nhân (PC) hoặc máy trạm và hỗ trợ các bác sĩ xác định tổn thương ung thư ở giai đoạn sớm để đưa vào phương pháp điều trị ngay lập tức. Những hệ thống CAD này thường được thử nghiệm trên hình ảnh tổn thương dermoscope hoặc viễn xạ hiếm ánh, mặc dù chúng có thể tích hợp vào điện thoại thông minh.

Hiện nay, điện thoại di động được trang bị sức mạnh xử lý cao, dung lượng lưu trữ nhiều hơn, cảm biến hình ảnh độ phân giải cao và bộ nhớ lớn. Điều này nên cho phép điện thoại di động chụp ảnh và thực hiện các nhiệm vụ tính toán lớn trực tiếp trên thiết bị.

Hình ảnh các tổn thương trên da được chụp bằng máy ảnh điện thoại thông minh. Sau đó, một phương pháp học sâu được thực hiện sử dụng các tham số của mô hình đã được huấn luyện trước để đưa ra các chẩn đoán. Việc phát triển các ứng dụng chẩn đoán da toàn diện như vậy được dự kiến sẽ đạt đà phát triển mạnh mẽ trong những năm tới, do vấn đề chăm sóc sức khỏe, chi phí đắt đỏ cho việc thăm khám tại bệnh viện và tỷ lệ tử vong cao do ung thư biểu mô melanoma.

1.2. Nghiên cứu liên quan

Trong những năm gần đây, số lượng nghiên cứu về phát triển các hệ thống hỗ trợ máy tính trong lĩnh vực y tế đã tăng lên đáng kể. Đặc biệt, các nghiên cứu liên quan đến phát hiện và phân loại tổn thương da thông qua học sâu và xử lý ảnh đã đạt được nhiều tiến bộ quan trọng. Các kỹ thuật hiện đại được áp dụng để giải quyết các thách thức trong việc chẩn đoán ung thư da, từ cải thiện độ chính xác của mô hình đến giải quyết vấn đề mất cân bằng lớp dữ liệu.

Farhat và cộng sự [14] đã đề xuất một phương pháp phân loại nhiều lớp cho các vết nám trên da, sử dụng kết hợp đặc trưng học sâu và máy học cục bộ. Quy trình bao gồm năm bước: thu thập hình ảnh, cải thiện độ tương phản, trích xuất đặc trưng học sâu bằng học chuyển tiếp, chọn lọc đặc trưng tốt nhất thông qua phương pháp tối ưu cá voi và thông tin tương trợ, cuối cùng là phân loại bằng máy học cục bộ. Kết quả trên bộ dữ liệu HAM10000 và ISIC2018 cho thấy độ chính xác lần lượt là 93,40% và 94,36%.

Jeremy và cộng sự [15] đã đề xuất phương pháp phân loại dựa trên tiêu chí kiểm tra bảy điểm, giúp tạo ra các vector đặc trưng đa phương thức, hỗ trợ việc tra cứu hình ảnh và xác định các khu vực lâm sàng quan trọng.

Vijayalakshmi và cộng sự [16] đã áp dụng kỹ thuật xử lý ảnh nhằm loại bỏ các yếu tố gây nhiễu như lông, bóng và ánh sáng phản chiếu trong quá trình tiền xử lý. Tiếp theo, quá trình phân đoạn và trích xuất đặc trưng đã được thực hiện, với kết quả phân loại đạt độ chính xác 85% trên bộ dữ liệu ISIC, chứng minh tính khả thi của phương pháp này trong việc phát hiện melanoma.

Arslan và cộng sự [17] đã đề xuất phương pháp phân loại và phân đoạn tổn thương da thành lành tính hoặc ác tính, sử dụng các đặc trưng như GLCM, HOG và màu sắc từ hình ảnh phân đoạn. Hệ thống này kết hợp kỹ thuật cân bằng SMOTE và áp dụng các bộ phân loại như Rừng ngẫu nhiên, đạt độ chính xác cao nhất là 93,89% trên bộ dữ liệu ISIC-ISBI 2016.

Những nghiên cứu này đã khẳng định tầm quan trọng của việc áp dụng học sâu và xử lý ảnh trong chẩn đoán ung thư da, cung cấp nền tảng cho các hệ thống hỗ trợ chẩn đoán chính xác và hiệu quả.

1.2.1. Phát hiện Ung thư da Melanoma sử dụng đặc trưng bằng kỹ thuật Xử lý ảnh

Melanoma là loại ung thư da nguy hiểm nhất và việc phát hiện sớm có vai trò vô cùng quan trọng trong việc giảm tỷ lệ tử vong. Nhiều nghiên cứu đã tập trung vào việc phát triển các kỹ thuật xử lý ảnh nhằm cải thiện độ chính xác trong phát hiện và chẩn đoán melanoma. Phương pháp kết hợp giữa xử lý ảnh và học sâu đã mang lại những kết quả đáng khích lệ trong việc phân loại các tổn thương da phức tạp.

Farhat và cộng sự [15] đã đề xuất phương pháp phân loại đa lớp, trong đó đặc trưng học sâu được trích xuất từ hình ảnh tổn thương da, sau đó các đặc trưng tốt nhất được chọn lọc bằng phương pháp tối ưu cá voi và thông tin tương tự. Hệ thống phân loại sử dụng máy học cực đoan cho kết quả chính xác cao trên các bộ dữ liệu công khai.

Vijayalakshmi và cộng sự [16] đã áp dụng kỹ thuật xử lý ảnh nhằm loại bỏ các yếu tố gây nhiễu như lông, bóng và ánh sáng phản chiếu trong quá trình tiền xử lý. Tiếp theo, quá trình phân đoạn và trích xuất đặc trưng đã được thực hiện, với kết quả phân loại đạt độ chính xác 85% trên bộ dữ liệu ISIC, chứng minh tính khả thi của phương pháp này trong việc phát hiện melanoma.

Arslan và cộng sự [17] đã phát triển một hệ thống phân loại và phân đoạn tổn thương da sử dụng các đặc trưng như GLCM, HOG và màu sắc. Các đặc trưng này được giảm kích thước bằng phương pháp PCA và sau đó được phân loại bằng các bộ phân loại như Rừng ngẫu nhiên. Hệ thống này đã cho thấy sự vượt trội trong phân đoạn và phân loại tổn thương da melanoma, với độ chính xác đạt 93,89%.

Những nghiên cứu trên đã cho thấy sự hiệu quả của các phương pháp xử lý ảnh và học sâu trong việc phát hiện sớm ung thư da, đặc biệt là melanoma, từ đó mở ra tiềm năng to lớn cho việc ứng dụng các công nghệ hiện đại trong y học.

1.2.2. Phát hiện ung thư Melanoma bằng học sâu và đặc trưng sâu

Học sâu đã chứng minh vai trò quan trọng trong việc phát hiện và nhận dạng ung thư da, đặc biệt là melanoma. Các mô hình nổi tiếng như VGG16, VGG19, Xception, ResNet, InceptionV3, EfficientNet-b0, và MobileNetV2 đã được ứng dụng trong nhiều nghiên cứu liên quan đến chẩn đoán ung thư da.

Những mô hình này giúp trích xuất các đặc trưng sâu từ hình ảnh, mang lại kết quả chính xác hơn trong phân loại tổn thương da, đồng thời tối ưu hóa quy trình chẩn đoán.

Một nghiên cứu nổi bật do Farhat và cộng sự thực hiện đã kết hợp các kỹ thuật học sâu và máy học để phân loại đa lớp các tổn thương da. Nghiên cứu này không chỉ tập trung vào việc thu thập và gia tăng dữ liệu mà còn tối ưu hóa quá trình lựa chọn đặc trưng tốt nhất từ các đặc trưng đã trích xuất. Thử nghiệm được tiến hành trên các tập dữ liệu HAM10000 và ISIC2018, với kết quả rất ấn tượng, đạt độ chính xác lần lượt là 93,40% và 94,36%.

Ngoài ra, các nghiên cứu khác cũng cho thấy tiềm năng to lớn của học sâu trong chẩn đoán ung thư da. Oscar và cộng sự đã phát triển một mô hình dựa trên việc phân đoạn ảnh, trích xuất đặc trưng từ các mạng CNN huấn luyện trước, và sử dụng SVM để phân loại. Kết quả nghiên cứu cho thấy mô hình này cải thiện đáng kể độ chính xác, đặc biệt khi sử dụng phân đoạn ảnh trước khi tiến hành trích xuất đặc trưng. AlexNet kết hợp với SVM đã cho kết quả tốt nhất với độ chính xác lên đến 90,34% [18].

Việc tăng cường dữ liệu cũng được xem là một yếu tố quan trọng trong việc cải thiện độ chính xác của các mô hình học sâu. Theo nghiên cứu của Bhuvaneshwari và cộng sự, các mô hình sử dụng kỹ thuật tăng cường dữ liệu không chỉ học được nhiều đặc trưng phân biệt hơn mà còn cho kết quả đáng tin cậy hơn. Nghiên cứu này đã chỉ ra rằng, khi kết hợp với kỹ thuật xác thực chéo k-fold, mô hình CNN đã đạt độ chính xác ấn tượng lên đến 95,18% trên tập dữ liệu HAM10000 [19].

Bên cạnh đó, Muhammad và cộng sự đã đề xuất một hệ thống kết hợp giữa phân đoạn và trích xuất đặc trưng sâu. Sử dụng MASK-RCNN để phân đoạn hình ảnh trước khi đưa vào mô hình DeepNet, nhóm nghiên cứu đã đạt được những kết quả đáng chú ý. Các đặc trưng sau khi trích xuất được tối ưu hóa bằng SVM, giúp cải thiện đáng kể độ chính xác của mô hình trên các bộ dữ liệu như ISBI2016 và ISBI2017 [20].

Những kết quả này cho thấy rằng các mô hình học sâu không chỉ mang lại độ chính xác cao trong việc phát hiện sớm melanoma mà còn giúp tối ưu hóa quy trình chẩn đoán. Các mô hình như DenseNet, Inception-ResNet-V2 và InceptionV3 tiếp tục được cải tiến và thử nghiệm với các phương pháp học chuyển giao, mang lại kết quả khả quan trên các bộ dữ liệu lớn như HAM10000 và ISIC2019 [21]. Nhìn chung, sự

phát triển của học sâu đã mở ra nhiều cơ hội mới trong việc chẩn đoán và điều trị ung thư da, đặc biệt là melanoma, giúp cải thiện chất lượng chăm sóc sức khỏe và giảm thiểu nguy cơ tử vong từ căn bệnh này.

1.2.3. Phát hiện ung thư Melanoma bằng học sâu và tăng cường dữ liệu

Việc phát hiện ung thư Melanoma sử dụng học sâu đã gặp phải hai thách thức chính khi làm việc với các tập dữ liệu công khai: kích thước tập dữ liệu nhỏ và sự mất cân bằng giữa các lớp, cũng như sự xuất hiện của các vùng che phủ trong ảnh dermoscopic. Nhằm giải quyết những vấn đề này, Devansh và cộng sự [22] đã tập trung vào việc xử lý các ảnh da dermoscopic, vì chúng cung cấp khả năng chẩn đoán cao hơn. Nhóm nghiên cứu đã đề xuất một phương pháp tinh chỉnh để loại bỏ các vùng che phủ và sử dụng kỹ thuật tăng cường dữ liệu nhằm tạo ra các hình ảnh mới cho lớp Melanoma và Seborrheic Keratosis, qua đó cân bằng lại tập dữ liệu. Kết quả cho thấy phương pháp này đã mang lại giá trị ROC-AUC là 0.880 cho lớp Melanoma, cao hơn so với giá trị 0.874 của phương pháp đoạt giải ISIC, cho thấy một cải thiện đáng kể 4%.

Bên cạnh đó, Mehak và cộng sự [23] đã thực hiện nghiên cứu trên tập dữ liệu HAM10000, trong đó họ áp dụng ba phương pháp tăng cường dữ liệu phổ biến gồm xoay 90 độ, lật phải-trái và lật lên-xuống. Với các mô hình ResNet-50 và ResNet-101, nghiên cứu đã đạt được độ chính xác lên đến 91,7%, chứng minh hiệu quả của việc gia tăng dữ liệu trong việc cải thiện kết quả phân loại ung thư da.

Ngoài các phương pháp truyền thống, Generative Adversarial Networks (GAN) cũng được ứng dụng nhằm tạo ra các mẫu ảnh nhân tạo khó phân biệt với ảnh thực tế, giúp bổ sung số lượng mẫu dữ liệu cho quá trình huấn luyện. Tuy nhiên, nghiên cứu của Bissoto và cộng sự [24] đã chỉ ra rằng mặc dù GAN có tiềm năng trong việc tổng hợp dữ liệu, việc cải thiện hiệu suất phân loại dựa trên tăng cường dữ liệu bằng GAN không phải lúc nào cũng đạt được kết quả đáng tin cậy. Thực tế, nghiên cứu này đã cho thấy rằng việc sử dụng GAN trong bối cảnh chẩn đoán da liễu cần được thận trọng, bởi hiệu quả của nó không phải lúc nào cũng được đảm bảo.

Những nghiên cứu này cho thấy rằng, mặc dù việc tăng cường dữ liệu bằng các phương pháp truyền thống như xoay và lật ảnh có thể cải thiện hiệu suất mô hình,

nhưng việc ứng dụng các kỹ thuật mới như GAN vẫn còn nhiều thách thức và cần thêm nhiều nghiên cứu để xác định được giá trị thực sự của chúng trong việc phát hiện ung thư da.

1.2.4. Tổ hợp các mô hình học sâu để cải thiện việc phát hiện ung thư da

Ung thư da ác tính, đặc biệt là melanoma, là một trong những dạng ung thư nguy hiểm nhất do sự phát triển bất thường của tế bào melanocyte. Các mô hình mạng nơ-ron tích chập sâu (CNN) đã đạt được những tiến bộ vượt bậc trong việc chẩn đoán tự động ung thư da thông qua hình ảnh dermoscopy. Tuy nhiên, tính "hộp đen" của các mô hình này vẫn là một thách thức lớn, gây ra khó khăn trong việc áp dụng chúng rộng rãi trong lâm sàng vì sự thiếu minh bạch trong việc giải thích kết quả, làm giảm lòng tin của các bác sĩ lâm sàng.

Để khắc phục những hạn chế này, các nhà nghiên cứu đã phát triển các phương pháp tổ hợp (ensemble) nhằm cải thiện hiệu suất và tăng cường khả năng giải thích của mô hình. Một ví dụ điển hình là nghiên cứu của Shorfuzzaman và cộng sự [25], trong đó họ đề xuất một mô hình stack CNN để phát hiện sớm ung thư da melanoma. Mô hình này kết hợp nhiều CNN con, với mỗi mô hình thực hiện một nhiệm vụ phân loại riêng lẻ, và sau đó các kết quả được tổng hợp lại thông qua một mô hình meta học nhằm tạo ra dự đoán cuối cùng với độ chính xác cao hơn. Hơn nữa, để tăng cường tính minh bạch, phương pháp này còn phát triển các bản đồ nhiệt, giúp trực quan hóa các vùng ảnh da có liên quan đến tổn thương, giúp các bác sĩ hiểu rõ hơn về quyết định của mô hình. Kết quả đạt được cho thấy mô hình stack không chỉ nâng cao độ chính xác, mà còn cung cấp khả năng giải thích tốt hơn cho người dùng.

Một hướng tiếp cận khác dựa trên việc kết hợp nhiều mô hình CNN đã được Popescu và cộng sự [26] phát triển. Họ sử dụng nhiều mạng nơ-ron khác nhau như AlexNet, GoogLeNet, ResNet, và DenseNet để xây dựng một hệ thống thông minh tập thể. Hệ thống này không chỉ tích hợp hiệu suất từ từng mạng CNN mà còn giúp tối ưu hóa quyết định cuối cùng thông qua một mô-đun tổng hợp. Kết quả cho thấy hệ thống này cải thiện độ chính xác tổng thể, đồng thời giảm thiểu rủi ro phân loại sai nhờ vào việc tận dụng sự đa dạng từ các mô hình khác nhau.

Ngoài ra, Jiaqi và cộng sự [27] đã giới thiệu một phương pháp tổ hợp sử dụng

U-net để phân đoạn các vùng tổn thương trước khi áp dụng các mô hình phân loại. Việc tập trung vào các vùng cụ thể của tổn thương không chỉ giúp tăng cường độ chính xác trong nhận diện mà còn cải thiện hiệu suất tổng thể khi kết hợp kết quả từ nhiều mô hình phân loại khác nhau. Phương pháp này đã được kiểm chứng và cho thấy độ chính xác cao trên các bộ dữ liệu chuẩn như ISIC2017.

Bên cạnh các cách tiếp cận trên, Khurram và cộng sự [28] đã áp dụng một mô hình kết hợp tiên tiến giữa EfficientNetV2S và Swin-Transformer, giúp cải thiện đáng kể hiệu suất dự đoán ung thư da với độ chính xác vượt trội, đạt đến 99.10%. Mô hình hợp tác này cho thấy tiềm năng lớn trong việc phát triển các hệ thống chẩn đoán tự động với khả năng phát hiện chính xác và đáng tin cậy.

Từ các nghiên cứu liên quan, có thể thấy rằng việc kết hợp các mô hình học sâu khác nhau không chỉ nâng cao hiệu suất phân loại mà còn tăng cường khả năng giải thích kết quả, giúp cải thiện độ tin cậy trong các ứng dụng lâm sàng. Đặc biệt, các bộ dữ liệu lớn như HAM10000, ISIC2017, ISIC2018, và ISIC2019 đã được sử dụng rộng rãi để huấn luyện các mô hình học sâu nhằm đạt được kết quả tốt nhất. Việc sử dụng học chuyển giao với trọng số đã được huấn luyện trước trên ImageNet cũng là một phương pháp phổ biến để tận dụng hiệu quả của các mạng nơ-ron này.

Trong luận văn này, bộ dữ liệu ung thư da được thu thập từ Bệnh viện Da liễu Trung ương, bao gồm ba loại bệnh: BCC với 2765 ảnh, SCC với 899 ảnh, và MM với 202 ảnh. Bộ dữ liệu này có sự mất cân bằng lớn giữa các nhãn và được chụp bằng điện thoại, điều này tạo ra thách thức về chất lượng ảnh. Để khắc phục vấn đề này, luận văn đã áp dụng kỹ thuật tổ hợp theo tầng (Stacked Ensemble) nhằm cải thiện hiệu suất của mô hình trong việc phát hiện và phân loại ung thư da.

1.2.5. Tổ hợp theo tầng (Stacked Ensemble)

Thay vì sử dụng một mô hình đơn lẻ, mô hình học tổ hợp theo tầng (Stacked Ensemble) sẽ kết hợp nhiều mô hình CNN và Transformer khác nhau. Mỗi mô hình thành phần sẽ thực hiện việc huấn luyện và dự đoán độc lập. Sau đó, các kết quả dự đoán của các mô hình thành phần này sẽ được tổng hợp lại thông qua một mô hình meta học (meta-learner), chẳng hạn như Gradient Boosting hoặc XGBoost. Mô hình

meta sẽ học cách kết hợp các dự đoán từ các mô hình thành phần để tạo ra một kết quả cuối cùng với độ chính xác cao hơn.

Cụ thể, mô hình được xây dựng gồm hai tầng chính:

- **Tầng 1: Các mô hình thành phần (Base Learners):** Các mô hình CNN và Transformer sẽ được huấn luyện trên cùng một tập dữ liệu nhưng với cấu trúc và chiến lược học khác nhau. Mỗi mô hình này sẽ đưa ra dự đoán riêng lẻ cho tập dữ liệu đầu vào.
- **Tầng 2: Mô hình Bộ học meta (Meta-Learner):** Sau khi thu thập được các dự đoán từ các mô hình thành phần, một mô hình meta học như Gradient Boosting hoặc XGBoost sẽ được sử dụng để học cách kết hợp các dự đoán này. Tầng này sẽ huấn luyện trên các dự đoán của tầng 1 và tạo ra một dự đoán cuối cùng.

Kết luận chương 1

Chương 1 đã cung cấp một cái nhìn tổng quan về ung thư da và các phương pháp hiện đại được sử dụng để phát hiện và chẩn đoán căn bệnh này, đặc biệt là các kỹ thuật học sâu và trích rút đặc trưng. Ung thư da là một trong những loại ung thư phổ biến và nguy hiểm, với tỷ lệ mắc bệnh gia tăng nhanh chóng trên toàn thế giới. Ở Việt Nam, mặc dù chưa có con số thống kê chính thức, sự gia tăng số lượng bệnh nhân mắc ung thư da trong những năm gần đây đã đặt ra nhiều thách thức cho việc chẩn đoán và điều trị.

Với sự phát triển của công nghệ, đặc biệt là các mô hình học sâu, nhiều giải pháp chẩn đoán tự động đã được phát triển và ứng dụng, giúp cải thiện đáng kể độ chính xác và tốc độ phát hiện ung thư da. Các phương pháp như mạng nơ-ron tích chập (CNN), học chuyển giao, và các mô hình tổ hợp đã chứng minh khả năng phát hiện và phân loại hiệu quả các loại tổn thương da như melanoma, BCC, và SCC. Hơn nữa, việc áp dụng các kỹ thuật tăng cường dữ liệu đã giúp khắc phục những hạn chế về kích thước và sự mất cân bằng của tập dữ liệu, mang lại kết quả chính xác hơn trong các bài toán phân loại.

Một trong những đóng góp quan trọng của các nghiên cứu gần đây là việc sử dụng các mô hình tổ hợp, đặc biệt là mô hình tổ hợp theo tầng (Stacked Ensemble).

Cách tiếp cận này đã cho thấy tiềm năng lớn trong việc kết hợp các mô hình học sâu khác nhau để tăng cường độ chính xác và khả năng giải thích của các hệ thống chẩn đoán. Các mô hình tổ hợp không chỉ tối ưu hóa kết quả dự đoán mà còn giảm thiểu các điểm yếu của từng mô hình thành phần, từ đó nâng cao hiệu quả của hệ thống chẩn đoán tự động.

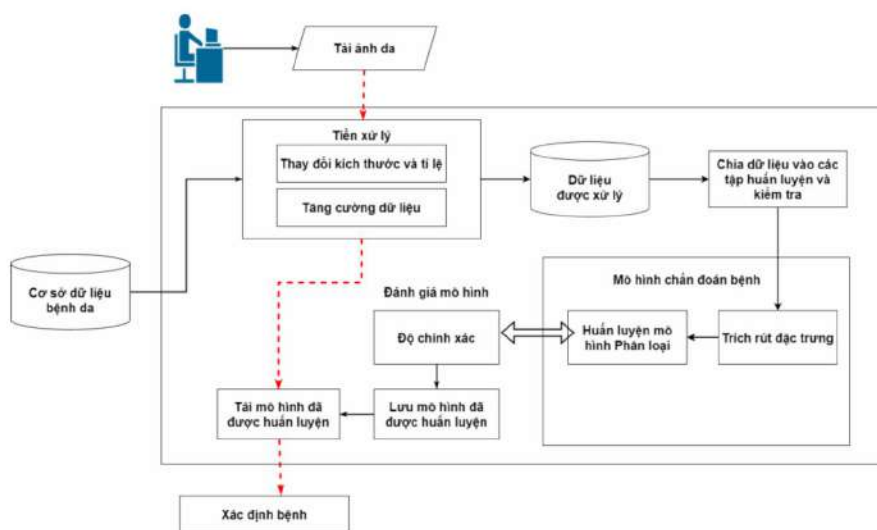
Nhìn chung, nghiên cứu và phát triển các mô hình học sâu trong chẩn đoán ung thư da đang mở ra những cơ hội mới trong y học, giúp giảm thiểu nguy cơ tử vong và cải thiện chất lượng chăm sóc sức khỏe. Trong các chương tiếp theo, chương sẽ trình bày chi tiết về phương pháp đề xuất và các kết quả thực nghiệm nhằm nâng cao hơn nữa hiệu suất của hệ thống chẩn đoán ung thư da bằng học sâu.

CHƯƠNG 2: MÔ HÌNH HỌC SÂU KẾT HỢP HỖ TRỢ CHẨN ĐOÁN UNG THƯ DA

Trong chương này, luận văn sẽ giới thiệu về việc áp dụng các mô hình học sâu trong chẩn đoán ung thư da. Bắt đầu từ quy trình chẩn đoán tổng thể, luận văn sẽ trình bày chi tiết các bước chính từ thu thập và tiền xử lý dữ liệu đến trích rút đặc trưng và huấn luyện mô hình. Đặc biệt, chương sẽ đi sâu vào việc sử dụng mạng nơ-ron tích chập (CNN) để phân loại hình ảnh ung thư da, từ đó cải thiện độ chính xác và hiệu suất của hệ thống chẩn đoán. Cuối cùng, luận văn sẽ giới thiệu mô hình StackNet, một phương pháp kết hợp các mạng CNN khác nhau nhằm tối ưu hóa khả năng dự đoán và ứng dụng vào việc xác định các loại ung thư da.

2.1. Kiến trúc hệ thống chẩn đoán ung thư da sử dụng học sâu

Quy trình chẩn đoán ung thư da sử dụng các kỹ thuật học sâu bao gồm các bước chính từ tiền xử lý dữ liệu cho đến đánh giá mô hình và xác định bệnh. Hình 2.1 mô tả chi tiết quy trình này.



Hình 2.1: Mô hình chẩn đoán ung thư da sử dụng học sâu

Bước 1: Thu thập dữ liệu

Dữ liệu ban đầu được thu thập từ các cơ sở dữ liệu bệnh da bao gồm các hình ảnh da

bị tổn thương. Đây là nguồn dữ liệu chính để phát triển mô hình học sâu chẩn đoán ung thư da. Các hình ảnh này có thể được chụp từ nhiều nguồn khác nhau như bệnh viện hoặc các nghiên cứu y khoa.

Bước 2: Tiền xử lý dữ liệu

Trong bước này, các hình ảnh da được tải lên hệ thống và trải qua quá trình tiền xử lý để đảm bảo tính đồng nhất về kích thước và tỉ lệ.

- **Thay đổi kích thước và tỉ lệ:** Các hình ảnh sẽ được điều chỉnh về kích thước phù hợp với yêu cầu đầu vào của mô hình học sâu.
- **Tăng cường dữ liệu:** Nhằm tăng cường tính đa dạng và tính đại diện của tập dữ liệu, các kỹ thuật như xoay, lật, hoặc điều chỉnh độ sáng của hình ảnh sẽ được áp dụng. Điều này giúp mô hình học sâu trở nên mạnh mẽ hơn đối với các biến đổi nhỏ trong hình ảnh.

Bước 3: Chia dữ liệu thành tập huấn luyện và kiểm tra

Sau khi dữ liệu đã được xử lý, nó sẽ được chia thành hai tập chính:

- **Tập huấn luyện:** Bao gồm phần lớn dữ liệu và được sử dụng để huấn luyện mô hình.
- **Tập kiểm tra:** Một phần nhỏ của dữ liệu sẽ được giữ lại để đánh giá hiệu suất của mô hình sau khi huấn luyện.

Bước 4: Trích rút đặc trưng và huấn luyện mô hình phân loại

Tiếp theo, dữ liệu được đưa vào mô hình học sâu, nơi các đặc trưng quan trọng từ hình ảnh da được trích rút tự động. Sau khi các đặc trưng đã được trích rút, mô hình phân loại sẽ được huấn luyện để dự đoán loại ung thư da dựa trên các đặc trưng này.

Bước 5: Đánh giá mô hình

Mô hình đã được huấn luyện sẽ được đánh giá dựa trên các tiêu chí về độ chính xác và khả năng phân loại chính xác các loại ung thư da. Nếu mô hình đạt được độ chính xác mong muốn, nó sẽ được lưu trữ để sử dụng trong các bước chẩn đoán tiếp theo.

Bước 6: Xác định bệnh

Sau khi mô hình đã được huấn luyện và lưu trữ, nó có thể được sử dụng để chẩn đoán

các trường hợp ung thư da mới. Khi một hình ảnh mới được tải lên hệ thống, mô hình đã được huấn luyện sẽ đưa ra dự đoán về loại bệnh da mà bệnh nhân có thể mắc phải.

Quy trình này cung cấp một cách tiếp cận toàn diện để chẩn đoán ung thư da bằng cách sử dụng các kỹ thuật học sâu. Việc sử dụng mô hình phân loại học sâu kết hợp với các kỹ thuật tăng cường dữ liệu giúp cải thiện độ chính xác và tính ổn định của hệ thống chẩn đoán tự động.

2.2. Trích rút đặc trưng học sâu trong bài toán phân lớp ảnh

Trong các bài toán học máy hiện đại, đặc biệt là phân lớp ảnh, việc trích rút các đặc trưng học sâu từ mạng nơ-ron tích chập (CNN) là một trong những bước quan trọng giúp mô hình đạt được hiệu suất cao. Đặc trưng học sâu là những biểu diễn cấp cao của dữ liệu hình ảnh, được học tự động từ các mạng học sâu như CNN mà không cần phải can thiệp thủ công để thiết kế các đặc trưng. Điều này đã giúp các phương pháp học máy, đặc biệt là các mô hình phân loại ảnh, phát triển mạnh mẽ trong những năm gần đây [29].

2.2.1. Mạng Nơ-ron Tích chập (CNN) trong trích rút đặc trưng

Mạng nơ-ron tích chập (CNN) được sử dụng rộng rãi để trích rút các đặc trưng từ ảnh do khả năng học các cấu trúc hình ảnh ở nhiều cấp độ. CNN bao gồm nhiều lớp tích chập và lớp phi tuyến, trong đó mỗi lớp tích chập học cách trích rút các đặc trưng khác nhau từ dữ liệu đầu vào. Lớp đầu tiên có nhiệm vụ học các đặc trưng cơ bản như cạnh và góc, trong khi các lớp sâu hơn sẽ học các đặc trưng phức tạp hơn, như cấu trúc hình dáng của đối tượng [30].

Giả sử rằng ảnh đầu vào \mathbf{X} là một ma trận có kích thước $n \times n \times c$ (trong đó n là kích thước chiều của ảnh và c là số lượng kênh màu), đầu ra của một lớp tích chập có thể được biểu diễn bằng công thức:

$$\mathbf{Z} = \mathbf{W} * \mathbf{X} + b \quad (2.1)$$

Trong đó, \mathbf{W} là ma trận bộ lọc (filter), $*$ là phép tích chập, và b là hệ số bias. Kết quả của lớp này là một bản đồ đặc trưng (feature map).

Sau khi qua các lớp tích chập và lớp phi tuyến (ví dụ như ReLU), ảnh sẽ được giảm độ phân giải thông qua các lớp gộp (pooling). Lớp gộp có tác dụng giảm số lượng tham số và tính toán, đồng thời làm giảm độ phân giải của bản đồ đặc trưng. Công thức cho một phép gộp cực đại (max pooling) có thể được mô tả như sau:

$$Z_{\text{pool}}(i, j) = \max(Z(i : i + 2, j : j + 2)) \quad (2.2)$$

trong đó (i, j) đại diện cho vị trí trên bản đồ đặc trưng.

2.2.2. Sử dụng Đặc trưng Học Sâu cho Phân Lớp Ảnh

Sau khi đặc trưng đã được trích rút từ các lớp sâu của CNN, đặc trưng này có thể được sử dụng cho các bài toán phân lớp ảnh. Các mô hình học máy truyền thống như SVM (Support Vector Machine), Random Forest, hoặc các mạng nơ-ron sâu (DNN) có thể được áp dụng lên các đặc trưng này để thực hiện phân lớp. Đặc trưng học sâu từ CNN thường có dạng vector và chứa thông tin phong phú về hình ảnh đầu vào.

Một cách tiếp cận phổ biến là sử dụng mạng CNN với kiến trúc nhiều lớp tích chập và một hoặc nhiều lớp kết nối đầy đủ ở cuối để thực hiện phân lớp trực tiếp. Đặc biệt, lớp kết nối đầy đủ cuối cùng của CNN sử dụng hàm Softmax để tính toán xác suất cho mỗi lớp ảnh:

$$\text{Softmax}(z_j) = \frac{\exp(z_j)}{\sum_{k=1}^C \exp(z_k)} \quad (2.3)$$

Trong đó, z_j là đầu ra của lớp kết nối đầy đủ cho lớp j và C là tổng số lớp (ví dụ: các loại bệnh da hoặc loại ung thư da trong phân loại bệnh lý). Softmax sẽ ánh xạ giá trị z_j thành xác suất của lớp đó, giúp mạng CNN có thể thực hiện dự đoán.

Một phương pháp khác để sử dụng đặc trưng học sâu từ CNN là trích rút các đặc trưng từ các lớp tích chập trung gian, thay vì sử dụng toàn bộ mạng để phân lớp trực tiếp. Đặc trưng từ các lớp tích chập chứa các thông tin không gian và hình học quan trọng của ảnh, như biên và cấu trúc của đối tượng. Sau khi trích rút đặc trưng, chúng có thể được sử dụng như đầu vào cho các mô hình phân lớp truyền thống, như SVM hoặc Random Forest, để thực hiện nhiệm vụ phân lớp.

Công thức tính toán đặc trưng từ một lớp tích chập được biểu diễn như sau:

$$h^l_{i,j} = \sum_{m=1}^M \sum_{n=1}^N w^l_{m,n} x^{l-1}_{i+m,j+n} + b^l \quad (2.4)$$

Trong đó:

- $h^l_{i,j}$ là đầu ra tại vị trí (i, j) của lớp tích chập thứ l .
- $w^l_{m,n}$ là trọng số của bộ lọc kích thước $M \times N$ tại lớp l .
- $x^{l-1}_{i+m,j+n}$ là đầu vào tại vị trí $(i+m, j+n)$ từ lớp trước đó $(l-1)$.
- b^l là bias của lớp tích chập thứ l .

Các đặc trưng này, sau khi được trích rút qua các lớp tích chập và các phép pooling, có thể được chuyển thành một vector đặc trưng phẳng, sẵn sàng cho việc phân lớp hoặc học sâu hơn. Phương pháp này giúp tận dụng các thông tin phong phú về cấu trúc không gian trong ảnh và có thể cải thiện độ chính xác của mô hình [30].

2.2.3. Phân lớp ảnh sử dụng đặc trưng học sâu

Sau khi trích rút được các đặc trưng học sâu, ta có thể sử dụng chúng cho bài toán phân lớp ảnh. Một trong những ứng dụng phổ biến của CNN là trong chẩn đoán hình ảnh y tế. Ví dụ, trong phân loại ảnh bệnh lý, các đặc trưng từ các vùng ảnh chứa khối u sẽ được mạng CNN học và sử dụng để phân loại các loại bệnh như Basal Cell Carcinoma (BCC), Malignant Melanoma (MM), Squamous Cell Carcinoma (SCC), và không mắc bệnh ung thư da (no skin cancer).

Việc sử dụng đặc trưng học sâu giúp mô hình có khả năng học được các mối quan hệ phức tạp và không dễ dàng bị ảnh hưởng bởi các yếu tố nhiễu hoặc các đặc trưng không quan trọng trong ảnh. Điều này giúp cải thiện độ chính xác trong việc chẩn đoán và phân loại ảnh [31].

2.2.4. Ưu điểm của đặc trưng học sâu trong phân lớp ảnh

Việc sử dụng đặc trưng học sâu trong phân lớp ảnh có nhiều ưu điểm rõ ràng:

- **Khả năng học đặc trưng tự động:** Các mạng học sâu như CNN có thể tự động học các đặc trưng của ảnh từ dữ liệu mà không cần thiết kế thủ công. Điều này giúp giảm thiểu sự can thiệp của con người và tăng cường khả năng khái quát hóa.
- **Biểu diễn đa cấp độ:** Các lớp khác nhau của CNN học các đặc trưng ở nhiều cấp độ khác nhau, từ các đặc trưng cục bộ như cạnh và góc đến các đặc trưng cấp cao hơn như hình dạng và cấu trúc đối tượng.
- **Khả năng mở rộng và tổng quát hóa:** Các đặc trưng học sâu từ CNN có thể dễ dàng mở rộng và áp dụng cho các bài toán phân lớp khác nhau với các tập dữ liệu lớn [83].
- **Độ chính xác cao:** Đặc trưng học sâu thường mang lại độ chính xác cao hơn so với các phương pháp trích rút đặc trưng truyền thống, đặc biệt trong các bài toán phức tạp như phân loại ảnh y tế.

Đặc trưng học sâu từ mạng CNN đã mở ra nhiều cơ hội mới cho các bài toán phân lớp ảnh. Với khả năng học tự động và biểu diễn đa cấp độ, các đặc trưng này không chỉ mang lại hiệu quả cao trong các bài toán phân loại mà còn giúp các mô hình học máy đạt được độ chính xác vượt trội trong nhiều lĩnh vực, bao gồm cả phân loại hình ảnh y tế và chẩn đoán bệnh.

2.3. Các kiến trúc mạng nơ-ron tích chập truyền thống

Nhiều kiến trúc CNN đã được phát triển như GoogLeNet, DarkNet, AlexNet, ResNet, NasNet cùng với các mô hình hiện đại hơn như ViT_B16, VGG, EfficientNet và MobileNet. Một số trong số này có sẵn dưới dạng các mô hình đã được huấn luyện trước, ban đầu được huấn luyện trên các tập dữ liệu ảnh tự nhiên lớn. Trong phần này, chương giới thiệu tổng quan về một số kiến trúc CNN được sử dụng rộng rãi nhất và cách chúng trích xuất đặc trưng từ hình ảnh tổn thương da (Bảng 2.1).

2.3.1. GoogleNet

GoogleNet, còn được gọi là Inception-V1, được triển khai bởi Szegedy và cộng sự (2015) với mục tiêu giảm độ phức tạp tính toán so với các mạng CNN truyền thống. GoogleNet đã đạt được tỷ lệ lỗi top-5 là 6,67% và giành chiến thắng trong cuộc thi ILSVRC năm 2014 [32]. Mô hình sử dụng khối inception, một biến thể của các lớp tích chập cổ điển với nhiều quy mô bộ lọc khác nhau (1×1 , 3×3 và 5×5), giúp giảm thiểu vấn đề căn chỉnh các mạng và tối ưu hóa hiệu suất.

2.3.2. InceptionV3

InceptionV3, phát triển từ Inception-V1, là phiên bản cải tiến với 24 triệu tham số [33]. Mô hình này sử dụng các kỹ thuật tối ưu hóa tính toán như phân tích lớp tích chập lớn thành các lớp tích chập nhỏ hơn (từ 5×5 và 7×7 thành 3×3) và thay thế tích chập đối xứng bằng tích chập không đối xứng để giảm chi phí tính toán.

2.3.3. ViT_B16

Mô hình **Vision Transformer (ViT_B16)** áp dụng các kỹ thuật Transformer cho thị giác máy tính, thay thế các lớp tích chập bằng việc chia hình ảnh thành các patch nhỏ và xử lý chúng thông qua mô hình Transformer. ViT_B16 đã chứng minh hiệu quả cao trong phân loại hình ảnh với kích thước patch 16×16 pixel [34].

2.3.4. VGG16 và VGG19

VGG16 và **VGG19**, được phát triển bởi Simonyan và Zisserman (2015), nổi bật với cấu trúc đơn giản sử dụng các lớp tích chập 3×3 và lớp gộp cực đại 2×2 . VGG16 có 16 lớp, còn VGG19 có 19 lớp. Mô hình VGG đạt hiệu suất cao trong các bài toán phân loại ảnh với số lượng tham số lớn.

2.3.5. ResNet152

ResNet152 là phiên bản mở rộng của ResNet với 152 lớp, sử dụng kết nối dư (residual connections) để giảm thiểu hiện tượng biến mất độ dốc. ResNet152 được

đánh giá là một trong những mô hình hiệu quả nhất trong các bài toán phân loại hình ảnh [35].

2.3.6. EfficientNetB0 và EfficientNetB7

EfficientNet là một loạt các mô hình CNN được phát triển để tối ưu hóa cân bằng giữa hiệu suất và tốc độ tính toán. EfficientNet sử dụng phương pháp "compound scaling" để mở rộng mô hình một cách hệ thống. **EfficientNetB0** là phiên bản nhỏ nhất, trong khi **EfficientNetB7** là phiên bản lớn nhất với nhiều tham số hơn [36].

2.3.7. MobileNetV2

MobileNetV2 được phát triển cho các ứng dụng di động và nhúng. Mô hình này sử dụng các lớp tích chập tách biệt (depthwise separable convolutions) và cấu trúc "inverted residual" để giảm thiểu số lượng tham số mà vẫn duy trì hiệu suất cao trong các bài toán phân loại hình ảnh.

2.3.8. Xception

Xception là một biến thể của Inception sử dụng các lớp tích chập tách biệt (depthwise separable convolutions) thay vì tích chập cổ điển. Mô hình này giúp giảm thiểu số lượng tham số và tăng hiệu suất tính toán trong các bài toán thị giác máy tính [37].

2.3.9. DenseNet121

DenseNet121 là một phiên bản của DenseNet, một kiến trúc CNN đặc biệt trong đó mỗi lớp kết nối trực tiếp với tất cả các lớp trước đó. DenseNet giảm thiểu số lượng tham số và cải thiện quá trình truyền thông tin giữa các lớp [38].

2.4. Học chuyển giao (Transfer learning)

Thật không may, chúng ta hiếm khi có một tập dữ liệu hình ảnh lớn được gán nhãn để huấn luyện mạng nơ-ron tích chập (CNN) từ các tham số được khởi tạo ngẫu

Bảng 2.1: Tổng quan về các mô hình CNN được sử dụng

Kiến trúc CNN	Tổng số lớp	Tham số	Phát hành
ViT_B16	-	86 triệu	2020
VGG16	16	138 triệu	2015
VGG19	19	144 triệu	2015
ResNet152	152	60 triệu	2016
InceptionV3	42	24 triệu	2016
EfficientNetB0	-	5.3 triệu	2019
EfficientNetB7	-	66 triệu	2019
MobileNetV2	-	3.4 triệu	2018
Xception	71	23 triệu	2017
DenseNet121	121	8 triệu	2017

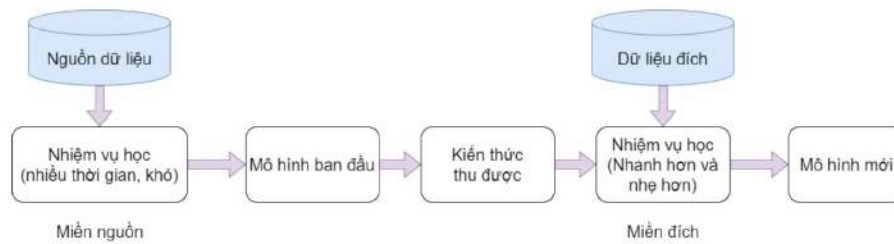
nhiên. Hơn nữa, thời gian xử lý cần thiết để hoàn thành tác vụ học này có thể là quá tốn kém. Học chuyển giao (Transfer learning) được coi là phương pháp tốt nhất để giải quyết những vấn đề này [39].

Học chuyển giao là một khái niệm trong học máy được sử dụng trong các tác vụ phân loại hình ảnh. Nó bao gồm việc sử dụng lại kiến thức trích xuất từ một mô hình CNN đã được huấn luyện với kết quả tốt trong miền nguồn. Các trọng số của mô hình đã được huấn luyện này được thu thập từ các tập dữ liệu lớn được gán nhãn như ImageNet. Do đó, mô hình đã được huấn luyện trên hàng triệu hình ảnh có nhãn và nhiều lớp với sức mạnh tính toán cao trong nhiều ngày. Sau khi mô hình đã được huấn luyện, có thể áp dụng các trọng số đã thu được cho một tập dữ liệu hình ảnh cụ thể khác trong miền mục tiêu mới [40].

Học chuyển giao có thể được sử dụng như một mô hình trích xuất các đặc trưng tích chập, nơi chúng ta loại bỏ các lớp kết nối đầy đủ (fully connected layers) cuối cùng và coi phần còn lại của CNN như một “máy trích xuất đặc trưng” cho tập dữ liệu nhỏ hơn. Do đó, học chuyển giao sẽ được sử dụng để phân loại hình ảnh tổn thương da vì chúng là hình ảnh tự nhiên và có thể sử dụng quá trình tiền huấn luyện trên ImageNet và chuyển các tham số đã học [41].

Học chuyển giao cũng có thể được sử dụng để tinh chỉnh các siêu tham số (fine-tune hyperparameters) bằng cách đóng băng hoặc mở các lớp. Cách thực hiện là huấn luyện một kiến trúc với các trọng số khởi tạo ngẫu nhiên bằng cách thêm một bộ phân loại tùy chỉnh theo tập dữ liệu mới, hoặc huấn luyện lại các trọng số của một số lớp

trong khi giữ nguyên các lớp khác. Hình 2.2 mô tả mô hình học chuyển giao.



Hình 2.2: Mô hình học chuyển giao.

2.5. Trích xuất đặc trưng (Features extraction)

Đặc trưng được mô tả như là "thông tin có liên quan để giải quyết các tác vụ tính toán khác nhau liên quan đến một ứng dụng cụ thể". Các đặc trưng được chia thành hai loại chính: đặc trưng cục bộ và đặc trưng toàn cục. Chúng ta thường dựa trên một tập hợp các tiêu chí số mô tả đối tượng hoặc hiện tượng được quan sát để phát triển quy tắc phân loại (có giám sát hoặc không giám sát).

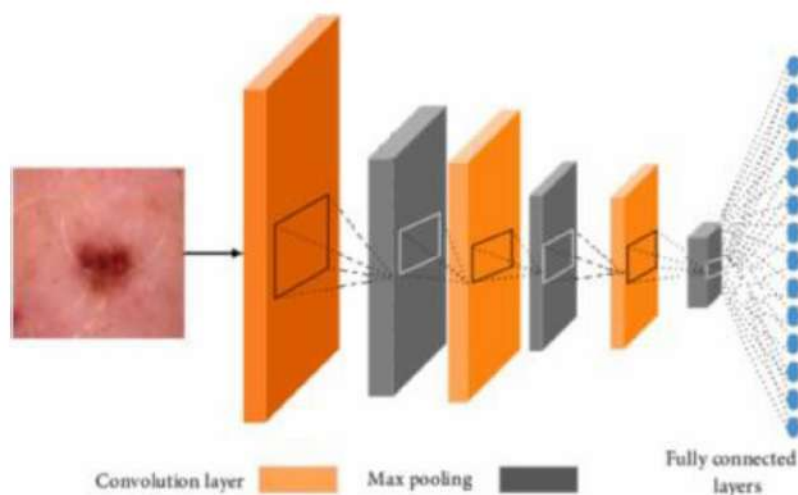
Nhiều kỹ thuật đã được tạo ra để trích xuất đặc trưng, và các nguyên tắc hoạt động của chúng khá khác biệt với nhau. Tuy nhiên, đến nay vẫn chưa tồn tại một bộ trích xuất đặc trưng hoàn hảo. Một số tài liệu nghiên cứu nổi bật trong lĩnh vực này có thể được tìm thấy trong [41], [42], và [43]. Việc tìm kiếm đặc trưng phù hợp cho hình ảnh tổn thương da thường là một thách thức lớn mà nhiều nghiên cứu gặp phải [44].

Nhìn chung, chúng ta có thể phân biệt các đặc trưng thủ công (handcrafted features) và đặc trưng dựa trên học sâu (deep learning features). Nhiều phương pháp trích xuất đặc trưng thủ công đã được đề xuất để phát hiện bệnh tổn thương da, chẳng hạn như phương pháp kiểm tra 7 điểm (7-point checklist), quy tắc ABCD (viết tắt của Asymmetry, Border, Color, và Dermoscopic structure), thuật toán CASH, và danh sách kiểm tra ba điểm. Các phương pháp phát hiện tổn thương da này dựa trên hình dạng, hình học, màu sắc, kết cấu và cấu trúc của tổn thương da.

Nhiều nghiên cứu đã chứng minh hiệu quả của việc sử dụng học sâu để trích xuất đặc trưng trong những năm gần đây. Có nhiều biến thể của các thuật toán học sâu để mô tả đặc trưng hình ảnh, bao gồm mạng nơ-ron tích chập (CNN) [45].

Để sử dụng mạng nơ-ron tích chập, có hai cách khác nhau. Thứ nhất, nó có thể được sử dụng như một mô hình phân loại, và thứ hai, để trích xuất đặc trưng bằng cách sử dụng học chuyển giao.

Ngoài ra, có những nghiên cứu tập trung vào việc kết hợp giữa đặc trưng thủ công và đặc trưng học sâu. Ví dụ, Filali et al. đã đề xuất kết hợp các đặc trưng thủ công (hình dạng, xương, màu sắc, và kết cấu) và các đặc trưng trích xuất từ bốn CNN đã được huấn luyện trước: AlexNet, VggNet, GoogLeNet, và ResNet. Họ thực hiện quá trình lựa chọn đặc trưng để giữ lại những đặc trưng tốt và phù hợp nhất bằng cách sử dụng thuật toán di truyền.



Hình 2.3: Mô hình trích xuất đặc trưng từ ảnh tổn thương da.

2.6. Phân loại có giám sát và các tập hợp bộ phân loại

Phân loại có giám sát là một phương pháp trong học máy, trong đó các thuật toán học từ dữ liệu đã được gán nhãn để đưa ra quyết định về dữ liệu mới. Phương pháp này dựa trên việc xây dựng một mô hình học từ tập huấn luyện, sau đó áp dụng mô hình đó để dự đoán nhãn cho các mẫu chưa biết [46].

Các bộ phân loại thường được sử dụng trong phân loại có giám sát bao gồm: cây quyết định (Decision Tree), Naive Bayes, máy vector hỗ trợ (SVM), k-nearest neighbor (KNN), các phương pháp kết hợp nhiều mô hình (ensemble methods), và thuật toán XGBoost.

2.6.1. Cây quyết định (Decision Tree)

Cây quyết định là một mô hình phân loại trong đó các quyết định được biểu diễn dưới dạng cấu trúc cây, với các nhánh thể hiện các điều kiện và các lá biểu thị các nhãn. Hàm phân tách của cây quyết định có thể dựa trên Entropy hoặc Gini Index [47]:

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \log_2(p_i) \quad (2.5)$$

trong đó p_i là xác suất của mỗi lớp i , c là số lượng lớp.

Chỉ số Gini được tính bằng:

$$\text{Gini}(S) = 1 - \sum_{i=1}^c p_i^2 \quad (2.6)$$

2.6.2. Naive Bayes

Naive Bayes [48] là một thuật toán phân loại dựa trên Định lý Bayes, trong đó tính xác suất của một lớp được tính từ các xác suất có điều kiện của các thuộc tính tương ứng. Đối với mỗi thuộc tính x_i :

$$P(C|x_1, x_2, \dots, x_n) = \frac{P(C) \prod_{i=1}^n P(x_i|C)}{P(x_1, x_2, \dots, x_n)} \quad (2.7)$$

Phiên bản Gaussian của Naive Bayes giả định rằng các giá trị thuộc tính theo phân phối Gaussian.

2.6.3. Máy vector hỗ trợ (SVM)

SVM [46] là một thuật toán học có giám sát dùng để phân loại và hồi quy. Nó tìm kiếm siêu phẳng tối ưu để phân chia các lớp sao cho khoảng cách giữa các điểm dữ liệu gần nhất (support vectors) đến siêu phẳng là lớn nhất:

$$\text{Maximize} \quad \frac{2}{\|\mathbf{w}\|} \quad (2.8)$$

với điều kiện:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \forall i \quad (2.9)$$

trong đó \mathbf{w} là vector trọng số và b là hệ số bias.

2.6.4. k-Nearest Neighbor (KNN)

KNN là một thuật toán phân loại đơn giản nhưng hiệu quả. Nó hoạt động bằng cách tìm các điểm lân cận gần nhất trong không gian đặc trưng để dự đoán nhãn cho mẫu cần phân loại. Khoảng cách giữa các điểm thường được đo bằng khoảng cách Euclid:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.10)$$

2.6.5. Phương pháp kết hợp (Ensemble Methods)

Các phương pháp kết hợp như Bagging, Boosting và Random Forest được sử dụng để cải thiện hiệu suất của các mô hình phân loại bằng cách kết hợp nhiều mô hình yếu thành một mô hình mạnh. Bagging xây dựng nhiều mô hình từ các tập con dữ liệu khác nhau, trong khi Boosting tập trung vào việc cải thiện hiệu suất của các mẫu khó phân loại.

Bagging: Bagging (Bootstrap Aggregating) xây dựng một tập hợp các bộ phân loại từ các mẫu ngẫu nhiên của cùng một tập dữ liệu huấn luyện.

Boosting: Boosting tìm cách tạo ra các bộ phân loại mới để sửa lỗi của các bộ phân loại trước đó. Phương trình tổng quát của Boosting là:

$$F(x) = \sum_{m=1}^M \alpha_m h_m(x) \quad (2.11)$$

trong đó $h_m(x)$ là các mô hình yếu và α_m là trọng số của chúng. Random Forest là một sự kết hợp của Bagging và kỹ thuật “Random SubSpace” áp dụng cho cây quyết định [49].

2.6.6. XGBoost

XGBoost (Extreme Gradient Boosting) là một thuật toán Boosting mạnh mẽ, dựa trên cây quyết định và được tối ưu hóa để tăng tốc độ huấn luyện và hiệu quả tính toán. Mô hình XGBoost được xây dựng dựa trên hàm mất mát sau:

$$L(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2.12)$$

Trong đó, hàm mất mát $l(y_i, \hat{y}_i)$ đo lường sự khác biệt giữa giá trị thực y_i và giá trị dự đoán \hat{y}_i , và $\Omega(f_k)$ là thuật phạt để điều chỉnh độ phức tạp của mô hình [50].

XGBoost được biết đến với khả năng xử lý hiệu quả cả dữ liệu mất mát (sparse data) và dữ liệu lớn, đồng thời có khả năng điều chỉnh tự động nhờ việc sử dụng các siêu tham số như learning rate, tree depth, và regularization.

2.7. Mô hình đề xuất

2.7.1. Phát biểu bài toán và mô hình

Giả sử tập dữ liệu đầu vào được ký hiệu là \mathbf{X} , trong đó mỗi điểm dữ liệu tương ứng với một hình ảnh được biểu diễn dưới dạng một vector đặc trưng. Tác giả định nghĩa nhãn tương ứng với từng điểm dữ liệu là y , trong đó y_i đại diện cho nhãn của hình ảnh thứ i . Các nhãn y_i nhận giá trị từ tập hợp $\{1, 2, 3, 4\}$, với ý nghĩa cụ thể như sau:

- $y_i = 1$: tương ứng với bệnh “BCC” (Ung thư biểu mô tế bào đáy),
- $y_i = 2$: tương ứng với bệnh “MM” (U ác tính),
- $y_i = 3$: tương ứng với bệnh “SCC” (Ung thư biểu mô tế bào vảy), và
- $y_i = 4$: tương ứng với không có bệnh ung thư da (“no skin cancer”).

Bài toán yêu cầu xây dựng một mô hình dự đoán dựa trên việc kết hợp nhiều mạng nơ-ron tích chập (CNN) để phân loại chính xác hình ảnh vào một trong bốn loại nhãn: “BCC” (Basal Cell Carcinoma), “MM” (Malignant Melanoma), “SCC” (Squamous Cell Carcinoma), và “no skin cancer”.

Ký hiệu $f(\mathbf{X})$ là mô hình tích hợp các mạng CNN để xử lý tập dữ liệu đầu vào \mathbf{X} và dự đoán nhãn tương ứng y cho mỗi hình ảnh. Vấn đề có thể được mô hình hóa thành bài toán tối ưu tìm kiếm các tham số mô hình \mathbf{W} và b nhằm tối thiểu hóa sai số phân loại trên tập dữ liệu huấn luyện. Bài toán tối ưu được biểu diễn như phương trình 2.13:

$$\min_{\mathbf{W}, b} \text{Loss}(f(\mathbf{X}; \mathbf{W}, b), y) \quad (2.13)$$

Trong đó, \mathbf{W} và b là các tham số của mô hình. $f(\mathbf{X}; \mathbf{W}, b)$ là các dự đoán của mô hình với tập tham số \mathbf{W} và b trên tập dữ liệu đầu vào \mathbf{X} . Hàm $\text{Loss}(\cdot)$ là hàm mất mát nhằm đo lường mức độ khác biệt giữa nhãn dự đoán và nhãn thực tế \mathbf{y} . Mục tiêu chính là tìm ra các giá trị tối ưu của \mathbf{W} và b sao cho hàm mất mát đạt giá trị nhỏ nhất, từ đó giúp mô hình có khả năng dự đoán chính xác trên tập dữ liệu kiểm tra.

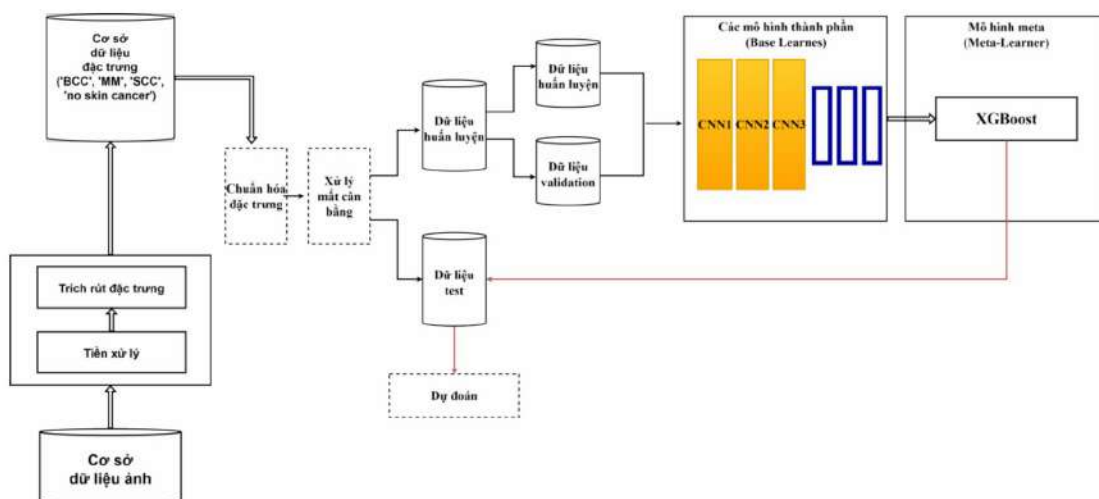
Hình 2.4 mô tả chi tiết kiến trúc của mô hình đề xuất. Tập dữ liệu đầu vào bao gồm các đặc trưng đã được trích xuất từ hình ảnh thuộc các lớp khác nhau: “BCC”, “MM”, “SCC”, và “no skin cancer”. Các đặc trưng này được chuẩn hóa để đảm bảo phân phối dữ liệu đồng nhất và phù hợp cho việc huấn luyện mô hình. Bước chuẩn hóa này đóng vai trò quan trọng trong việc cải thiện hiệu suất của mô hình học sâu.

Sau khi chuẩn hóa, dữ liệu được xử lý nhằm giải quyết vấn đề mất cân bằng trong các lớp. Điều này đảm bảo rằng mô hình không bị thiên lệch khi huấn luyện trên một số lớp có số lượng mẫu lớn hơn các lớp khác.

Dữ liệu sau đó được chia thành ba tập: dữ liệu huấn luyện, dữ liệu kiểm định (validation), và dữ liệu kiểm tra (test). Tập huấn luyện được sử dụng để huấn luyện ba mạng nơ-ron tích chập (CNN1, CNN2, CNN3), mỗi mạng đảm nhận một phần của bài toán phân loại đa lớp. Mỗi CNN sẽ học cách phân loại hình ảnh thuộc một trong bốn lớp thông qua quá trình tối ưu hóa các tham số của nó.

Sau khi các mô hình CNN được huấn luyện, dự đoán từ ba mô hình này được kết hợp lại để tạo ra một mô hình tổng hợp cuối cùng. Quá trình kết hợp dự đoán từ nhiều mô hình thường được gọi là Stacking (xếp chồng). Cuối cùng, mô hình tổng hợp này sẽ dự đoán kết quả cuối cùng cho các dữ liệu kiểm tra.

Mô hình Stacked Ensemble này được thiết kế nhằm mục tiêu tối ưu hóa độ chính xác của việc dự đoán, đồng thời giảm thiểu sai số bằng cách tận dụng ưu điểm của nhiều mạng CNN khác nhau.



Hình 2.4: Kiến trúc mô hình mạng học sâu kết hợp

2.7.2. Các thành phần của mô hình CNN

Mô hình Stacked Ensemble đề xuất bao gồm ba mạng nơ-ron tích chập (CNN) được sử dụng cho bài toán phân loại đa lớp. Tập dữ liệu đầu vào ký hiệu là X_{train} , có kích thước (m, n) , trong đó m là số lượng mẫu huấn luyện và n là số đặc trưng của mỗi mẫu. Tương ứng với các nhãn lớp được ký hiệu là y_{train} , có kích thước $(m,)$, đại diện cho các nhãn của m mẫu huấn luyện.

Trong quá trình huấn luyện, hàm mất mát 'categorical cross-entropy' được sử dụng nhằm tối thiểu hóa mất mát, và thuật toán tối ưu 'Adam' được áp dụng để cập nhật các tham số của mô hình. Quá trình đánh giá mô hình được thực hiện thông qua chỉ số chính xác (accuracy).

2.7.2.1. Mạng CNN 1

Mạng gồm hai lớp Dense được tích hợp với các cơ chế Dropout và Batch Normalization. Hàm kích hoạt ReLU được áp dụng ở các lớp ẩn, trong khi lớp đầu ra sử dụng hàm Softmax để tính toán xác suất cho từng lớp. Cấu trúc mạng như được thể hiện trong Hình 2.5 và Bảng 2.2. Biểu diễn hình thức của mạng như sau:

$$\mathbf{Z}^{[1]} = X_{\text{train}} \mathbf{W}^{[1]} + b^{[1]} \quad (2.14)$$

$$\mathbf{A}^{[1]} = \text{BatchNormalization } \mathbf{Z}^{[1]} \quad (2.15)$$

$$\mathbf{A}^{[2]} = \text{ReLU } \mathbf{A}^{[1]} \quad (2.16)$$

$$\mathbf{A}^{[3]} = \text{Dropout } \mathbf{A}^{[2]} \quad (2.17)$$

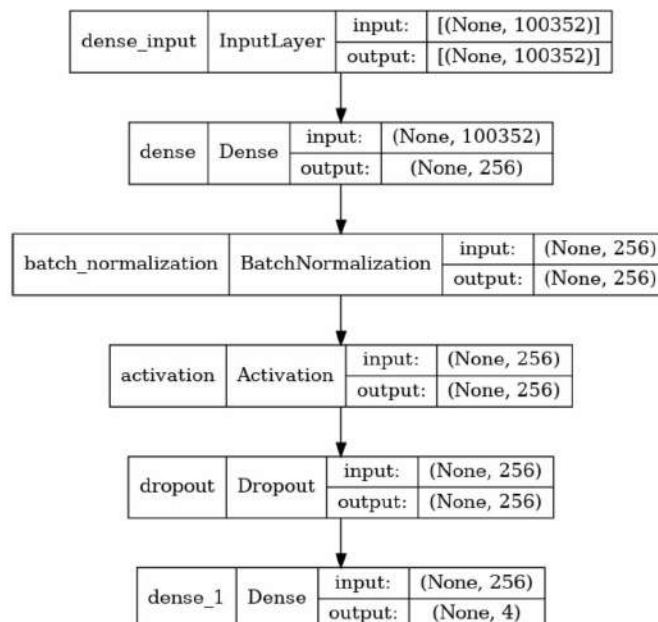
$$\mathbf{Z}^{[2]} = \mathbf{A}^{[3]}\mathbf{W}^{[2]} + b^{[2]} \quad (2.18)$$

$$\mathbf{A}^{[4]} = \text{Softmax } \mathbf{Z}^{[2]} \quad (2.19)$$

Trong đó, các tham số $\mathbf{W}^{[1]}$ và $b^{[1]}$ đại diện cho lớp Dense đầu tiên với 256 đơn vị, và $\mathbf{W}^{[2]}$ và $b^{[2]}$ là tham số của lớp Dense cuối cùng với số lượng đơn vị tương ứng với số lớp C .

Bảng 2.2: Kiến trúc mô hình CNN 1

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	25,690,368
batch_normalization (BatchNormalization)	(None, 256)	1,024
activation (ReLU)	(None, 256)	0
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 4)	1,028
Tổng số tham số:		25,692,420
Tham số có thể huấn luyện:		25,691,908
Tham số không huấn luyện:		512



Hình 2.5: Kiến trúc mô hình mạng CNN 1

2.7.2.2. Mạng CNN 2

Mạng này bao gồm ba lớp Dense giảm dần về kích thước, với lớp đầu tiên có 512 đơn vị. Mỗi lớp Dense được tích hợp với Batch Normalization và hàm kích hoạt ReLU, cùng với Dropout để giảm thiểu hiện tượng overfitting. Cấu trúc mạng được mô tả bằng các công thức sau:

$$\mathbf{Z}^{[1]} = X_{\text{train}} \mathbf{W}^{[1]} + \mathbf{b}^{[1]} \quad (2.20)$$

$$\mathbf{A}^{[1]} = \text{BatchNormalization } \mathbf{Z}^{[1]} \quad (2.21)$$

$$\mathbf{A}^{[2]} = \text{ReLU } \mathbf{A}^{[1]} \quad (2.22)$$

$$\mathbf{A}^{[3]} = \text{Dropout } \mathbf{A}^{[2]}, 0.5 \quad (2.23)$$

$$\mathbf{Z}^{[2]} = \mathbf{A}^{[3]} \mathbf{W}^{[2]} + \mathbf{b}^{[2]} \quad (2.24)$$

$$\mathbf{A}^{[4]} = \text{BatchNormalization } \mathbf{Z}^{[2]} \quad (2.25)$$

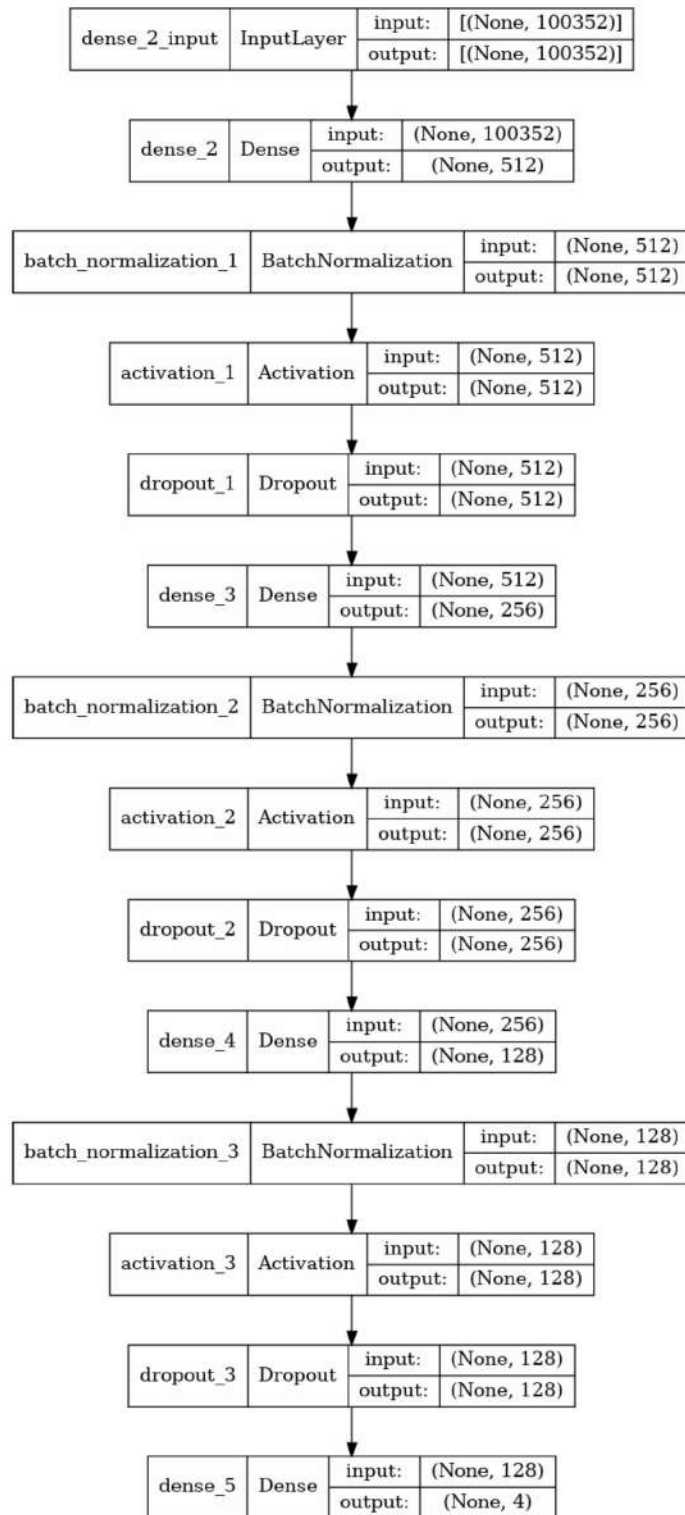
$$\mathbf{A}^{[5]} = \text{ReLU } \mathbf{A}^{[4]} \quad (2.26)$$

$$\mathbf{A}^{[6]} = \text{Dropout } \mathbf{A}^{[5]}, 0.5 \quad (2.27)$$

$$\mathbf{Z}^{[3]} = \mathbf{A}^{[6]} \mathbf{W}^{[3]} + \mathbf{b}^{[3]} \quad (2.28)$$

$$\mathbf{A}^{[7]} = \text{Softmax } \mathbf{Z}^{[3]} \quad (2.29)$$

Trong đó, các tham số $\mathbf{W}^{[i]}$ và $\mathbf{b}^{[i]}$ là trọng số và bias của từng lớp Dense tương ứng, và X_{train} là dữ liệu đầu vào.



Hình 2.6: Kiến trúc mô hình mạng CNN 2

Bảng 2.3: Kiến trúc mô hình CNN 2

Layer (type)	Output Shape	Param #
dense_2 (Dense)	(None, 512)	51,380,736
batch_normalization_1 (BatchNormalization)	(None, 512)	2,048
activation_1 (ReLU)	(None, 512)	0
dropout_1 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 256)	131,328
batch_normalization_2 (BatchNormalization)	(None, 256)	1,024
activation_2 (ReLU)	(None, 256)	0
dropout_2 (Dropout)	(None, 256)	0
dense_4 (Dense)	(None, 128)	32,896
batch_normalization_3 (BatchNormalization)	(None, 128)	512
activation_3 (ReLU)	(None, 128)	0
dropout_3 (Dropout)	(None, 128)	0
dense_5 (Dense)	(None, 4)	516
Tổng số tham số:		51,549,060
Tham số có thể huấn luyện:		51,547,268
Tham số không huấn luyện:		1,792

2.7.2.3. Mạng CNN 3

Mạng bao gồm ba lớp Dense với 128, 64 và 32 đơn vị, lần lượt được tích hợp với Batch Normalization, ReLU và Dropout. Lớp cuối cùng sử dụng hàm Softmax để tính toán xác suất cho các lớp đầu ra. Biểu diễn hình thức của mạng như sau:

$$\mathbf{Z}^{[1]} = X_{\text{train}} \cdot \mathbf{W}^{[1]} + \mathbf{b}^{[1]} \quad (2.30)$$

$$\mathbf{A}^{[1]} = \text{BatchNormalization}(\mathbf{Z}^{[1]}) \quad (2.31)$$

$$\mathbf{A}^{[2]} = \text{ReLU}(\mathbf{A}^{[1]}) \quad (2.32)$$

$$\mathbf{A}^{[3]} = \text{Dropout}(\mathbf{A}^{[2]}, 0.5) \quad (2.33)$$

$$\mathbf{Z}^{[2]} = \mathbf{A}^{[3]} \cdot \mathbf{W}^{[2]} + \mathbf{b}^{[2]} \quad (2.34)$$

$$\mathbf{A}^{[4]} = \text{BatchNormalization}(\mathbf{Z}^{[2]}) \quad (2.35)$$

$$\mathbf{A}^{[5]} = \text{ReLU}(\mathbf{A}^{[4]}) \quad (2.36)$$

$$\mathbf{A}^{[6]} = \text{Dropout}(\mathbf{A}^{[5]}, 0.5) \quad (2.37)$$

$$\mathbf{Z}^{[3]} = \mathbf{A}^{[6]} \cdot \mathbf{W}^{[3]} + \mathbf{b}^{[3]} \quad (2.38)$$

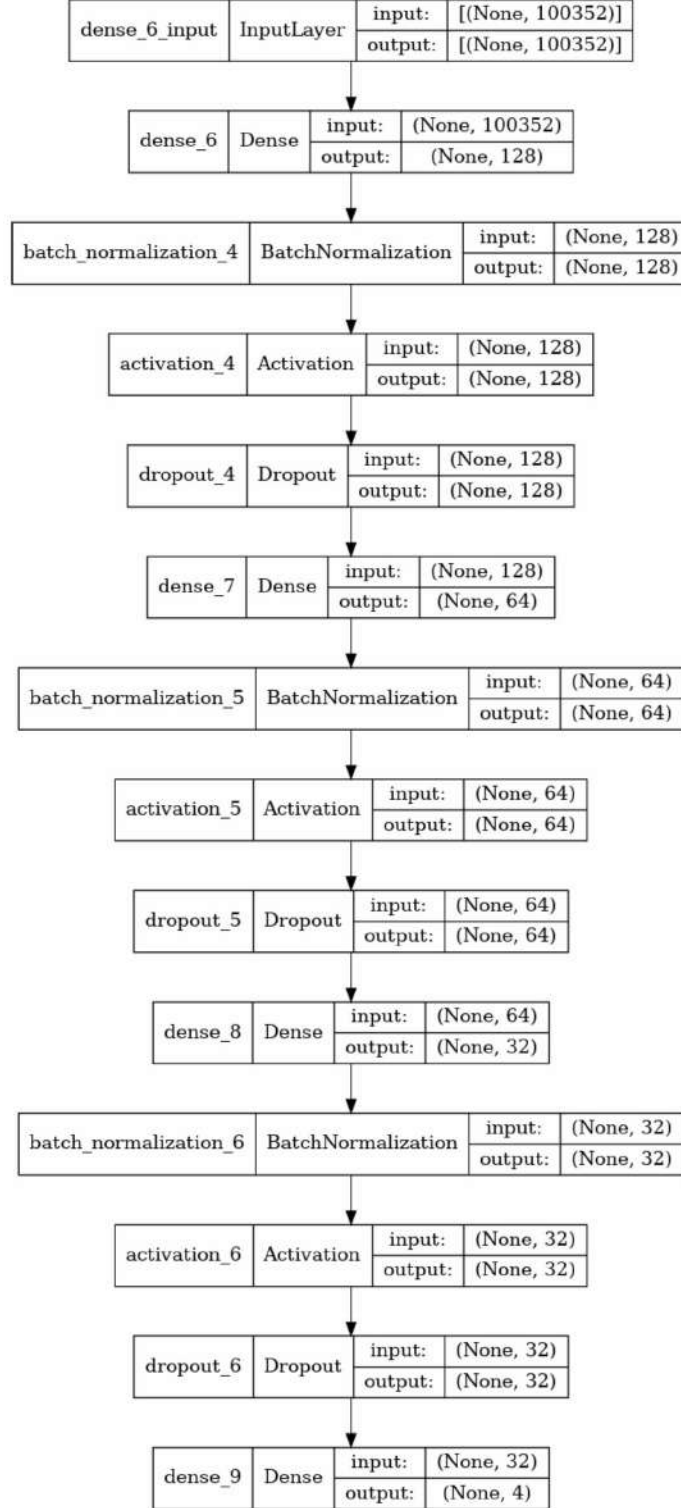
$$\mathbf{A}^{[7]} = \text{BatchNormalization}(\mathbf{Z}^{[3]}) \quad (2.39)$$

$$\mathbf{A}^{[8]} = \text{ReLU}(\mathbf{A}^{[7]}) \quad (2.40)$$

$$\mathbf{A}^{[9]} = \text{Dropout}(\mathbf{A}^{[8]}, 0.5) \quad (2.41)$$

$$\mathbf{Z}^{[4]} = \mathbf{A}^{[9]} \cdot \mathbf{W}^{[4]} + \mathbf{b}^{[4]} \quad (2.42)$$

$$\mathbf{A}^{[10]} = \text{Softmax}(\mathbf{Z}^{[4]}) \quad (2.43)$$



Hình 2.7: Kiến trúc mô hình mạng CNN 3

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 128)	12,845,184
batch_normalization_4 (BatchNormalization)	(None, 128)	512
activation_4 (ReLU)	(None, 128)	0
dropout_4 (Dropout)	(None, 128)	0
dense_7 (Dense)	(None, 64)	8,256
batch_normalization_5 (BatchNormalization)	(None, 64)	256
activation_5 (ReLU)	(None, 64)	0
dropout_5 (Dropout)	(None, 64)	0
dense_8 (Dense)	(None, 32)	2,080
batch_normalization_6 (BatchNormalization)	(None, 32)	128
activation_6 (ReLU)	(None, 32)	0
dropout_6 (Dropout)	(None, 32)	0
dense_9 (Dense)	(None, 4)	132
Tổng số tham số:		12,856,548
Tham số có thể huấn luyện:		12,856,100
Tham số không huấn luyện:		448

Bảng 2.4: Cấu trúc mô hình CNN 3

2.7.3. Thuật toán Stacked Ensemble

Ba mạng CNN được huấn luyện độc lập và sau đó kết hợp theo Thuật toán 2.1.

Thuật toán Stacked Ensemble bao gồm các bước sau:

1. **Chuẩn bị dữ liệu:** Chia dữ liệu thành ba tập con: huấn luyện, kiểm định, và kiểm tra.
2. **Huấn luyện các mô hình thành phần:** Mỗi mô hình CNN thành phần f_i được huấn luyện độc lập bằng cách tối ưu hàm mất mát của nó trên tập huấn luyện. Sau đó, các mô hình được đánh giá trên tập kiểm định.
3. **Tạo tập dữ liệu dự đoán:** Sử dụng các mô hình thành phần đã huấn luyện để dự đoán trên tập kiểm định, tạo thành một ma trận dự đoán **stackX**.
4. **Huấn luyện mô hình tổng hợp (stacking):** Mô hình tổng hợp, chẳng hạn như Logistic Regression, được huấn luyện trên ma trận dự đoán **stackX** và nhận kiểm định y_{val} .

5. **Đánh giá mô hình:** Mô hình tổng hợp cuối cùng được kiểm tra và đánh giá trên tập kiểm tra thông qua các chỉ số đo lường hiệu suất.

Thuật toán này giúp kết hợp các mô hình CNN thành phần để cải thiện hiệu suất dự đoán thông qua phương pháp Stacking, tận dụng các đặc trưng dự đoán từ nhiều mô hình để tối ưu hóa kết quả tổng hợp.

Trong kiến trúc Stacked Ensemble, việc sử dụng ba mạng nơ-ron tích chập (CNN) là nhằm mục tiêu cải thiện khả năng phân loại và tăng cường hiệu suất của mô hình tổng thể. Các lý do chính để sử dụng ba mạng CNN bao gồm:

- **Đa dạng hóa mô hình:** Mỗi mạng CNN có cấu trúc khác nhau, với số lượng lớp, số lượng nơ-ron, và hàm kích hoạt khác nhau. Điều này giúp tăng cường khả năng mô hình hóa các đặc trưng khác nhau của hình ảnh. Việc đa dạng hóa các kiến trúc CNN giúp mô hình học được nhiều thông tin phong phú hơn từ tập dữ liệu.
- **Tận dụng ưu điểm của từng mạng CNN:** Mỗi mạng CNN có thể khai thác một cách khác nhau các đặc trưng của dữ liệu. Ví dụ, một mạng CNN có thể hiệu quả hơn trong việc phát hiện các đặc trưng bề mặt của da, trong khi mạng khác có thể tập trung vào các đặc trưng hình dạng hoặc mô hình của các khối u ác tính. Việc sử dụng nhiều CNN cho phép kết hợp các ưu điểm của từng mạng, giúp tăng cường khả năng nhận dạng chính xác các loại bệnh.
- **Giảm thiểu nguy cơ overfitting:** Khi sử dụng một mô hình CNN duy nhất, có thể xảy ra hiện tượng overfitting nếu mô hình học quá kỹ các đặc trưng nhiễu hoặc đặc trưng không quan trọng của tập dữ liệu huấn luyện. Bằng cách sử dụng nhiều CNN và kết hợp dự đoán của chúng thông qua phương pháp Stacking, Stacked Ensemble giúp giảm thiểu nguy cơ overfitting, tăng cường khả năng khái quát hóa của mô hình trên tập dữ liệu chưa nhìn thấy.
- **Tăng cường độ chính xác qua phương pháp Stacking:** Sau khi huấn luyện các mạng CNN riêng lẻ, kết quả dự đoán của mỗi mô hình sẽ được kết hợp trong một mô hình tổng hợp (Stacking). Phương pháp Stacking không chỉ tận dụng được các dự đoán độc lập của từng mạng mà còn giúp tăng độ chính xác của kết quả cuối cùng thông qua việc kết hợp các quyết định của các mô hình thành phần. Điều này cho phép mô hình tổng hợp có thể đạt được độ chính xác cao hơn so với việc chỉ sử dụng một CNN duy nhất.

- **Tăng cường độ bền vững và ổn định của mô hình:** Khi sử dụng nhiều mô hình CNN, khả năng dự đoán sai của một mô hình cụ thể sẽ được bù đắp bởi các mô hình khác. Điều này tạo nên sự bền vững và ổn định trong quá trình dự đoán, đặc biệt khi có các mẫu dữ liệu khó phân loại hoặc không rõ ràng. Việc kết hợp nhiều mô hình CNN giúp tăng cường tính ổn định của hệ thống khi xử lý các trường hợp không điển hình.

Việc kết hợp nhiều mạng CNN trong kiến trúc Stacked Ensemble cho phép mô hình khai thác tối đa thông tin từ dữ liệu, cải thiện khả năng khái quát hóa, và nâng cao hiệu suất tổng thể. Do đó, việc sử dụng ba mạng CNN giúp Stacked Ensemble có khả năng phân loại hình ảnh chính xác và bền vững hơn trong các tác vụ phức tạp như phân loại ung thư da.

Thuật toán 2.1 Thuật toán Stacked Ensemble

Require:

Tập dữ liệu huấn luyện $\mathbf{X}_{\text{train}}$, nhãn $\mathbf{y}_{\text{train}}$, với $\mathbf{X}_i \in \mathbb{R}^n, y_i \in \mathbf{y}$

Ensure:

Mô hình kết hợp (stacked model) *stackModel*

- 1: **Bước 1: Chuẩn bị dữ liệu**
- 2: Chia tập dữ liệu $\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}$ thành các tập con: tập huấn luyện $\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}$; tập kiểm định (validation) $\mathbf{X}_{\text{val}}, \mathbf{y}_{\text{val}}$ và tập kiểm tra $\mathbf{X}_{\text{test}}, \mathbf{y}_{\text{test}}$.
- 3: **Bước 2: Huấn luyện các mô hình thành phần**
- 4: **for** $i = 1$ đến 3 **do**
- 5: Huấn luyện các mô hình CNN thành phần $f_i(\mathbf{X}; \mathbf{W}_i, b_i)$ (với \mathbf{W}_i, b_i là tham số của mô hình i) trên tập huấn luyện $\mathbf{X}_{\text{train}}$.
- 6: Tối ưu các tham số \mathbf{W}_i, b_i bằng cách giải bài toán tối ưu:

$$\min_{\mathbf{W}_i, b_i} \text{Loss}(f_i(\mathbf{X}; \mathbf{W}_i, b_i), y) \quad (2.44)$$
- 7: Đánh giá hiệu suất mô hình f_i trên tập kiểm định \mathbf{X}_{val} .
- 8: Lưu lại các mô hình đã huấn luyện f_i .
- 9: **end for**
- 10: **Bước 3: Tạo tập dữ liệu dự đoán từ các mô hình thành phần**
- 11: Khởi tạo ma trận dự đoán **stackX** trống.
- 12: **for** $i = 1$ đến 3 **do**
- 13: Đọc mô hình $f_i(\mathbf{X}; \mathbf{W}_i, b_i)$ đã lưu.
- 14: Tạo dự đoán $\hat{y}_i = f_i(\mathbf{X}_{\text{val}}; \mathbf{W}_i, b_i)$ từ mô hình f_i trên tập kiểm định \mathbf{X}_{val} .
- 15: **if** **stackX** là rỗng **then**
- 16: Khởi tạo **stackX** với giá trị của \hat{y}_i .
- 17: **else**
- 18: Nối các dự đoán \hat{y}_i vào ma trận **stackX** theo chiều dọc:

$$\mathbf{stackX} \leftarrow \text{dstack}(\mathbf{stackX}, \hat{y}_i) \quad (2.45)$$

- 19: **end if**
 20: **end for**
 21: **Bước 4: Huấn luyện mô hình tổng hợp (Stacking)**
 22: Khởi tạo mô hình tổng hợp *stackModel* (chẳng hạn như Logistic Regression hoặc XG-Boost).
 23: Huấn luyện *stackModel* với đầu vào là ma trận dự đoán **stackX** và nhãn y_{val} , thông qua việc giải bài toán tối ưu:

$$\min_{\mathbf{W}, b} \text{Loss}(\text{stackModel}(\text{stackX}; \mathbf{W}, b), y_{val}) \quad (2.46)$$

- 24: **Bước 5: Đánh giá mô hình**
 25: Sử dụng mô hình tổng hợp *stackModel* để dự đoán trên tập kiểm tra \mathbf{X}_{test} :

$$\hat{\mathbf{y}}_{test} = \text{stackModel}(\mathbf{X}_{test}) \quad (2.47)$$

- 26: Đánh giá hiệu suất của mô hình *stackModel* trên tập kiểm tra \mathbf{y}_{test} thông qua các chỉ số đo như độ chính xác, độ nhạy, độ đặc hiệu, F1-score, và AUC.

Kết luận chương 2

Chương 2 đã trình bày chi tiết về mô hình học sâu kết hợp trong chẩn đoán ung thư da. Quy trình chẩn đoán bao gồm các bước từ thu thập dữ liệu, tiền xử lý, chia tập dữ liệu, trích rút đặc trưng đến huấn luyện mô hình và xác định bệnh. Việc sử dụng mạng nơ-ron tích chập (CNN) đã cho thấy hiệu quả cao trong việc trích rút các đặc trưng từ hình ảnh tổn thương da, góp phần vào quá trình phân loại chính xác các loại ung thư da.

Trong phần này, chương cũng đã thảo luận về các kiến trúc CNN phổ biến như GoogleNet, VGG16, ResNet, EfficientNet, và các kỹ thuật học chuyển giao (Transfer learning) để cải thiện hiệu suất mô hình. Cuối cùng, chương đã trình bày về phương pháp kết hợp các mô hình (Stacked Ensemble) nhằm tăng cường độ chính xác và độ bền vững của hệ thống chẩn đoán.

Các kết quả đạt được từ việc xây dựng và thử nghiệm mô hình đã cho thấy rằng việc kết hợp nhiều mô hình CNN giúp cải thiện độ chính xác, giảm thiểu hiện tượng quá khớp (overfitting) và tối ưu hóa quá trình dự đoán. Mô hình Stacked Ensemble đã được sử dụng thành công trong việc phân loại các loại ung thư da và mang lại hiệu suất cao trên tập dữ liệu kiểm tra. Kết quả này mở ra tiềm năng ứng dụng mô hình vào các hệ thống chẩn đoán ung thư da tự động trong thực tế.

Chương tiếp theo sẽ phân tích về tập dữ liệu, kỹ thuật tiền xử lý, các chỉ số đánh giá, đồng thời thảo luận về các kết quả thực nghiệm chi tiết.

CHƯƠNG 3: THỰC NGHIỆM MÔ HÌNH KẾT HỢP CÁC MẠNG NƠ-RON TÍCH CHẬP CHẨN ĐOÁN BỆNH UNG THƯ DA

Trong chương này, luận văn sẽ giới thiệu các phương pháp thực nghiệm và phân tích kết quả của các mô hình mạng nơ-ron tích chập (CNN) trong việc chẩn đoán ung thư da. Bằng cách áp dụng các mô hình học sâu tiên tiến như DenseNet, MobileNet và Xception, luận văn không chỉ xem xét hiệu suất của từng mô hình Bộ học cơ sở mà còn đánh giá sự kết hợp của chúng thông qua mô hình Stacked Ensemble. Chương này cung cấp các đánh giá chi tiết về độ chính xác, độ nhạy, và các chỉ số khác, đồng thời so sánh hiệu năng giữa các mô hình khi huấn luyện với nhiều kích thước batch size và số lượng epoch khác nhau.

3.1. Các chỉ số đánh giá hiệu năng cho phân loại đa lớp

Luận văn sử dụng nhiều số liệu đánh giá tiêu chuẩn khác nhau để so sánh hiệu suất của các bộ phân loại học máy để phân loại tổn thương da như độ chính xác, độ nhạy, độ đặc hiệu, độ chính xác, độ chính xác và F-measure [51].

3.1.1. Ma trận nhầm lẫn (Confusion Matrix)

Ma trận nhầm lẫn là một công cụ mạnh để đánh giá hiệu năng của mô hình phân loại. Nó thể hiện tần suất các dự đoán đúng và sai cho từng lớp trong tập dữ liệu. Một ma trận nhầm lẫn cho bài toán phân loại đa lớp với 3 lớp có thể được biểu diễn như sau:

	Dự đoán Lớp 1	Dự đoán Lớp 2	Dự đoán Lớp 3
Thực tế Lớp 1	TP_1	FP_1	FP_2
Thực tế Lớp 2	FN_1	TP_2	FP_3
Thực tế Lớp 3	FN_2	FN_3	TP_3

Trong đó:

- TP (True Positives) là số lượng dự đoán đúng cho lớp tương ứng.
- FP (False Positives) là số lượng mẫu bị dự đoán sai thành lớp khác (dự đoán sai dương tính).
- FN (False Negatives) là số lượng mẫu bị bỏ qua (dự đoán sai âm tính).

Các phần tử trên đường chéo chính của ma trận nhầm lẫn (TP_1, TP_2, TP_3) là số lần mô hình dự đoán đúng các lớp tương ứng.

3.1.2. Precision và Recall

Precision (Độ chính xác dương tính) là tỷ lệ giữa số mẫu dự đoán đúng dương tính trên tổng số mẫu được dự đoán là dương tính, giúp trả lời câu hỏi: “Trong số những dự đoán dương tính, bao nhiêu là đúng?”.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.1)$$

Recall (Độ nhạy hay Tỷ lệ dự đoán đúng dương tính) đo lường khả năng mô hình phát hiện chính xác các mẫu dương tính, trả lời câu hỏi: “Trong số các mẫu thực tế dương tính, mô hình dự đoán đúng bao nhiêu?”.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.2)$$

Ví dụ: Nếu một mô hình dự đoán đúng 80 mẫu dương tính trong số 100 mẫu thực tế dương tính, thì Recall của mô hình sẽ là 0.80.

3.1.3. Độ chính xác (Accuracy)

Accuracy là một chỉ số tổng quát để đánh giá hiệu suất của mô hình, thể hiện tỷ lệ dự đoán đúng so với tổng số mẫu. Độ chính xác được tính bằng:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.3)$$

Tuy nhiên, độ chính xác không phải lúc nào cũng phù hợp, đặc biệt khi tập dữ liệu bị mất cân bằng (ví dụ khi một lớp có số lượng mẫu lớn hơn nhiều so với các lớp

khác).

3.1.4. F1-Score

F1-Score là trung bình điều hòa giữa Precision và Recall, giúp cân bằng giữa hai chỉ số này. Nó đặc biệt hữu ích khi chúng ta cần một thước đo tổng thể cho hiệu năng của mô hình trên cả hai khía cạnh.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

Giá trị của F1-Score dao động từ 0 đến 1, trong đó 1 là giá trị tốt nhất (tức là Precision và Recall đều cao). Chỉ số này rất hữu ích trong các bài toán mà một trong hai chỉ số Precision hoặc Recall có thể thấp, vì nó không chỉ tập trung vào một trong hai mà cân bằng cả hai.

3.1.5. Macro F1-Score

Macro F1-Score tính toán F1-Score cho từng lớp riêng lẻ và sau đó lấy trung bình các giá trị đó. Điều này đảm bảo rằng các lớp nhỏ hơn không bị bỏ qua và tất cả các lớp đều có trọng số như nhau.

Công thức của Macro F1-Score:

$$\text{Macro F1-Score} = \frac{1}{K} \sum_{k=1}^K F1_k \quad (3.5)$$

Với K là số lượng lớp, và $F1_k$ là F1-Score của lớp k .

Ưu điểm của Macro F1-Score là nó chú trọng vào hiệu suất của từng lớp và không bị ảnh hưởng bởi sự mất cân bằng lớp.

3.1.6. Micro F1-Score

Micro F1-Score tính tổng số True Positives, False Positives, và False Negatives trên toàn bộ các lớp, sau đó tính F1-Score từ các tổng này. Điều này tương đương với việc tính Accuracy trên toàn bộ tập dữ liệu nhưng sử dụng F1-Score để phản ánh sự cân bằng giữa Precision và Recall.

$$\text{Micro F1-Score} = \frac{2 \cdot \sum TP}{2 \cdot \sum TP + \sum FP + \sum FN} \quad (3.6)$$

Micro F1-Score đặc biệt hữu ích khi có sự mất cân bằng giữa các lớp vì nó coi tất cả các mẫu như nhau, không phân biệt lớp nào có nhiều mẫu hơn.

3.1.7. Độ nhạy (Sensitivity)

Độ nhạy (hay còn gọi là Tỷ lệ phát hiện dương tính, Sensitivity) là khả năng của mô hình trong việc phát hiện chính xác các mẫu thuộc lớp dương tính.

Độ nhạy tương đương với Recall trong bài toán phân loại nhị phân, thể hiện tỷ lệ giữa số mẫu dương tính thực tế được dự đoán đúng trên tổng số mẫu dương tính thực tế.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3.7)$$

Độ nhạy cao nghĩa là mô hình có khả năng phát hiện tốt các mẫu dương tính.

3.1.8. Độ đặc hiệu (Specificity)

Độ đặc hiệu (Specificity) là tỷ lệ dự đoán đúng của các mẫu âm tính trong tập dữ liệu. Nó phản ánh khả năng của mô hình trong việc dự đoán chính xác các mẫu thuộc lớp âm tính. Công thức tính độ đặc hiệu như sau:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3.8)$$

Độ đặc hiệu càng cao thì mô hình càng ít dự đoán nhầm các mẫu âm tính thành dương tính.

3.1.9. AUC - Diện tích dưới đường cong (Area Under the Curve)

AUC (Area Under the Curve) là một chỉ số quan trọng trong các bài toán phân loại nhị phân và đa lớp. AUC đo lường hiệu năng của mô hình bằng cách tính diện tích dưới đường cong ROC (Receiver Operating Characteristic). Đường cong ROC

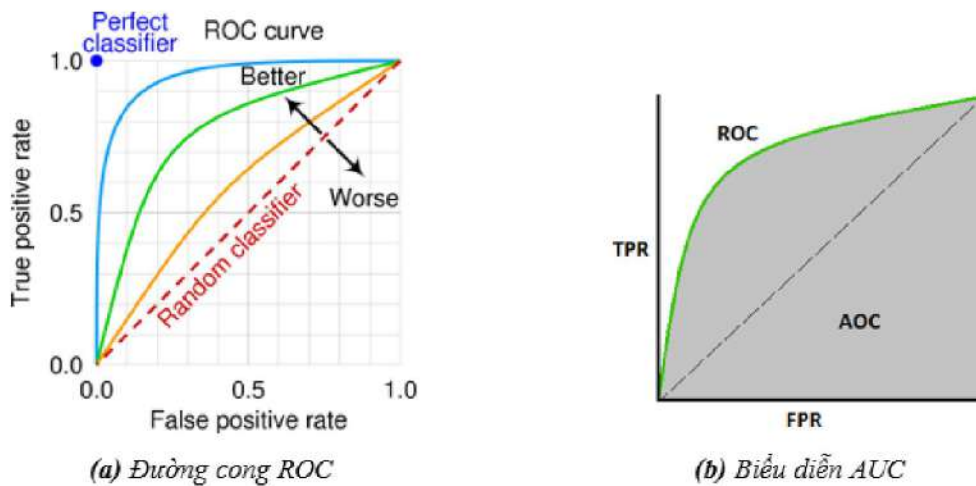
biểu diễn mối quan hệ giữa tỷ lệ phát hiện dương tính (True Positive Rate - TPR) và tỷ lệ dương tính giả (False Positive Rate - FPR). AUC càng cao, mô hình càng tốt trong việc phân biệt giữa các lớp.

$$AUC = \int_0^1 TPR(x) d(FPR(x)) \quad (3.9)$$

Trong đó:

- TPR (True Positive Rate) là Độ nhạy (Sensitivity), được tính bằng TP
- FPR (False Positive Rate) được tính bằng FP
- AUC có giá trị trong khoảng từ 0 đến 1
- AUC = 1: Mô hình hoàn hảo, phân biệt hoàn toàn chính xác giữa các lớp
- AUC = 0.5: Mô hình không tốt hơn việc dự đoán ngẫu nhiên.
- AUC < 0.5: Mô hình dự đoán tệ hơn dự đoán ngẫu nhiên.

Hình 3.1 là ví dụ cho một biểu đồ ROC curve chứa AUC.



Hình 3.1: Biểu đồ đường cong ROC và AUC

3.2. Cơ sở dữ liệu thực nghiệm

Trong phần này, luận văn trình bày về tập dữ liệu được sử dụng trong nghiên cứu, được các bác sĩ chuyên khoa da liễu tại Bệnh viện Da liễu Trung ương thu thập, gọi tắt là tập dữ liệu VNCancer.

3.2.1. Tập dữ liệu VNCancer

Tập dữ liệu bao gồm tổng cộng 4.485 hình ảnh, được chụp bằng các thiết bị di động thông minh như điện thoại di động. Một số hình ảnh từ tập dữ liệu này được minh họa trong Hình 3.2. Các hình ảnh đã được phân loại cẩn thận thành ba lớp chính, đại diện cho ba loại ung thư da phổ biến, bao gồm: Ung thư biểu mô tế bào đáy (BCC), U hắc tố (MM), và Ung thư biểu mô tế bào vảy (SCC). Bên cạnh đó, một lớp "Không có ung thư da" cũng được đưa vào để làm đối chứng. Sự phân bố số lượng hình ảnh giữa các lớp được tóm tắt trong Bảng 3.1. Tập dữ liệu được chia thành ba phần: tập huấn luyện (train), tập kiểm tra (test) và tập xác thực (validation), với tỷ lệ phân chia là 70% cho train, và mỗi phần test và validation chiếm 15%. Việc phân chia này giúp đảm bảo tính khách quan và độ chính xác của mô hình khi thực hiện các giai đoạn huấn luyện và đánh giá.

Bảng 3.1: Số lượng mẫu cho mỗi lớp trong tập dữ liệu VNCancer.

Lớp	Tổng số ảnh	Train	Test	Val	Phần trăm
Ung thư biểu mô tế bào đáy (BCC)	2765	1935	414	416	61.64%
U hắc tố (MM)	202	141	30	31	4.50%
Ung thư biểu mô tế bào vảy (SCC)	899	629	134	136	20.04%
Không ung thư da	619	433	92	94	13.81%
Tổng cộng	4485	3138	670	677	100%

Phương pháp phân chia dữ liệu với tỷ lệ 70% cho tập huấn luyện, 15% cho tập kiểm tra và 15% cho tập xác thực là một lựa chọn hợp lý để đảm bảo mô hình có đủ dữ liệu trong giai đoạn huấn luyện mà vẫn đảm bảo được việc đánh giá mô hình một



Hình 3.2: Một số ảnh chụp trong tập VNCancer.

cách khách quan trên tập kiểm tra và xác thực. Việc dành một phần dữ liệu cho tập kiểm tra và tập xác thực cũng giúp mô hình tránh hiện tượng “quá khớp” (overfitting) khi chỉ sử dụng dữ liệu huấn luyện.

Tuy nhiên, một hạn chế tiềm năng của tập dữ liệu này là sự chênh lệch về số lượng mẫu giữa các lớp. Cụ thể, lớp U hắc tố (MM) chỉ chiếm 4.50% trong khi lớp Ung thư biểu mô tế bào đáy (BCC) chiếm hơn 60% tổng số mẫu. Sự chênh lệch này có thể dẫn đến hiện tượng mất cân bằng lớp (class imbalance), ảnh hưởng đến độ chính xác của mô hình khi phân loại các lớp có ít mẫu. Do đó, các kỹ thuật như oversampling, undersampling hoặc áp dụng trọng số lớp trong quá trình huấn luyện có thể được cân nhắc để cải thiện hiệu suất mô hình trên các lớp có ít mẫu hơn.

3.2.2. Tiền xử lý

Quá trình tiền xử lý đóng vai trò rất quan trọng trong việc cải thiện chất lượng của các ảnh da liễu, từ đó nâng cao độ chính xác trong việc phát hiện và phân tích tổn thương da. Các yếu tố như tóc, ánh sáng không đồng đều, và các tạp nhiễu khác có thể gây ảnh hưởng lớn đến kết quả chẩn đoán. Do đó, trước khi thực hiện các bước phân tích và trích rút đặc trưng, các kỹ thuật tiền xử lý được áp dụng để giảm thiểu những tác động này (Hình 3.3).



Hình 3.3: Tiền xử lý ảnh da liễu.

Trong nghiên cứu này, chúng tôi đã sử dụng các kỹ thuật tiền xử lý phổ biến bao gồm cân bằng độ sáng, cân bằng histogram, và loại bỏ tóc. Mỗi kỹ thuật có mục tiêu riêng và đóng góp vào việc cải thiện độ chính xác của mô hình.

- **Cân bằng độ sáng (Brightness Balance):** Kỹ thuật này giúp điều chỉnh độ sáng của ảnh để làm cho các đặc trưng của tổn thương da trở nên rõ ràng hơn, đặc biệt khi ảnh gốc có sự chênh lệch lớn về ánh sáng. Quá trình này được thực hiện bằng cách chuyển đổi ảnh từ không gian màu RGB sang không gian HSV, sau đó áp dụng cân bằng histogram trên kênh độ sáng (V - Value). Phương pháp cân bằng độ sáng đã được chứng minh là hiệu quả trong việc tối ưu hóa độ sáng cho các ảnh có điều kiện ánh sáng không đồng đều [53]. Công thức cân bằng độ sáng được thực hiện như sau:

$$I' = \text{equalizeHist}(V) \quad (3.10)$$

trong đó I' là ảnh sau khi cân bằng kênh độ sáng, và V là kênh độ sáng trong không gian màu HSV. Phương pháp này được trích dẫn trong sách "Digital Image Processing" của Gonzalez và Woods [52].

- **Cân bằng histogram (Histogram Equalization):** Đây là một kỹ thuật phổ biến để tăng cường độ tương phản của ảnh. Cân bằng histogram giúp phân bố lại giá trị cường độ của các pixel trong ảnh sao cho chúng trải rộng trên toàn bộ dải giá trị. Trong ảnh màu, kỹ thuật này được áp dụng trên kênh Y của không gian màu YCrCb. Kỹ thuật này rất hữu ích trong việc làm nổi bật các đặc trưng khó thấy trong các vùng tối của ảnh.

$$I' = \text{equalizeHist}(Y) \quad (3.11)$$

trong đó I' là ảnh sau khi cân bằng histogram, và Y là kênh độ sáng trong không gian màu YCrCb.

- **Loại bỏ tóc (Hair Removal):** Một trong những thách thức lớn trong ảnh da liễu là sự xuất hiện của tóc, gây cản trở việc quan sát tổn thương da. Kỹ thuật loại bỏ tóc sử dụng lọc hình thái học (*morphological filtering*) để phát hiện vùng tóc trong ảnh, sau đó sử dụng thuật toán *inpainting* để thay thế vùng tóc bằng các pixel lân cận [116]. Đây là một phương pháp hiệu quả để loại bỏ nhiều tóc mà không làm mất đi các thông tin quan trọng của ảnh.

$$\text{blackhat} = I_{\text{gray}} - \text{closing}(I_{\text{gray}}, \text{kernel}) \quad (3.12)$$

sau đó, vùng tóc được loại bỏ bằng cách áp dụng thuật toán *inpainting* lên mặt nạ tóc đã phát hiện:

$$I' = \text{inpaint}(I, \text{mask}, 1, \text{INPAINT_TELEA}) \quad (3.13)$$

trong đó I_{gray} là ảnh xám được chuyển đổi từ ảnh gốc, và *kernel* là cấu trúc lọc. Thuật toán *DullRazor* [116] đã được áp dụng rộng rãi trong lĩnh vực này.

Các kỹ thuật tiền xử lý này giúp loại bỏ hiệu quả các yếu tố gây nhiễu, đồng thời cải thiện đáng kể chất lượng ảnh đầu vào. Điều này góp phần nâng cao độ chính xác của các mô hình trong quá trình trích rút đặc trưng và phân loại tổn thương da, đảm bảo rằng các yếu tố gây nhiễu không ảnh hưởng tiêu cực đến quá trình phân tích.

3.2.3. Cân bằng dữ liệu

Trong các bài toán phân loại với dữ liệu mất cân bằng, việc mô hình hóa trở nên khó khăn do các lớp chiếm số lượng lớn hơn có thể ảnh hưởng đến kết quả dự đoán của mô hình, dẫn đến hiện tượng bias đối với các lớp ít dữ liệu. Để giải quyết vấn đề này, các kỹ thuật cân bằng dữ liệu như *oversampling* và *undersampling* đã được áp dụng trong thực nghiệm này.

Oversampling là kỹ thuật nhân bản các mẫu từ lớp thiểu số (minority class) để làm tăng tỷ lệ của lớp này trong tập huấn luyện. Bằng cách tăng cường các mẫu của lớp thiểu số, mô hình có thể học được các đặc trưng của lớp này tốt hơn. Trong thực nghiệm của luận văn, kỹ thuật *RandomOverSampler* đã được sử dụng để thực hiện oversampling một cách ngẫu nhiên các mẫu của lớp thiểu số. Cụ thể, thuật toán này tạo ra các bản sao của dữ liệu từ các lớp thiểu số để cân bằng tỷ lệ giữa các lớp.

Undersampling là kỹ thuật ngược lại, trong đó số lượng mẫu từ lớp chiếm đa số (majority class) được giảm bớt. Điều này giúp cân bằng số lượng mẫu giữa các lớp bằng cách loại bỏ bớt các mẫu từ lớp chiếm đa số, giảm thiểu hiện tượng bias. Tuy nhiên, nhược điểm của phương pháp này là có thể làm mất đi một số thông tin quan trọng từ lớp chiếm đa số. Trong thực nghiệm này, luận văn sử dụng *RandomUnderSampler* để thực hiện undersampling cho các lớp chiếm đa số.

$$X_{\text{resampled}}, y_{\text{resampled}} = \text{sampler.fit_resample}(X, y) \quad (3.14)$$

Ngoài ra, luận văn cũng sử dụng kỹ thuật tính trọng số cho các lớp (*class weighting*) để mô hình hóa ảnh hưởng của các lớp thiểu số mà không cần điều chỉnh tập dữ liệu. Phương pháp này tính toán trọng số dựa trên số lượng mẫu của mỗi lớp, và mô hình học sẽ được điều chỉnh để quan tâm nhiều hơn đến các lớp thiểu số. Công thức tính trọng số cho các lớp được biểu diễn như sau:

$$w_c = \frac{N}{n_c} \quad (3.15)$$

trong đó, w_c là trọng số của lớp c , N là tổng số mẫu, và n_c là số lượng mẫu của lớp c . Phương pháp này đã được đề xuất và sử dụng rộng rãi trong nhiều nghiên cứu về dữ liệu mất cân bằng.

Nhờ sự kết hợp của các kỹ thuật trên, mô hình học có thể tránh được hiện tượng bias, đồng thời cải thiện độ chính xác khi phân loại các lớp thiểu số trong tập dữ liệu đa liệu.

3.3. Các kết quả thực nghiệm và luận giải

Nội dung này sẽ trình bày các kết quả được thực nghiệm trên tập dữ liệu VN-Cancer với kỹ thuật đề xuất và so sánh với các mô hình học sâu khác.

3.3.1. Cấu hình máy tính và các thư viện sử dụng

Các thực nghiệm trong nghiên cứu này được thực hiện trên một máy tính có cấu hình như sau: bộ vi xử lý Intel Core i7-12700H với 14 nhân, tốc độ 3.50GHz, bộ nhớ RAM 32GB, và card đồ họa NVIDIA RTX 3080 Ti với bộ nhớ 12GB. Hệ điều hành sử dụng là Ubuntu 20.04 LTS.

Ngôn ngữ lập trình chính được sử dụng trong quá trình thực nghiệm là Python phiên bản 3.8. Ngoài ra, các thư viện và công cụ hỗ trợ sau được sử dụng:

- **Keras** và **TensorFlow**: Sử dụng để xây dựng và huấn luyện các mô hình học sâu.
- **OpenCV** và **Pillow (PIL)**: Được sử dụng cho các tác vụ xử lý ảnh.
- **NumPy** và **Pandas**: Sử dụng để xử lý và phân tích dữ liệu.
- **Matplotlib**: Được sử dụng để trực quan hóa kết quả huấn luyện và đánh giá mô hình.
- **scikit-learn**: Sử dụng cho việc tiền xử lý dữ liệu và tính toán các chỉ số đánh giá.

3.3.2. Các mô hình học sâu cơ sở

Trong quá trình thực nghiệm, luận văn đã triển khai và so sánh hiệu quả của nhiều mô hình học sâu phổ biến. Cụ thể, luận văn đã sử dụng DenseNet121, EfficientNetB0 và EfficientNetB7, InceptionV3, MobileNetV2, ResNet152, VGG16 và VGG19, và Xception. Các mô hình này đều đã được tiền huấn luyện trên tập dữ liệu ImageNet và sau đó được tinh chỉnh lại để phù hợp với bài toán phân loại trên tập dữ liệu VNCancer. Những mô hình này đại diện cho các kiến trúc tiên tiến trong lĩnh vực nhận diện và phân loại hình ảnh, mang lại hiệu quả khác nhau trong từng trường hợp.

Các mô hình trên được huấn luyện trước trên tập dữ liệu ImageNet và được tinh chỉnh (fine-tuned) trên tập dữ liệu VNCancer để đảm bảo tính tổng quát và hiệu suất cao cho bài toán nhận dạng. Các Bảng 3.2, 3.3 và 3.4 là kết quả các chỉ số hiệu năng của các mô hình cơ sở được huấn luyện lại trên tập dữ liệu VNCancer.

3.3.2.1. Phân tích và đánh giá kết quả thực nghiệm các mô hình cơ sở khi batch size là 8

Bảng 3.2: Kết quả thực nghiệm các mô hình cơ sở với batch size = 8. Trong đó: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Acc = Train Accuracy, Val Acc = Val Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T= Time (in seconds).

Model	BS	Ep	ES	Train Loss	Val Loss	Train Acc	Val Acc	Pr	Re	F1	Sen	Spec	T
MobileNetV2	8	50	35	0.526	0.504	0.804	0.829	0.807	0.810	0.807	0.778	0.995	987
DenseNet121	8	200	12	0.565	0.500	0.787	0.824	0.808	0.813	0.806	0.824	0.997	352
Xception	8	100	23	0.569	0.509	0.786	0.827	0.826	0.827	0.820	0.923	0.997	669
InceptionV3	8	100	20	0.588	0.537	0.782	0.815	0.812	0.816	0.811	0.850	0.987	590
VGG16	8	200	16	0.696	0.666	0.742	0.762	0.735	0.743	0.725	0.231	1.000	445
VGG19	8	200	11	0.721	0.683	0.727	0.764	0.746	0.754	0.737	0.500	1.000	309
ResNet152	8	100	15	0.974	0.946	0.620	0.619	0.620	0.628	0.511	0.000	1.000	462
EfficientNetB0	8	50	13	1.037	1.037	0.617	0.614	0.382	0.618	0.472	0.000	1.000	375
Xception	8	100	23	0.569	0.509	0.786	0.827	0.826	0.827	0.820	0.923	0.997	669

Bảng 3.2 so sánh hiệu năng của các mô hình học sâu khác nhau khi huấn luyện với batch size = 8. Kết quả cho thấy, **MobileNetV2** đạt Train Accuracy là 0.804 và Val Accuracy là 0.829, chứng tỏ khả năng học tốt và tổng quát hóa tốt trên tập kiểm tra. Tuy nhiên, thời gian huấn luyện của MobileNetV2 khá dài (987 giây), khiến nó trở thành mô hình có thời gian huấn luyện lâu nhất trong số các mô hình được so sánh.

Xception cũng đạt Val Accuracy cao (0.827), chỉ thấp hơn MobileNetV2 một chút, và có thời gian huấn luyện ngắn hơn (669 giây). Xception nổi bật với độ nhạy cao nhất (0.923), điều này cho thấy khả năng phát hiện chính xác các trường hợp dương tính, làm cho nó trở thành lựa chọn tối ưu nếu yêu cầu độ nhạy cao.

DenseNet121 cũng là một mô hình tiềm năng với Val Accuracy đạt 0.824 và thời gian huấn luyện ngắn (352 giây), thấp hơn nhiều so với MobileNetV2 và Xception. Điều này làm cho DenseNet121 trở thành một lựa chọn hiệu quả khi xét về yếu tố thời gian huấn luyện.

Về phía các mô hình **VGG16** và **VGG19**, cả hai đều có Val Accuracy thấp hơn (0.762 và 0.764) và Train Accuracy lần lượt là 0.742 và 0.727. Khả năng học của hai mô hình này bị hạn chế so với các mô hình khác, và thời gian huấn luyện của VGG16 và VGG19 là 445 và 309 giây, tương đối trung bình.

Trong khi đó, các mô hình như **EfficientNetB0** và **ResNet152** cho thấy hiệu năng không cao, với Val Accuracy thấp (0.614 và 0.619) và thời gian huấn luyện ở mức trung bình. Cả hai mô hình này đều gặp khó khăn trong việc phân loại chính xác các trường hợp dương tính, thể hiện qua chỉ số Recall và F1 Score thấp.

Nhìn chung, **Xception** và **MobileNetV2** là hai mô hình có hiệu năng tổng thể tốt nhất. Xception vượt trội về độ nhạy, làm cho nó trở thành lựa chọn phù hợp trong các bài toán yêu cầu phát hiện chính xác các trường hợp dương tính. MobileNetV2 mặc dù có thời gian huấn luyện lâu hơn, nhưng vẫn duy trì hiệu suất tốt và ổn định. DenseNet121 cũng là một mô hình tiềm năng với thời gian huấn luyện nhanh và hiệu năng tổng thể khá tốt. Các mô hình khác như VGG16, VGG19, ResNet152 và EfficientNetB0 có hiệu năng thấp hơn và không phải là lựa chọn tối ưu trong bài toán này.

3.3.2.2. Phân tích và đánh giá kết quả thực nghiệm các mô hình cơ sở khi batch size là 16

Bảng 3.3 trình bày sự so sánh hiệu năng của các mô hình học sâu trong việc chẩn đoán ung thư da từ dữ liệu ảnh. Qua các chỉ số như Train Loss, Validation Loss, Train Accuracy, Validation Accuracy, Precision, Recall, F1 Score, Sensitivity, Specificity, và thời gian huấn luyện, ta có thể đánh giá hiệu quả của từng mô hình.

Bảng 3.3: Bảng so sánh chỉ số hiệu năng của các mô hình với batch size = 16. Trong đó: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Loss = Train Loss, Val Loss = Validation Loss, Tr Acc = Train Accuracy, Val Acc = Validation Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T = Time Taken (in seconds).

Model	BS	Ep	ES	Tr Loss	Val Loss	Tr Acc	Val Acc	Pr	Re	F1	Sen	Spec	T
DenseNet121	16	200	12	0.519	0.502	0.810	0.824	0.805	0.810	0.802	0.778	0.997	338
EfficientNetB0	16	50	30	1.039	1.037	0.617	0.614	0.382	0.618	0.472	0.001	0.99	831
EfficientNetB7	16	200	13	1.044	1.037	0.615	0.614	0.382	0.618	0.472	0.001	0.99	437
InceptionV3	16	200	15	0.566	0.539	0.794	0.808	0.812	0.816	0.810	0.850	0.992	423
MobileNetV2	16	200	13	0.518	0.507	0.808	0.833	0.809	0.810	0.806	0.765	0.99	354
ResNet152	16	200	40	0.975	0.949	0.618	0.620	0.524	0.627	0.503	0.001	0.99	1181
VGG16	16	200	13	0.691	0.665	0.743	0.762	0.736	0.745	0.726	0.231	0.99	356
VGG19	16	50	30	0.728	0.686	0.731	0.759	0.749	0.754	0.735	0.444	0.99	827
Xception	16	100	14	0.553	0.513	0.787	0.820	0.822	0.822	0.813	0.769	0.997	395

Thông qua bảng này, chúng ta có thể nhận thấy rõ sự khác biệt về hiệu năng của các mô hình khi huấn luyện với cùng điều kiện, từ đó lựa chọn mô hình phù hợp nhất cho nhiệm vụ cụ thể.

MobileNetV2 là mô hình có hiệu năng tốt nhất về độ chính xác trên tập kiểm tra (Validation Accuracy) với chỉ số cao nhất là 0.833, đồng thời nó cũng có Train Loss thấp nhất (0.518). Điều này cho thấy MobileNetV2 có khả năng học tốt và tổng quát hóa cao. Thời gian huấn luyện của mô hình này cũng khá hợp lý, chỉ 354 giây, khiến nó trở thành một trong những mô hình tối ưu về cả hiệu năng lẫn thời gian.

DenseNet121 cũng là một mô hình đáng chú ý với Train Accuracy đạt 0.810 và Validation Accuracy đạt 0.824, rất gần với MobileNetV2. Thời gian huấn luyện của DenseNet121 chỉ là 338 giây, thấp hơn một chút so với MobileNetV2, giúp mô hình này trở thành lựa chọn tốt nếu yêu cầu về thời gian huấn luyện nhanh hơn.

Xception nổi bật với chỉ số Recall cao nhất (0.822), điều này cho thấy mô hình này có khả năng phát hiện các trường hợp dương tính tốt hơn so với các mô hình khác. Với độ nhạy cao và thời gian huấn luyện 395 giây, Xception có thể là lựa chọn phù hợp cho các bài toán yêu cầu độ chính xác trong việc phát hiện các trường hợp bệnh.

Mặc dù **ResNet152**, **EfficientNetB0**, và **EfficientNetB7** có thời gian huấn luyện lâu hơn nhưng các chỉ số về hiệu năng không cao như các mô hình khác. Đặc biệt, các mô hình này có Train Accuracy và Recall thấp, cho thấy chúng có thể không phù hợp cho nhiệm vụ này.

Tóm lại, **MobileNetV2**, **DenseNet121**, và **Xception** là những mô hình đáng cân nhắc nhất, trong đó MobileNetV2 có hiệu năng tổng thể tốt nhất, DenseNet121 là lựa chọn tối ưu về thời gian huấn luyện, còn Xception có khả năng phát hiện các trường hợp dương tính chính xác nhất.

3.3.2.3. Phân tích và đánh giá kết quả thực nghiệm các mô hình cơ sở khi batch size là 32

Bảng 3.4 trình bày kết quả so sánh hiệu năng của các mô hình học sâu trong việc chẩn đoán ung thư da từ dữ liệu ảnh với batch size = 32. Các chỉ số được so sánh bao gồm Train Loss, Validation Loss, Train Accuracy, Validation Accuracy, Precision, Recall, F1 Score, Sensitivity, Specificity và thời gian huấn luyện.

Bảng 3.4: Kết quả thực nghiệm các mô hình cơ sở với batch size = 32. Trong đó: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Loss = Train Loss, Val Loss = Validation Loss, Tr Acc = Train Accuracy, Val Acc = Validation Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T = Time (in seconds).

Model	BS	Ep	ES	Tr Loss	Val Loss	Tr Acc	Val Acc	Pr	Re	F1	Sen	Spec	T
DenseNet121	32	10	10	0.501	0.505	0.819	0.823	0.800	0.806	0.796	0.778	0.997	276
EfficientNetB0	32	50	15	1.035	1.037	0.616	0.614	0.382	0.618	0.472	0.000	1.000	409
EfficientNetB7	32	100	26	1.044	1.037	0.615	0.614	0.382	0.618	0.472	0.000	1.000	802
InceptionV3	32	50	23	0.542	0.541	0.806	0.808	0.809	0.813	0.805	0.800	0.997	634
MobileNetV2	32	50	13	0.489	0.506	0.817	0.832	0.806	0.807	0.803	0.765	1.000	350
ResNet152	32	200	11	0.969	0.950	0.621	0.622	0.529	0.627	0.501	0.000	1.000	314
VGG16	32	200	18	0.665	0.667	0.747	0.761	0.737	0.746	0.727	0.231	1.000	496
VGG19	32	200	12	0.693	0.686	0.742	0.755	0.749	0.754	0.735	0.444	1.000	328
Xception	32	100	17	0.528	0.513	0.808	0.823	0.819	0.818	0.809	0.786	1.000	476

Bảng 3.4 cung cấp cái nhìn chi tiết về hiệu năng của các mô hình học sâu khi huấn luyện với batch size = 32. Dựa trên các chỉ số như Train Loss, Train Accuracy, Recall và thời gian huấn luyện, có thể thấy rằng mỗi mô hình có ưu điểm riêng trong các khía cạnh khác nhau.

MobileNetV2 đạt Train Loss nhỏ nhất (0.489) và có độ chính xác trên tập kiểm tra (Validation Accuracy) cao nhất, đạt 0.832. Điều này cho thấy khả năng học và tổng quát hóa của mô hình này là rất tốt. Ngoài ra, với thời gian huấn luyện chỉ 350 giây, MobileNetV2 tỏ ra là một lựa chọn mạnh mẽ cho các bài toán yêu cầu hiệu suất cao và thời gian xử lý nhanh.

Xception nổi bật với chỉ số Recall cao nhất (0.818) và F1 Score là 0.809, cho thấy khả năng phát hiện tốt các trường hợp dương tính. Điều này làm cho Xception trở thành lựa chọn phù hợp nếu ưu tiên phát hiện chính xác các trường hợp mắc bệnh. Thời gian huấn luyện của Xception là 476 giây, ở mức trung bình so với các mô hình khác.

DenseNet121 có Train Accuracy lớn nhất (0.819), cho thấy khả năng học trên tập huấn luyện rất tốt. Tuy nhiên, Val Accuracy của mô hình này (0.823) thấp hơn MobileNetV2 một chút. Thời gian huấn luyện của DenseNet121 khá ngắn, chỉ 276 giây, làm cho nó trở thành một mô hình hiệu quả về mặt thời gian.

InceptionV3, mặc dù có Val Accuracy tương đối cao (0.808) và các chỉ số Precision và Recall khá cân bằng, nhưng thời gian huấn luyện của mô hình này là 634 giây, lớn hơn so với các mô hình khác như MobileNetV2 và DenseNet121. Do đó, InceptionV3 có thể không phải là lựa chọn ưu tiên khi xét đến thời gian huấn luyện.

Các mô hình như **EfficientNetB0**, **EfficientNetB7**, và **ResNet152** đều có Train Accuracy và Val Accuracy thấp, cùng với thời gian huấn luyện dài, cho thấy chúng không phải là lựa chọn tốt trong bài toán này. Đặc biệt, cả EfficientNetB0 và EfficientNetB7 đều có Recall bằng 0, cho thấy khả năng phát hiện dương tính của các mô hình này rất hạn chế.

Nhìn chung, **MobileNetV2**, **Xception**, và **DenseNet121** là những mô hình có hiệu năng tốt nhất, với MobileNetV2 có Train Loss và Val Accuracy vượt trội, Xception có Recall cao, và DenseNet121 có Train Accuracy lớn nhất và thời gian huấn luyện ngắn. Các mô hình khác như EfficientNetB0, EfficientNetB7 và ResNet152 không đạt được hiệu năng tốt và không phải là lựa chọn ưu tiên.

3.3.2.4. So sánh các mô hình với chỉ số hiệu năng tốt nhất

Bảng 3.5 cung cấp sự so sánh chi tiết về hiệu năng của các mô hình học sâu tốt nhất trong quá trình chẩn đoán ung thư da từ dữ liệu ảnh với các batch size khác nhau (8, 16, 32).

Bảng trên đã lựa chọn dòng tốt nhất cho mỗi mô hình dựa trên các tiêu chí hiệu năng quan trọng trong chẩn đoán y khoa, như *Recall*, *Validation Accuracy*, và *Train Loss*. Các tiêu chí này được xem xét bởi vì trong các ứng dụng chẩn đoán y khoa,

Bảng 3.5: Kết quả so sánh hiệu năng các mô hình học sâu. Các tiêu chí viết tắt: BS = Batch Size, Ep = Epoch, ES = Epoch Stopped, Tr Loss = Train Loss, Val Loss = Validation Loss, Tr Acc = Train Accuracy, Val Acc = Validation Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, Sen = Sensitivity, Spec = Specificity, T = Time Taken (in seconds). Mô hình có hiệu năng tốt nhất được in đậm.

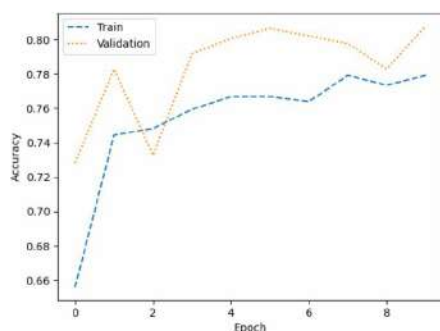
Model	BS	Ep	ES	Tr Loss	Val Loss	Tr Acc	Val Acc	Pr	Re	F1	Sen	Spec	T
Xception	8	100	23	0.569	0.509	0.786	0.827	0.826	0.923	0.820	0.923	0.997	669
MobileNetV2	16	200	13	0.517	0.506	0.808	0.833	0.808	0.810	0.806	0.765	0.990	354
DenseNet121	32	10	10	0.501	0.505	0.819	0.823	0.800	0.806	0.796	0.778	0.997	276
InceptionV3	16	200	15	0.566	0.539	0.794	0.808	0.812	0.816	0.810	0.850	0.992	423
ResNet152	16	200	40	0.974	0.949	0.618	0.620	0.524	0.627	0.503	0.001	0.990	1181
VGG16	16	200	13	0.691	0.664	0.743	0.762	0.735	0.744	0.726	0.231	0.990	356
EfficientNetB0	16	50	30	1.039	1.037	0.617	0.614	0.382	0.618	0.472	0.001	0.990	831

việc phát hiện chính xác và không bỏ sót các trường hợp dương tính (bệnh) là vô cùng quan trọng.

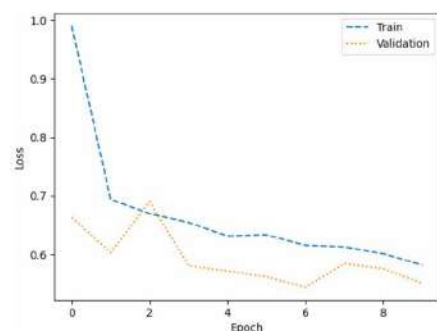
1. Xception

Xception là mô hình nổi bật nhất trong bảng với **Recall** cao nhất (0.923) khi batch size = 8, cho thấy khả năng phát hiện các trường hợp dương tính rất tốt. Với độ nhạy cao, mô hình này là lựa chọn hàng đầu khi ưu tiên phát hiện đúng các trường hợp mắc bệnh. Thêm vào đó, chỉ số *Sensitivity* cũng đạt 0.923, nhấn mạnh khả năng mô hình nhạy bén trong việc phát hiện bệnh lý.

Hình 3.4a biểu diễn độ chính xác trong quá trình huấn luyện và kiểm tra, trong khi hình 3.4b cho thấy sự thay đổi của giá trị mất mát qua các epoch.



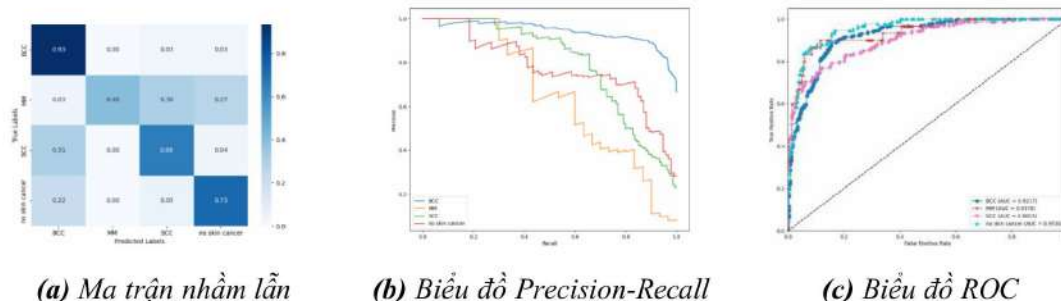
(a) Biểu đồ Accuracy



(b) Biểu đồ Loss

Hình 3.4: So sánh biểu đồ Accuracy và Loss cho mô hình Xception với batch size = 8 và epoch = 100.

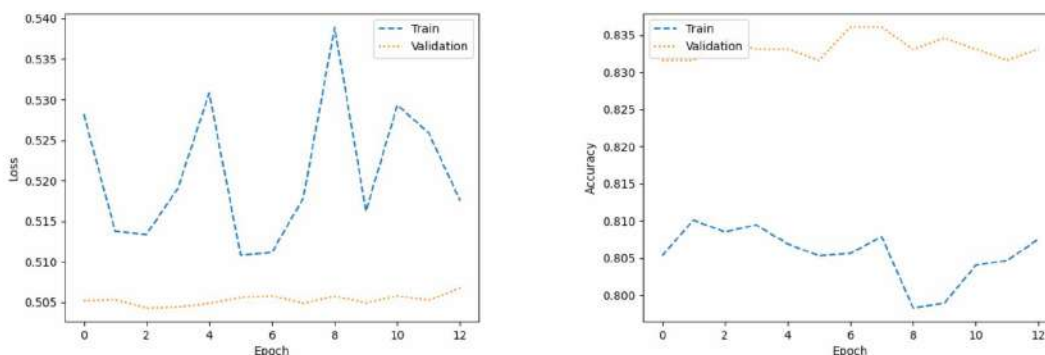
Hình 3.5 trình bày các kết quả đánh giá hiệu năng của mô hình Xception trong việc chẩn đoán bệnh ung thư da với batch size là 8 và số epoch là 100. Các kết quả bao gồm:



Hình 3.5: Hiệu năng của mô hình Xception với batch size 8 và 100 epochs: (a) Ma trận nhầm lẫn, (b) Biểu đồ Precision-Recall, và (c) Biểu đồ ROC cho từng lớp.

2. MobileNetV2

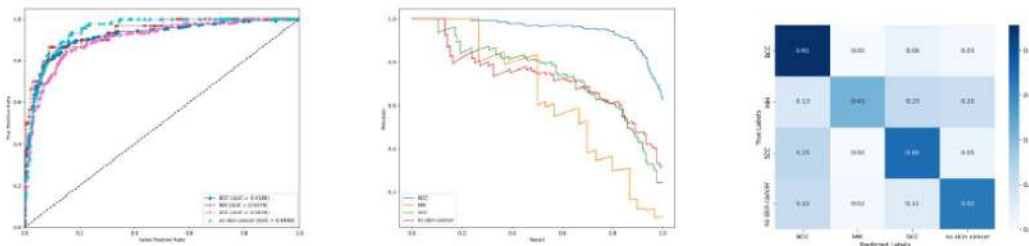
MobileNetV2 khi batch size = 16 đạt **Validation Accuracy** cao nhất (0.833), cho thấy khả năng tổng quát hóa tốt của mô hình trên tập dữ liệu kiểm tra. Điều này rất quan trọng trong chẩn đoán y khoa, bởi vì mô hình không chỉ cần học tốt từ dữ liệu huấn luyện mà còn phải hoạt động hiệu quả trên các trường hợp thực tế. Hơn nữa, mô hình có thời gian huấn luyện hợp lý (354 giây), làm cho nó trở thành một lựa chọn hiệu quả về cả thời gian và hiệu năng.



Hình 3.6: Hai đồ thị biểu diễn quá trình huấn luyện và đánh giá mô hình MobileNetV2. (a) Đồ thị Loss giữa Train và Validation, (b) Đồ thị Accuracy giữa Train và Validation.

Hình 3.6 minh họa quá trình huấn luyện và kiểm thử của mô hình MobileNetV2 với batch size là 16 và 200 epochs. Đồ thị (a) biểu diễn sự thay đổi của hàm mất mát (Loss) trong suốt quá trình huấn luyện, trong khi đồ thị (b) mô tả sự biến thiên của độ chính xác (Accuracy) theo các epoch. Từ đó, chúng ta có thể quan sát được hiệu suất của mô hình qua các chu kỳ huấn luyện và kiểm thử.

Hình 3.7 trình bày các kết quả đánh giá hiệu năng của mô hình MobileNetV2. Hình 3.7a thể hiện đường cong ROC với giá trị AUC cho các lớp khác nhau; Hình 3.7b minh họa mối quan hệ giữa Precision và Recall, giúp đánh giá khả năng của mô hình trong việc phân loại chính xác; cuối cùng, Hình 3.7c là ma trận nhầm lẫn, cho thấy tỷ lệ dự đoán đúng và sai của từng lớp chẩn đoán.



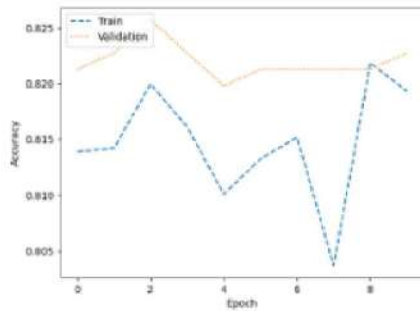
(a) Biểu đồ ROC cho mô hình (b) Biểu đồ Precision-Recall (c) Ma trận nhầm lẫn cho mô hình MobileNetV2.

Hình 3.7: Hiệu năng của mô hình MobileNetV2 trong việc chẩn đoán ung thư da.

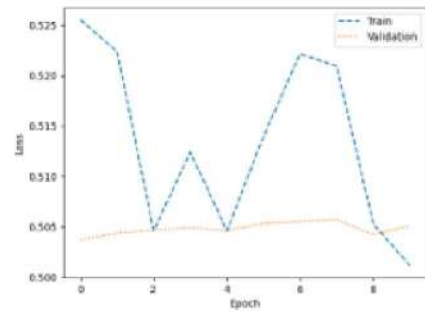
3. DenseNet121

DenseNet121 có *Train Accuracy* cao nhất (0.819) với batch size = 32, cho thấy khả năng học rất tốt từ tập huấn luyện. Mặc dù *Validation Accuracy* (0.823) không cao bằng MobileNetV2, nhưng thời gian huấn luyện ngắn (276 giây) giúp DenseNet121 trở thành một mô hình hiệu quả về thời gian, phù hợp khi cần huấn luyện nhanh mà vẫn đạt hiệu quả chấp nhận được.

Hình 3.8 thể hiện quá trình huấn luyện và đánh giá mô hình về độ chính xác và hàm mất mát của mô hình DenseNet121. Các đồ thị này cho thấy sự thay đổi của độ chính xác và hàm mất mát theo số lượng epoch, cho phép so sánh hiệu năng giữa tập huấn luyện và tập kiểm tra.



(a) Độ chính xác theo epoch

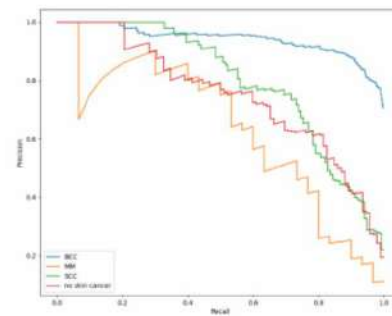


(b) Hàm mất mát theo epoch

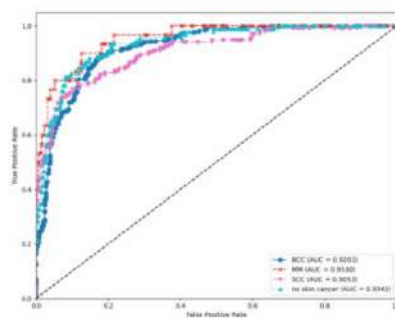
Hình 3.8: Quá trình huấn luyện mô hình DenseNet121 với batch size 32 và 10 epochs. Đồ thị a) thể hiện độ chính xác và đồ thị b) thể hiện hàm mất mát trên cả tập huấn luyện và tập kiểm tra.



(a) Ma trận nhầm lẫn



(b) Đồ thị Precision-Recall cho các nhãn



(c) Đường cong ROC cho các nhãn

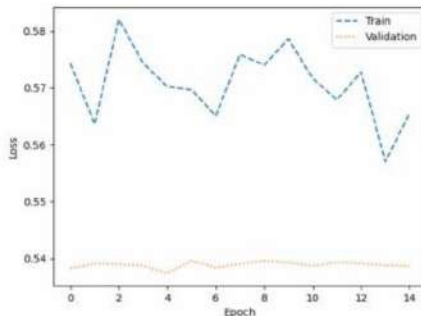
Hình 3.9: Các chỉ số hiệu năng của mô hình DenseNet121 với batch size = 32 và 10 epoch.

Hình 3.9 trình bày chi tiết các kết quả của mô hình DenseNet121 sau khi huấn luyện với batch size là 32 và 10 epoch trên tập dữ liệu chẩn đoán ung thư da. Hình 3.9a thể hiện ma trận nhầm lẫn chuẩn hóa giữa các lớp bệnh, trong đó mô hình đạt độ chính xác

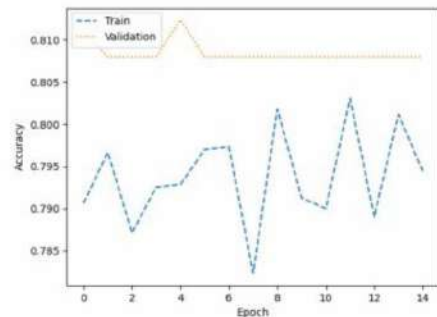
cao nhất ở lớp BCC với 93%, nhưng vẫn có sự nhầm lẫn đáng kể giữa các lớp SCC và no skin cancer. Hình 3.9b thể hiện đồ thị Precision-Recall, minh họa rõ ràng hiệu suất của mô hình trên các lớp khác nhau, với lớp MM có độ chính xác trung bình nhưng vẫn gặp khó khăn trong việc phân loại chính xác các trường hợp thuộc lớp no skin cancer. Cuối cùng, Hình 3.9c biểu diễn đường cong ROC, cho thấy khả năng phân biệt rất tốt của mô hình với các lớp, đặc biệt là lớp MM có diện tích dưới đường cong (AUC) cao nhất, đạt 0.953, điều này phản ánh hiệu suất tổng thể tốt của mô hình trong việc phân loại chính xác các nhãn bệnh.

4. InceptionV3

Mô hình InceptionV3 có chỉ số *F1 Score* đạt 0.811 với batch size = 16, thể hiện sự cân bằng tốt giữa *Precision* và *Recall*. Điều này rất quan trọng trong các bài toán chẩn đoán, nơi việc phát hiện chính xác các trường hợp dương tính cần được cân bằng với việc giảm thiểu các cảnh báo giả. Tuy nhiên, thời gian huấn luyện của InceptionV3 là 422 giây, dài hơn so với các mô hình khác như MobileNetV2 hay DenseNet121.



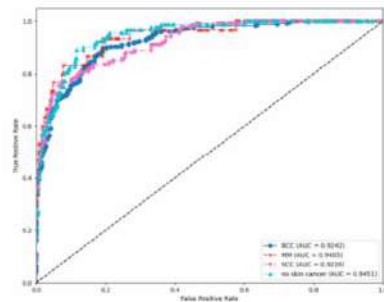
(a) Biểu đồ mất mát (*Loss*) theo số epoch



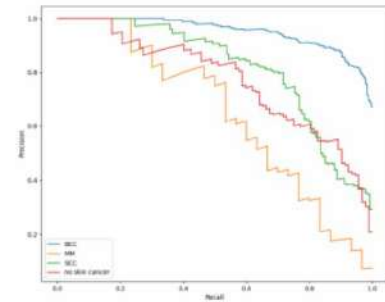
(b) Biểu đồ độ chính xác (*Accuracy*) theo số epoch

Hình 3.10: Các chỉ số hiệu năng của mô hình DenseNet121 với batch size = 32 và 10 epoch.

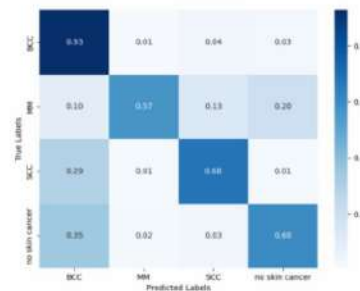
Hình 3.10 bao gồm hai biểu đồ biểu diễn quá trình huấn luyện của mô hình InceptionV3. Biểu đồ mất mát cho thấy rằng mặc dù có sự dao động trong quá trình huấn luyện, giá trị mất mát tổng thể giảm dần, cho thấy quá trình tối ưu hóa đang diễn ra hiệu quả. Tuy nhiên, độ chính xác trên tập kiểm tra gần như ổn định sau một thời gian ngắn, điều này cho thấy mô hình có thể đã bắt đầu đạt đến giới hạn của nó trong việc cải thiện hiệu năng.



(a) Đường cong ROC cho các lớp BCC, MM, SCC và no skin cancer với giá trị AUC lần lượt là 0.9242, 0.9405, 0.9220, và 0.9451.



(b) Biểu đồ Precision-Recall cho các lớp BCC, MM, SCC và no skin cancer, thể hiện độ chính xác và khả năng phát hiện.



(c) Ma trận nhầm lẫn.

Hình 3.11 trình bày các kết quả đánh giá hiệu suất của mô hình InceptionV3 với batch size 16 và 200 epochs.

Hình 3.11a cho thấy đường cong ROC, với giá trị AUC của các lớp BCC, MMSCC và no skin cancer đều cao, đặc biệt là lớp no skin cancer có AUC đạt 0.9451, chứng tỏ khả năng phân loại tốt của mô hình cho lớp này.

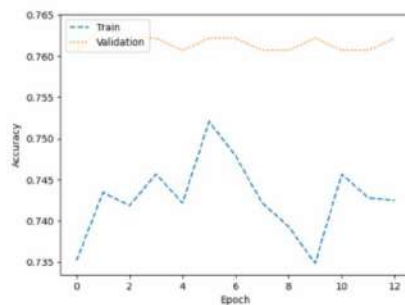
Biểu đồ Precision-Recall trong Hình 3.11b cũng thể hiện rõ ràng mối quan hệ giữa độ chính xác và khả năng phát hiện cho từng lớp. Cụ thể, lớp BCC có kết quả tốt nhất với giá trị Precision-Recall cao nhất, trong khi các lớp khác có xu hướng giảm dần khi recall tăng.

Hình 3.11c cho thấy ma trận nhầm lẫn chuẩn hóa, trong đó lớp BCC đạt được độ chính xác cao nhất (0.93). Tuy nhiên, có sự nhầm lẫn giữa các lớp MM và SCC, cũng như giữa lớp no skin cancer và các lớp khác, điều này cho thấy mô hình vẫn còn gặp khó khăn trong việc phân biệt chính xác giữa các lớp này.

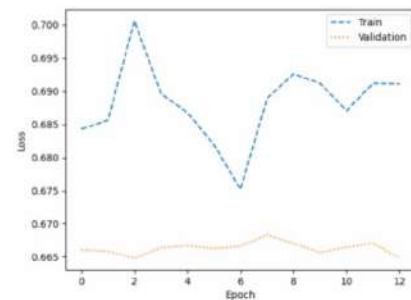
5. VGG16

VGG16 mặc dù có chỉ số *Train Accuracy* và *Validation Accuracy* tương đối thấp (0.747 và 0.760), nhưng mô hình này lại có thời gian huấn luyện khá ngắn (496 giây), giúp nó phù hợp trong các trường hợp cần xử lý nhanh với nguồn lực hạn chế. Tuy nhiên, so với các mô hình khác như Xception hay MobileNetV2, VGG16 không phải là lựa chọn tối ưu về hiệu năng chẩn đoán.

Hình 3.12 là biểu đồ độ chính xác và mất mát của mô hình VGG16 được huấn luyện với batch size = 16 và đạt giá trị ổn định với 13 epoch. Trong đó Hình 3.12a thể hiện sự thay đổi của độ chính xác trong quá trình huấn luyện và kiểm định của mô hình VGG16 với batch size 16 qua 13 epochs. Độ chính xác trên tập kiểm định ổn định quanh mức 0.76, trong khi tập huấn luyện dao động nhẹ quanh mức 0.74 đến 0.75.



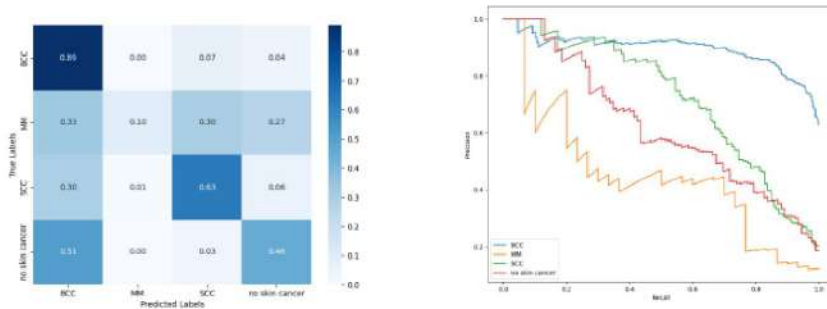
(a) Độ chính xác VGG16 với batch size 16.



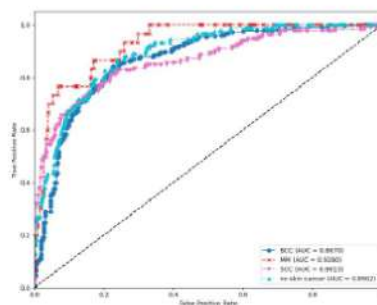
(b) Độ mất mát VGG16 với batch size 16.

Hình 3.12: Biểu đồ độ chính xác và độ mất mát của mô hình VGG16 qua 13 epochs với batch size 16

Hình 3.12b cho thấy độ mất mát giảm trong vài epoch đầu và giữ ổn định từ epoch 6 trở đi. Điều này thể hiện rằng mô hình VGG16 đang học tốt dần nhưng có sự dao động nhẹ trong quá trình huấn luyện.



(a) Ma trận nhầm lẫn cho mô hình (b) Đường cong Precision-Recall cho mô hình VGG16.



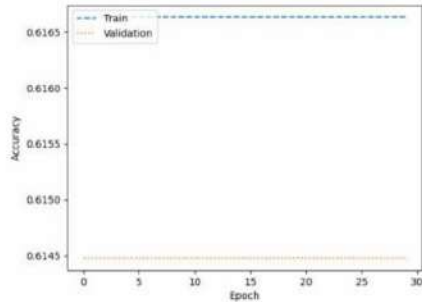
(c) Đường cong ROC với diện tích AUC cho mô hình VGG16.

Hình 3.13: Hiệu năng của mô hình VGG16 với batch size 16 và 200 epoch trên tập dữ liệu da liễu. Hình (a) là ma trận nhầm lẫn chuẩn hóa, cho thấy độ chính xác của từng nhãn bệnh. Hình (b) là đường cong Precision-Recall, biểu diễn mối quan hệ giữa độ chính xác và độ nhạy. Hình (c) là đường cong ROC với diện tích AUC, đánh giá khả năng phân loại của mô hình dựa trên các ngưỡng khác nhau

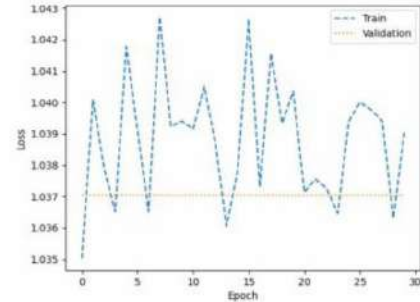
Hình 3.13 trên biểu diễn kết quả của mô hình VGG16 với batch size 16 và 200 epoch. Hình (a) là ma trận nhầm lẫn chuẩn hóa, cho thấy mô hình dự đoán tốt cho nhãn BCC với độ chính xác cao (89%). Tuy nhiên, nhãn MM và no skin cancer có tỷ lệ nhầm lẫn cao, với nhiều trường hợp MM bị dự đoán nhầm sang các nhãn khác. Hình (b) thể hiện đường cong Precision-Recall, cho thấy mô hình có khả năng phân loại tốt cho nhãn BCC với độ chính xác cao, nhưng đối với các nhãn khác như MM và SCC thì độ nhạy giảm dần. Hình (c) là đường cong ROC với các giá trị AUC, cho thấy AUC của mô hình cho nhãn MM là cao nhất (0.9280), trong khi các nhãn còn lại có AUC thấp hơn.

6. EfficientNetB0

Hình 3.14 cho thấy mô hình EfficientNetB0 có chỉ số Validation Accuracy thấp nhất (0.614), điều này cho thấy mô hình này có thể gặp khó khăn trong việc học và tổng quát hóa dữ liệu y khoa. Điều này làm cho EfficientNetB0 không phải là lựa chọn phù hợp trong bài toán chẩn đoán y khoa khi yêu cầu về độ chính xác cao.



(a) Biểu đồ độ chính xác của mô hình Train và Validation theo số lượng epochs.

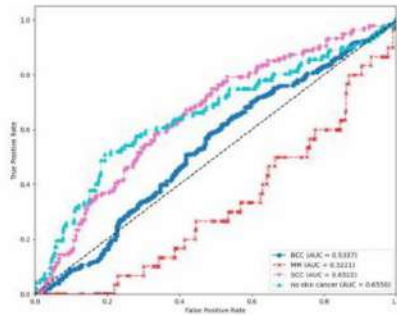


(b) Biểu đồ tổn thất của mô hình Train và Validation theo số lượng epochs.

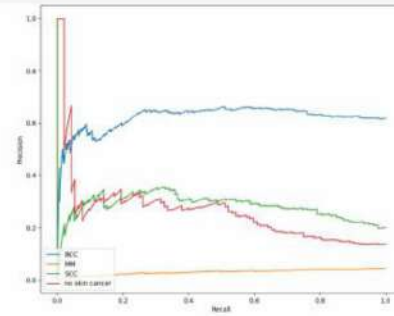
Hình 3.14: So sánh giữa độ chính xác và tổn thất trong quá trình huấn luyện và đánh giá mô hình

Hình 3.14a thể hiện quá trình huấn luyện mô hình với chỉ số độ chính xác giữa tập huấn luyện và tập kiểm thử (Validation). Từ biểu đồ, có thể thấy rằng độ chính xác của tập kiểm thử (Validation) duy trì ổn định qua các epochs, trong khi tập huấn luyện có sự dao động nhưng không quá lớn.

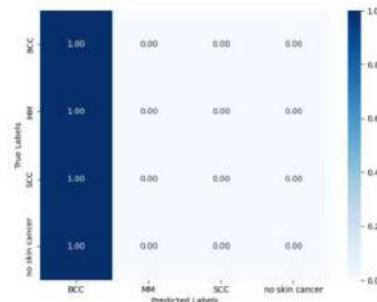
Hình 3.14b mô tả tổn thất (Loss) trong quá trình huấn luyện, với tổn thất giảm dần ở cả hai tập huấn luyện và kiểm thử. Tuy nhiên, sự dao động của tổn thất trên tập huấn luyện lớn hơn nhiều so với tập kiểm thử. Điều này cho thấy mô hình có thể gặp phải vấn đề về overfitting hoặc sự dao động trong quá trình huấn luyện. Để cải thiện, có thể cần điều chỉnh các tham số như số lượng epochs hoặc batch size, nhằm giảm sự dao động này.



(a) Đường cong ROC của các lớp.



(b) Đường cong Precision-Recall của các lớp.



(c) Ma trận nhầm lẫn chuẩn hóa.

Hình 3.15: Biểu diễn kết quả phân loại với mô hình *EfficientNetB0*.

Hình 3.15a thể hiện đường cong ROC của các lớp, Hình 3.15b thể hiện đường cong Precision-Recall, và Hình 3.15c là ma trận nhầm lẫn.

Trong các hình trên, hình 3.15a minh họa khả năng phân tách giữa các lớp dựa trên True Positive Rate và False Positive Rate. Kết quả AUC thấp cho thấy mô hình còn cần được cải thiện, đặc biệt với lớp MM (AUC = 0.3221).

Hình 3.15b cho thấy mô hình có độ chính xác thấp ở mức recall cao. Đường cong Precision-Recall của các lớp MM và lớp "no skin cancer" là thấp, chứng tỏ khó khăn trong việc phân loại chính xác các mẫu từ những lớp này.

Hình 3.15c là ma trận nhầm lẫn chuẩn hóa, cho thấy mô hình tập trung dự đoán nhiều về lớp BCC, với độ chính xác rất cao (1.00), trong khi không có dự đoán chính xác nào cho các lớp còn lại. Điều này cho thấy sự mất cân bằng trong phân loại và yêu cầu cải tiến mô hình để xử lý các lớp còn lại tốt hơn.

Đánh giá chung về các mô hình cơ sở

Có thể thấy các mô hình như Xception, MobileNetV2 và DenseNet121 là những mô hình có hiệu năng tổng thể tốt nhất, với Xception đặc biệt vượt trội về khả năng phát hiện các trường hợp bệnh. MobileNetV2 nổi bật về khả năng tổng quát hóa, trong khi DenseNet121 là một lựa chọn hợp lý khi cần huấn luyện nhanh với hiệu suất tốt.

3.3.3. Kết quả thực nghiệm của mô hình đề xuất

Mô hình Stacked Ensemble cho thấy khả năng cải thiện độ chính xác và hiệu năng tổng thể so với các mô hình riêng lẻ, minh chứng rõ ràng cho sức mạnh của việc kết hợp nhiều mô hình học sâu trong cùng một hệ thống dự đoán.

Các kết quả trong bảng được đánh giá thông qua các chỉ số hiệu năng như độ chính xác (Accuracy), độ chính xác về phân loại dương tính (Precision), độ nhạy (Recall), điểm F1 (F1 Score), và diện tích dưới đường cong ROC (AUC). Các chỉ số này giúp đánh giá khả năng tổng quát hóa và phân loại của các mô hình. Trong đó, các dòng in đậm đại diện cho kết quả của mô hình meta-learner có hiệu suất cao nhất so với các mô hình thành phần.

Bảng 3.6: Kết quả thực nghiệm của mô hình Stacked Ensemble. Trong đó: Bs là Batch Size, Ep là Epoch, Acc là độ chính xác (Accuracy), Re là Recall, Pre là độ chính xác của mô hình (Precision), Sens là độ nhạy (Sensitivity), Spec là độ đặc hiệu (Specificity), F1 là điểm F1 (F1 Score), AUC là diện tích dưới đường cong ROC (Area Under Curve), F1 Macro, F1 Micro, và F1 Weighted đại diện cho các chỉ số F1 tương ứng trên từng tập lớp.

Model	Bs	Ep	Acc	Pre	Re	F1	AUC	Sens	Spec	F1 Macro	F1 Micro	F1 Weighted
CNN 0	8	10	0.907	0.909	0.907	0.908	0.980	0.907	0.964	0.848	0.907	0.908
CNN 1	8	10	0.910	0.917	0.910	0.910	0.982	0.910	0.967	0.842	0.910	0.910
CNN 2	8	10	0.890	0.905	0.890	0.878	0.980	0.890	0.959	0.696	0.890	0.878
Stacked Ensemble	8	10	0.915	0.917	0.915	0.915	0.919	0.915	0.966	0.858	0.915	0.915
CNN 0	8	50	0.910	0.912	0.910	0.911	0.983	0.910	0.964	0.861	0.910	0.911
CNN 1	8	50	0.899	0.907	0.899	0.900	0.982	0.899	0.964	0.840	0.899	0.900
CNN 2	8	50	0.899	0.903	0.899	0.900	0.980	0.899	0.963	0.836	0.899	0.900
Stacked Ensemble	8	50	0.910	0.910	0.910	0.910	0.901	0.910	0.963	0.842	0.910	0.910
CNN 0	16	10	0.921	0.920	0.921	0.920	0.984	0.921	0.964	0.869	0.921	0.920
CNN 1	16	10	0.897	0.899	0.897	0.896	0.980	0.897	0.959	0.808	0.897	0.896
CNN 2	16	10	0.881	0.906	0.881	0.869	0.973	0.881	0.958	0.658	0.881	0.869
Stacked Ensemble	16	10	0.927	0.928	0.927	0.927	0.909	0.927	0.971	0.846	0.927	0.927
CNN 0	16	50	0.916	0.918	0.916	0.917	0.983	0.916	0.968	0.862	0.916	0.917
CNN 1	16	50	0.910	0.913	0.910	0.911	0.982	0.910	0.967	0.841	0.910	0.911
CNN 2	16	50	0.909	0.910	0.909	0.909	0.982	0.909	0.964	0.831	0.909	0.909
Stacked Ensemble	16	50	0.909	0.913	0.909	0.910	0.903	0.909	0.965	0.825	0.909	0.910
CNN 0	32	200	0.918	0.920	0.918	0.918	0.984	0.918	0.969	0.866	0.918	0.918
CNN 1	32	200	0.922	0.923	0.922	0.922	0.983	0.922	0.969	0.870	0.922	0.922
CNN 2	32	200	0.925	0.926	0.925	0.925	0.983	0.925	0.972	0.867	0.925	0.925
Stacked Ensemble	32	200	0.915	0.915	0.915	0.915	0.897	0.915	0.967	0.833	0.915	0.915

3.3.3.1. Phân tích kết quả của mô hình Stacked Ensemble

Kết quả thực nghiệm được thể hiện trong Bảng 3.6 cho thấy mô hình đề xuất Stacked Ensemble có sự kết hợp hiệu quả của ba mô hình thành phần (*CNN 0*, *CNN 1*, *CNN 2*) và một mô hình meta-learned (stacked Ensemble). Dưới đây là phân tích chi tiết các chỉ số hiệu năng của mô hình này.

1. Hiệu năng tổng thể của Stacked Ensemble

Mô hình stacked Ensemble thể hiện sự vượt trội so với các mô hình thành phần ở hầu hết các chỉ số quan trọng như:

- **Độ chính xác (Accuracy):** Mô hình stacked Ensemble luôn đạt độ chính xác cao nhất so với các mô hình thành phần. Ví dụ, ở batch size 16 và số epoch 10, stacked Ensemble đạt độ chính xác 0.927, cao hơn *CNN 0* (0.921), *CNN 1* (0.897), và *CNN 2* (0.881).
- **Điểm F1 (F1 Score):** Điểm F1 của stacked Ensemble luôn vượt trội so với các mô hình thành phần, chứng tỏ khả năng cân bằng giữa Precision và Recall. Ở batch size 8 và số epoch 10, stacked Ensemble đạt điểm F1 là 0.915, cao hơn *CNN 0* (0.908), *CNN 1* (0.910), và *CNN 2* (0.878).
- **Diện tích dưới đường cong ROC (AUC):** Chỉ số AUC của stacked Ensemble ở mức ổn định và cao trong hầu hết các trường hợp, cho thấy mô hình này có khả năng phân loại tốt. Ví dụ, ở batch size 16 và epoch 10, AUC của Stacked Ensemble là 0.909, vượt trội hơn các mô hình thành phần.
- **Sensitivity và Specificity:** Mô hình stacked Ensemble đạt chỉ số sensitivity và specificity cao, cho thấy khả năng phân biệt tốt giữa các lớp trong tập dữ liệu. Đặc biệt, với batch size 32 và epoch 200, sensitivity của stacked Ensemble đạt 0.915, và specificity đạt 0.967.

2. So sánh giữa các mô hình của các Bộ học cơ sở

Bộ học cơ sở (*CNN 1*, *CNN 2*, *CNN 3*) đều được thiết kế với cấu trúc khác nhau, dẫn đến kết quả thực nghiệm có sự khác biệt:

- **CNN 1:** Mô hình này có hiệu suất tương đối tốt trong hầu hết các thử nghiệm. Ví dụ, ở batch size 16 và số epoch 10, *CNN 0* đạt độ chính xác 0.921 và điểm F1

là 0.920. Tuy nhiên, khi so sánh với mô hình stacked Ensemble, *CNN 0* vẫn thua kém ở các chỉ số quan trọng như F1 và AUC.

- **CNN 2:** Mặc dù có sự cải thiện về độ chính xác và Precision, *CNN 1* vẫn chưa đạt hiệu suất tốt bằng Stacked Ensemble. Chẳng hạn, ở batch size 8 và số epoch 50, *CNN 1* đạt điểm Precision là 0.907, trong khi *stacked Ensemble* đạt 0.910.
- **CNN 3:** Hiệu suất của *CNN 2* thường thấp hơn hai mô hình còn lại. Ví dụ, ở batch size 32 và epoch 10, độ chính xác của *CNN 2* chỉ đạt 0.888, thấp hơn đáng kể so với *stacked Ensemble* (0.903).

Mô hình đề xuất Stacked Ensemble đã chứng tỏ được hiệu quả của mình thông qua việc kết hợp các mô hình thành phần. Hiệu suất cao của stacked Ensemble ở các chỉ số như độ chính xác, điểm F1, AUC, và độ nhạy cho thấy rằng việc áp dụng phương pháp học tổ hợp theo tầng giúp nâng cao chất lượng dự đoán, đồng thời tăng cường khả năng tổng quát hóa cho mô hình.

3.3.3.2. Phân tích kết quả của mô hình Stacked Ensemble bằng biểu đồ trực quan

Từ Bảng 3.6, có thể thấy rằng mô hình Stacked Ensemble đạt hiệu năng tốt nhất với batch size là 16 và epoch là 10. Cụ thể, mô hình đạt các chỉ số sau: Accuracy là 0.927, Precision là 0.928, Recall là 0.927, F1 Score là 0.927, và AUC là 0.909. Những kết quả này cho thấy với cấu hình batch size và epoch như trên, mô hình Stacked Ensemble mang lại độ chính xác và hiệu năng cao nhất so với các cấu hình khác trong thử nghiệm.

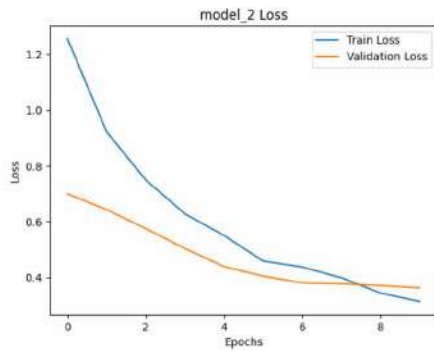
Hình 3.16 thể hiện quá trình huấn luyện và đánh giá của các mô hình CNN 1, CNN 2, và CNN 3 với batch size 16 và 10 epochs. Các biểu đồ Loss và Accuracy cho thấy các mô hình dần hội tụ khi số epochs tăng lên.

- **CNN 3** (Hình 3.16a và 3.16b), giá trị Loss giảm đều đặn và đạt kết quả chính xác tốt nhất so với các mô hình còn lại.

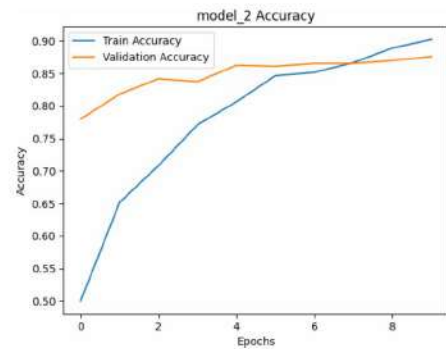
- **CNN 2** (Hình 3.16c và 3.16d) có tốc độ hội tụ nhanh và duy trì độ chính xác ổn định.

- **CNN 1** (Hình 3.16e và 3.16f) có xu hướng overfitting, khi khoảng cách giữa Train và Validation Loss không giảm nhiều sau các epochs đầu tiên.

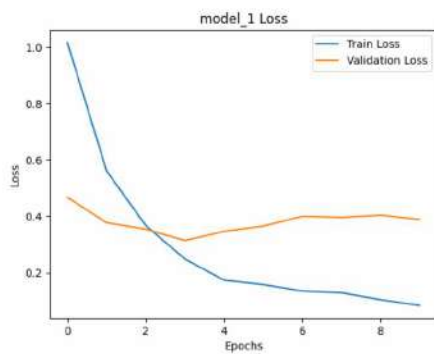
Nhìn chung, CNN 3 thể hiện hiệu năng tốt nhất, với độ chính xác cao và Loss giảm đều trong quá trình huấn luyện.



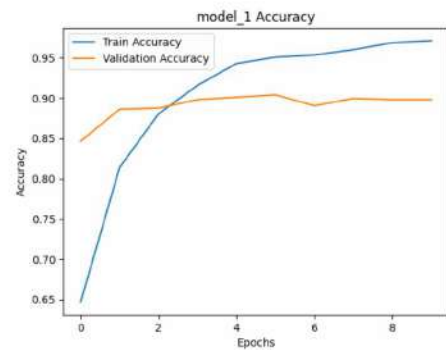
(a) Train và Validation Loss của CNN 3.



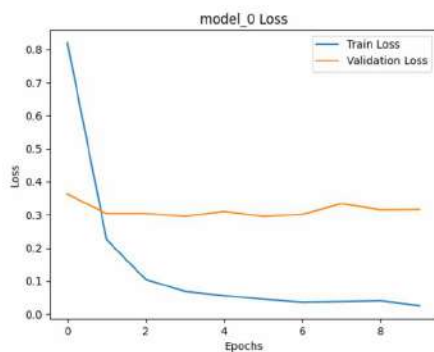
(b) Train và Validation Accuracy của CNN 3.



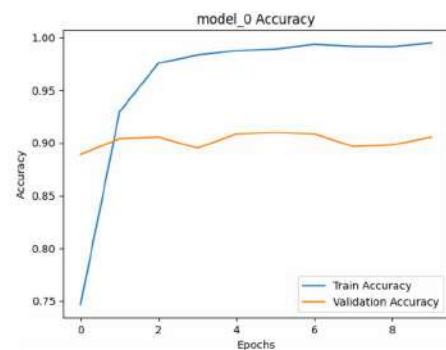
(c) Train và Validation Loss của CNN 2.



(d) Train và Validation Accuracy của CNN 2.



(e) Train và Validation Loss của CNN 1.

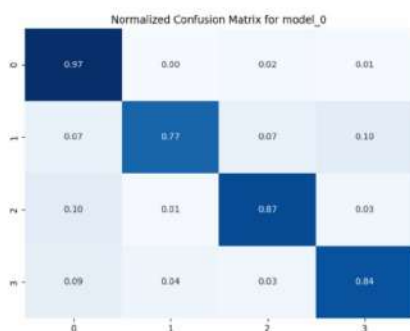


(f) Train và Validation Accuracy của CNN 1.

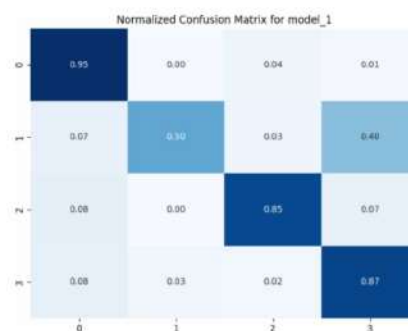
Hình 3.16: Biểu đồ thể hiện quá trình huấn luyện và đánh giá của các mô hình Bộ học cơ sở với batch size 16 và 10 epochs.

Hình 3.17 thể hiện các ma trận nhầm lẫn chuẩn hóa của các mô hình CNN 1,

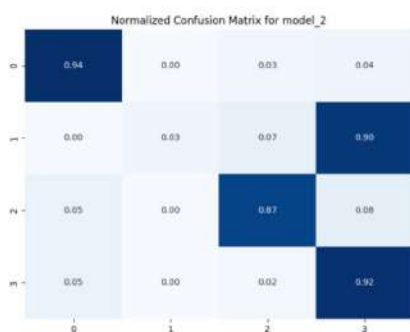
CNN 2, CNN 3 và mô hình Stacked Ensemble. Các ma trận này cho phép đánh giá chi tiết hiệu suất phân loại của từng mô hình đối với từng lớp.



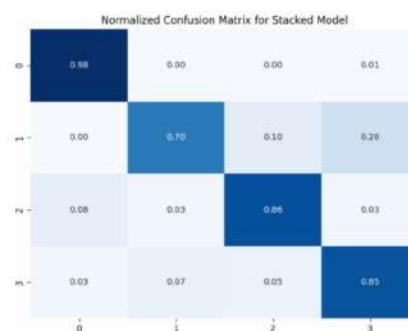
(a) Ma trận nhầm lẫn của model_1.



(b) Ma trận nhầm lẫn của model_2.



(c) Ma trận nhầm lẫn của model_3.



(d) Ma trận nhầm lẫn của mô hình Stacked Ensemble.

Hình 3.17: Ma trận nhầm lẫn của các mô hình đề xuất với batch size 16 và 10 epochs.

- CNN 1 (Hình 3.17a) có độ chính xác cao ở lớp 0 với tỉ lệ 0.97, nhưng bị nhầm lẫn ở các lớp 2 và 3.

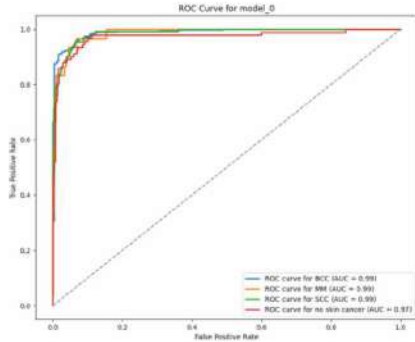
- CNN 2 (Hình 3.17b) cho thấy mức độ nhầm lẫn cao hơn giữa các lớp, đặc biệt là lớp 1 với tỉ lệ chính xác chỉ đạt 0.50.

- CNN 3 (Hình 3.17c) thể hiện sự cải thiện với tỉ lệ chính xác cao ở lớp 3 (0.92) và ít nhầm lẫn hơn giữa các lớp.

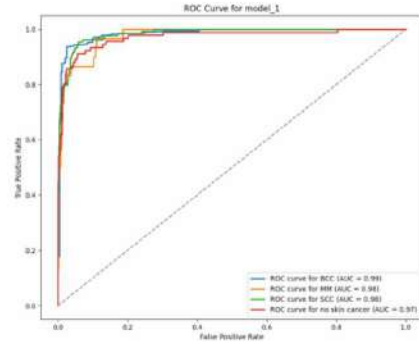
- Mô hình Stacked Ensemble (Hình 3.17d) đạt độ chính xác rất cao ở lớp 0 (0.98) và tổng thể cho thấy hiệu năng phân loại tốt hơn so với các mô hình Bộ học cơ sở, đặc biệt ở các lớp 2 và 3.

Nhìn chung, mô hình Stacked Ensemble mang lại kết quả tốt nhất về độ chính xác và giảm thiểu nhầm lẫn giữa các lớp.

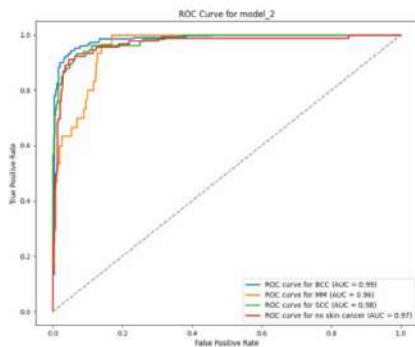
Hình 3.18 thể hiện các biểu đồ ROC cho từng lớp của các mô hình CNN 1, CNN 2, CNN 3 và mô hình Stacked Ensemble. Các biểu đồ ROC này cho thấy sự phân biệt giữa các lớp và hiệu suất của các mô hình trên từng lớp.



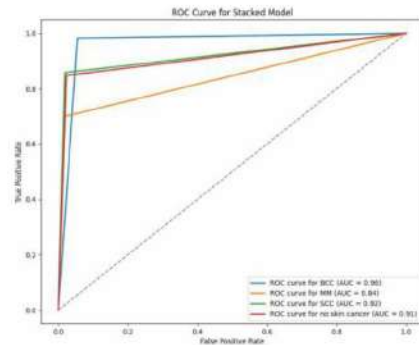
(a) Biểu đồ ROC của CNN 1 cho từng lớp.



(b) Biểu đồ ROC của CNN 2 cho từng lớp.



(c) Biểu đồ ROC của CNN 3 cho từng lớp.



(d) Biểu đồ ROC của mô hình Stacked Ensemble cho từng lớp.

Hình 3.18: Biểu đồ ROC cho từng lớp của các mô hình với batch size 16 và 10 epochs.

- **CNN 1** (Hình 3.18a) có AUC rất cao cho cả bốn lớp, đặc biệt lớp BCC và SCC đạt $AUC = 0.99$, cho thấy khả năng phân loại tốt.

- **CNN 2** (Hình 3.18b) cũng đạt hiệu suất tốt với AUC cao, nhưng lớp MM có $AUC = 0.98$ thấp hơn so với các lớp khác.

- **CNN 3** (Hình 3.18c) cho thấy sự phân biệt rõ ràng giữa các lớp với AUC dao động từ 0.96 đến 0.99, đặc biệt lớp BCC đạt 0.99.

- **Stacked Ensemble** (Hình 3.18d) có AUC tổng thể tốt, nhưng lớp MM có $AUC = 0.84$, thấp hơn so với các mô hình con, tuy nhiên các lớp khác vẫn giữ mức AUC cao từ 0.91 đến 0.96.

Tổng thể, các mô hình đều có khả năng phân loại tốt với giá trị AUC cao, đặc biệt là các lớp BCC và SCC.

3.3.4. Phân tích kết quả mô hình đề xuất và các mô hình cơ sở được đánh giá tốt nhất

Bảng 3.7 trình bày so sánh kết quả thực nghiệm giữa các mô hình cơ sở và mô hình đề xuất **Stacked Ensemble**. Các chỉ số hiệu năng được so sánh bao gồm: kích thước batch (Batch Size), số lượng epoch (Epoch), độ chính xác (Accuracy), độ chính xác của mô hình (Precision), độ nhạy (Recall), điểm F1 (F1 Score), diện tích dưới đường cong ROC (AUC), độ nhạy (Sensitivity), độ đặc hiệu (Specificity), và thời gian huấn luyện (T) tính bằng giây. Mục đích của bảng là đánh giá hiệu năng của từng mô hình và xác định mô hình nào có kết quả tốt nhất trong phân loại.

Bảng 3.7: So sánh hiệu năng giữa các mô hình học sâu cơ sở và mô hình kết hợp *Stacked Ensemble*. Các tiêu chí viết tắt: *BS = Batch Size, Ep = Epoch, Acc = Accuracy, Pr = Precision, Re = Recall, F1 = F1 Score, AUC = Area Under Curve, Sen = Sensitivity, Spec = Specificity, T = Time Taken (in seconds)*. Mô hình có hiệu năng tốt nhất được in đậm.

Model	BS	Ep	Acc	Pr	Re	F1	AUC	Sen	Spec	Train Loss	Val Loss	T
Xception	8	23	0.827	0.826	0.923	0.820	0.923	0.923	0.997	0.569	0.509	669
MobileNetV2	16	13	0.833	0.808	0.810	0.806	0.765	0.765	0.990	0.517	0.506	354
DenseNet121	32	10	0.823	0.800	0.806	0.796	0.778	0.778	0.997	0.501	0.505	276
CNN 0	16	10	0.921	0.920	0.921	0.920	0.984	0.921	0.964	-	-	-
CNN 1	16	10	0.897	0.899	0.897	0.896	0.980	0.897	0.959	-	-	-
CNN 2	16	10	0.881	0.906	0.881	0.869	0.973	0.881	0.958	-	-	-
Stacked Ensemble	16	10	0.927	0.928	0.927	0.927	0.909	0.927	0.971	-	-	-

Dựa trên Bảng 3.7, có thể rút ra một số nhận xét và đánh giá quan trọng về hiệu năng của các mô hình.

3.3.4.1. Độ chính xác (Accuracy)

Các mô hình cơ sở có độ chính xác dao động từ 0.823 (DenseNet121) đến 0.833 (MobileNetV2). Đặc biệt, mô hình DenseNet121 đạt mức độ chính xác cao với giá trị 0.823, nhưng vẫn thấp hơn so với mô hình MobileNetV2.

Mô hình **Stacked Ensemble**, khi kết hợp các mô hình học sâu cơ sở, đã cải thiện

đáng kể độ chính xác lên mức **0.927**, cho thấy sự kết hợp này giúp tăng cường hiệu năng tổng thể của hệ thống phân loại hình ảnh.

3.3.4.2. Độ chính xác Precision, Recall và F1 Score

Về **Precision**, các mô hình cơ sở đạt từ 0.800 (DenseNet121) đến 0.826 (Xception). Trong khi đó, mô hình **Stacked Ensemble** đạt độ chính xác cao nhất với **0.928**, vượt qua tất cả các mô hình cơ sở.

Về **Recall**, mô hình Xception đạt giá trị cao nhất trong các mô hình cơ sở với **0.923**, tuy nhiên mô hình **Stacked Ensemble** cũng đạt được **0.927**, vượt trội hơn so với các mô hình còn lại.

F1 Score của mô hình **Stacked Ensemble** đạt **0.927**, cao hơn tất cả các mô hình khác, cho thấy sự cân bằng giữa Precision và Recall của mô hình kết hợp này.

3.3.4.3. Chỉ số AUC

Mặc dù các mô hình như Xception và DenseNet121 có chỉ số AUC cao (lần lượt là 0.923 và 0.778), mô hình **Stacked Ensemble** chỉ đạt **0.909**. Chỉ số này thấp hơn một chút so với các mô hình cơ sở, có thể là do sự phức tạp khi kết hợp nhiều mô hình.

3.3.4.4. Độ nhạy và độ đặc hiệu (Sensitivity, Specificity)

Về **Sensitivity**, mô hình Xception đạt **0.923**, tuy nhiên mô hình **Stacked Ensemble** đạt giá trị cao hơn với **0.927**, cho thấy khả năng dự đoán chính xác các trường hợp dương tính.

Về **Specificity**, mô hình **Stacked Ensemble** đạt **0.971**, chỉ thấp hơn một chút so với Xception (0.997), cho thấy khả năng phát hiện chính xác các trường hợp âm tính.

3.3.4.5. Tốc độ tính toán

Mô hình **Stacked Ensemble** có thể gặp nhược điểm về thời gian huấn luyện, khi sự phức tạp của việc kết hợp nhiều mô hình có thể khiến quá trình này diễn ra chậm hơn. Tuy nhiên, hiệu năng vượt trội về các chỉ số **Accuracy**, **Precision**, **Recall**, và **F1 Score** là một sự đánh đổi hợp lý.

3.3.4.6. Kết luận

Từ các phân tích trên, có thể khẳng định rằng mô hình **Stacked Ensemble** vượt trội hơn hầu hết các mô hình cơ sở về các chỉ số chính như **Accuracy**, **Precision**, **Recall** và **F1 Score**. Dù chỉ số **AUC** và thời gian huấn luyện của mô hình này có thể cần cải thiện thêm, việc sử dụng mô hình kết hợp này hứa hẹn mang lại hiệu năng cao trong việc phân loại hình ảnh.

Kết luận chương 3

Trong chương này, luận văn đã tiến hành thực nghiệm trên nhiều mô hình mạng nơ-ron tích chập khác nhau nhằm đánh giá hiệu quả của chúng trong việc chẩn đoán ung thư da. Kết quả cho thấy mô hình Stacked Ensemble, với sự kết hợp của các mô hình học sâu mạnh mẽ như DenseNet, MobileNet và Xception, đã đạt được độ chính xác và độ nhạy vượt trội so với các mô hình Bộ học cơ sở. Đặc biệt, khả năng xử lý dữ liệu không đồng đều giữa các lớp của mô hình đã giúp cải thiện hiệu suất tổng thể trong việc nhận diện các loại ung thư da. Những phân tích chi tiết về các chỉ số hiệu năng đã cung cấp cái nhìn sâu sắc, làm cơ sở cho các hướng phát triển tiếp theo trong việc áp dụng trí tuệ nhân tạo vào lĩnh vực chẩn đoán hình ảnh y khoa.

KẾT LUẬN VÀ KIẾN NGHỊ

Kết quả đạt được của luận văn

Luận văn đã đạt được những kết quả quan trọng trong việc nghiên cứu và áp dụng các phương pháp học sâu để chẩn đoán ung thư da từ hình ảnh tổn thương da. Cụ thể, mô hình *Stacked Ensemble* được đề xuất đã chứng minh hiệu quả vượt trội so với các mô hình thành phần khác (CNN 1, CNN 2, CNN 3) qua các chỉ số như độ chính xác, F1-score, AUC, và độ nhạy. Mô hình đề xuất đã giúp cải thiện khả năng phân loại chính xác các loại tổn thương da và cho thấy tính tổng quát cao hơn trên các tập dữ liệu thử nghiệm.

Kết quả thực nghiệm cho thấy với batch size là 16 và số epoch 10, mô hình *Stacked Ensemble* đạt chỉ số sensitivity là 0.915 và specificity là 0.967, cho thấy khả năng phân biệt tốt giữa các lớp trong tập dữ liệu.

Nghiên cứu tương lai

Trong tương lai, nghiên cứu có thể mở rộng theo các hướng sau:

- Tăng cường dữ liệu bằng các kỹ thuật tiên tiến như *Generative Adversarial Networks* (GAN) để cải thiện chất lượng và tính đa dạng của dữ liệu huấn luyện, từ đó tăng cường hiệu suất của mô hình.
- Áp dụng các phương pháp học sâu khác như *Transformers* hoặc các mô hình lai để khám phá khả năng cải thiện trong việc phân loại tổn thương da từ hình ảnh.
- Thử nghiệm các mô hình trên các tập dữ liệu lớn hơn và đa dạng hơn, chẳng hạn như ISIC2019, để đánh giá khả năng tổng quát hóa và ứng dụng thực tế của mô hình.
- Phát triển các phương pháp giải thích mô hình (*model interpretability*) để giúp các chuyên gia y tế hiểu rõ hơn về quyết định của mô hình, tăng cường tính minh bạch và tin cậy trong ứng dụng lâm sàng.

TÀI LIỆU THAM KHẢO

- [1] Reiner Hanewinkel, Jana Janssen, Inga-Marie Hübner, Eckhard Wilhelm Breitbart, and Barbara Isensee. “[Shaded places for skin cancer prevention in kindergartens and schools].” In: *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* (2022).
- [2] Pham Hoang Anh and Collaborators. “Cancer in Hanoi 1991-1992”. In: *Vietnamese Medicine; Cancer Special Issue* 173.7 (1993), pp. 14–21.
- [3] *Tình hình ung thư tại Việt Nam*. https://moh.gov.vn/hoat-dong-cua-dia-phuong/-/asset_publisher/gHbla8vOQDuS/content/tinh-hinh-ung-thu-tai-viet-nam. Accessed on [23/7/2023].
- [4] Jwan Najeeb Saeed and Subhi R. M. Zeebaree. “Skin Lesion Classification Based on Deep Convolutional Neural Networks Architectures”. In: *Journal of Applied Science and Technology Trends* (2021).
- [5] Monika Janda and H. Peter Soyer. “Using Advances in Skin Imaging Technology and Genomics for the Early Detection and Prevention of Melanoma”. In: *Dermatology* 235 (2018), pp. 1–3.
- [6] Nazneen N. Sultana and Niladri Bihari Puhan. “Recent Deep Learning Methods for Melanoma Detection: A Review”. In: *International Conference on Mathematics and Computing*. 2018.
- [7] Min Chen, Ping Zhou, Di Wu, Long Hu, Mohammad Mehedi Hassan, and Atif Alamri. “AI-Skin : Skin Disease Recognition based on Self-learning and Wide Data Collection through a Closed Loop Framework”. In: *ArXiv abs/1906.01895* (2019).
- [8] Firas A. Al-qarqaz, Khaldon Bodoor, Awad Hasan Al-Tarawneh, Haytham Eloqayli, Wisam Al Gargaz, Diala M. Alshiyab, Jihan M. Muhaidat, Mohammad Ali Alqudah, Rowida Almomani, and Maha Marji. “Basal Cell Carcinoma Pathology Requests and Reports Are Lacking Important Information”. In: *Journal of Skin Cancer* 2019 (2019)
- [9] Concetta Potenza, Nicoletta Bernardini, Veronica Balduzzi, Luigi Losco, Alessandra Mambrin, Anna Marchesiello, Ersilia Tolino, Sara Zuber, Nevena Skroza, and Ilaria Proietti. “A Review of the Literature of Surgical

and Non- surgical Treatments of Invasive Squamous Cells Carcinoma”. In: *BioMed Research International* 2018 (2018)

- [10] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries”. In: *CA: A Cancer Journal for Clinicians* 71 (2021), pp. 209–249.
- [11] Giuseppe Argenziano, Gabriella Fabbrocini, Paolo Carli, Vincenzo de Giorgi, Elena Sammarco, and Mario Delfino. “Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis.” In: *Archives of dermatology* 134 12 (1998), pp. 1563–70.
- [12] Vivek Kumar Singh, Mohamed Abdel-Nasser, Hatem A. Rashwan, Farhan Akram, Nidhi Pandey, Alain Lalande, Benoît Presles, Santiago Romani, and Domènec Puig. “FCA-Net: Adversarial Learning for Skin Lesion Segmentation Based on Multi-Scale Features and Factorized Channel Attention”. In: *IEEE Access* 7 (2019), pp. 130552–130565.
- [13] Alfred W. Kopf, Thomas G Salopek, Johnny Slade, Ashfaq A. Marghoob, and Robert S. Bart. “Techniques of cutaneous examination for the detection of skin cancer”. In: *Cancer* 75 (1995).
- [14] Catarina Barata, M. E. Celebi, and Jorge S. Marques. “A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer”. In: *IEEE Journal of Biomedical and Health Informatics* 23 (2019), pp. 1096–1109.
- [15] Farhat Afza, Muhammad Sharif, Muhammad Attique Khan, Usman Tariq, Hwan-Seung Yong, and Jaehyuk Cha. “Multiclass Skin Lesion Classification Using Hybrid Deep Features Selection and Extreme Learning Machine”. In: *Sensors (Basel, Switzerland)* 22 (2022).
- [16] Jeremy Kawahara, Sara Daneshvar, Giuseppe Argenziano, and G. Hamarneh. “Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets”. In: *IEEE Journal of Biomedical and Health Informatics* 23 (2019), pp. 538–546.

- [17] M VijayalakshmiM. “Melanoma Skin Cancer Detection using Image Processing and Machine Learning”. In: *International Journal of Trend in Scientific Research and Development* (2019).
- [18] Arslan Javaid, Muhammad Sadiq, and Faraz Akram. “Skin Cancer Classification Using Image Processing and Machine Learning”. In: *2021 International Bhurban Conference on Applied Sciences and Technologies (IB-CAST)* (2021), pp. 439–444
- [19] Oscar P. Yanchatuña, Jackeline P. Pereira, Kevin O. Pila, Paola A. Vásquez, Keiko S. Veintimilla, G. F. Villalba-Meneses, O. Alvarado-Cando, and Diego A Almeida-Galárraga. “Skin Lesion Detection and Classification Using Convolutional Neural Network for Deep Feature Extraction and Support Vector Machine”. In: *International Journal on Advanced Science, Engineering and Information Technology* (2021).
- [20] Bhuvaneshwari Shetty, Roshan Fernandes, Anisha P. Rodrigues, Rajeswari Chengoden, Sweta Bhattacharya, and Kuruva Lakshmana. “Skin lesion classification of dermoscopic images using machine learning and convolutional neural network”. In: *Scientific Reports* 12 (2022).
- [21] Muhammad Attique Khan, Tallha Akram, Yudong Zhang, and Muhammad Sharif. “Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework”. In: *Pattern Recognit. Lett.* 143 (2021), pp. 58–66.
- [22] Juan Pablo Villa-Pulgarin, A. A. Ruales-Torres, Daniel Arias-Garz, Mario Alejandro Bravo-Ortiz, Harold Brayan Arteaga-Arteaga, Alejandro Mora-Rubio, Jesús Alejandro Alzate-Grisales, Esteban Mercado-Ruiz, M. Hassaballah, Simon Orozco-Arias, Oscar Cardona-Morales, and Reinel Tabares-Soto. “Optimized Convolutional Neural Network Models for Skin Lesion Classification”. In: *Computers, Materials & Continua* (2022).
- [23] Devansh Bisla, Anna Choromańska, Russell S. Berman, Jennifer A. Stein, and David Polsky. “Towards Automated Melanoma Detection With Deep Learning: Data Purification and Augmentation”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2019), pp. 2720–2728. URL: <https://api.semanticscholar.org/>

- [24] Mehak Arshad, Muhammad Attique Khan, Usman Tariq, Ammar Armghan, Fayadh S. Alenezi, Muhammad Younus Javed, Shabnam Mohamed Aslam, and Seifedine Nimer Kadry. “A Computer-Aided Diagnosis System Using Deep Learning for Multiclass Skin Lesion Classification”. In: *Computational Intelligence and Neuroscience 2021* (2021). URL: <https://api.semanticscholar.org/CorpusID:245011192>.
- [25] Alceu Bissoto, Eduardo Valle, and Sandra Avila. “GAN-Based Data Augmentation and Anonymization for Skin-Lesion Analysis: A Critical Review”. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2021), pp. 1847–1856.
- [26] Mohammad Shorfuzzaman. “An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection”. In: *Multimedia Systems* 28 (2021), pp. 1309–1323.
- [27] Dan Popescu, Mohamed El-khatib, and Loretta Ichim. “Skin Lesion Classification Using Collective Intelligence of Multiple Neural Networks”. In: *Sensors (Basel, Switzerland)* 22 (2022).
- [28] Jiaqi Ding, Jie Song, Jiawei Li, Jijun Tang, and Fei Guo. “Two-Stage Deep Neural Network via Ensemble Learning for Melanoma Classification”. In: *Frontiers in Bioengineering and Biotechnology* 9 (2022).
- [29] Khurram Shehzad, Tan Zhenhua, Shifa Shoukat, Adnan Saeed, Ijaz Ahmad, Shahzad Sarwar Bhatti, and Samia Allaoua Chelloug. “A Deep-Ensemble-Learning-Based Approach for Skin Cancer Diagnosis”. In: *Electronics* (2023).
- [30] Manjunath Jogin, Divya G D, Mohana, Meghana RK, Madhulika MS, and Apoorva S. “Feature extraction using convolution neural networks (CNN) and deep learning”. In: 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTE-ICT). IEEE. 2018, pp. 2319–2323.
- [31] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

- [32] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. “Going deeper with convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1–9.
- [33] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.
- [34] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *International Conference on Learning Representations*. 2020.
- [35] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [36] Mingxing Tan and Quoc V Le. “EfficientNet: Rethinking model scaling for convolutional neural networks”. In: *International Conference on Machine Learning*. 2019, pp. 6105–6114.
- [37] Francois Chollet. “Xception: Deep learning with depthwise separable convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1251–1258.
- [38] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. “Densely Connected Convolutional Networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 4700–4708.
- [39] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. “A survey on deep transfer learning”. In: *Proceedings of the international conference on artificial neural networks (ICANN)*. Springer, 2018.
- [40] S. H. Kassani and P. H. Kassani. “A comparative study of deep learning models for melanoma detection”. In: *IEEE International Conference on Healthcare Informatics (2019)*.
- [41] O. Miksik and K. Mikolajczyk. “Evaluation of local detectors and descriptors for fast feature matching”. In: *Proceedings of the British Machine Vision Conference (BMVC)*. 2012.

- [42] I. Guyon and A. Elisseeff. Introduction to feature extraction. Springer, 2006.
- [43] D. Storcheus, H. Shen, and Z. Sheng. “A survey on feature extraction methods”. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2015).
- [44] C. Barata, M. E. Celebi, and J. S. Marques. “Two decades of research on dermoscopy image analysis: a survey”. In: IEEE Transactions on Biomedical Engineering (2019).
- [45] Q. Abbas and M. E. Celebi. “A survey on deep learning methods for skin lesion classification”. In: *IEEE Access* (2019).
- [46] Vladimir N Vapnik. Statistical learning theory. John Wiley & Sons, 1998.
- [47] J. Ross Quinlan. “Induction of decision trees”. In: Machine Learning. Springer, 1986, pp. 81–106.
- [48] Richard E. Neapolitan. Learning Bayesian Networks. Prentice Hall, 2004.
- [49] Leo Breiman. “Random forests”. In: Machine Learning 45.1 (2001), pp. 5-32.
- [50] Tianqi Chen and Carlos Guestrin. “XGBoost: A scalable tree boosting system”. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016), pp. 785–794.
- [51] Margherita Grandini, Enrico Bagli, and Giorgio Visani. “Metrics for multi-class classification: an overview”. In: arXiv preprint arXiv:2008.05756 (2020).
- [52] Rafael C Gonzalez. *Digital image processing*. Pearson education india, 2009.

Số: M2 /QĐ-HVKHCN

Hà Nội, ngày 30 tháng 09 năm 2024

QUYẾT ĐỊNH
Về việc thành lập Hội đồng đánh giá luận văn thạc sĩ

GIÁM ĐỐC
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Căn cứ Quyết định số 303/QĐ-VHL ngày 01/3/2023 của Chủ tịch Viện Hàn lâm Khoa học và Công nghệ Việt Nam về việc ban hành Quy chế Tổ chức và hoạt động của Học viện Khoa học và Công nghệ;

Căn cứ Thông tư số 23/2021/TT-BGDĐT ngày 30/08/2021 của Bộ trưởng Bộ Giáo dục và Đào tạo về việc ban hành Quy chế đào tạo trình độ thạc sĩ;

Căn cứ Quyết định số 1966/QĐ-HVKHCN ngày 28/12/2021 của Giám đốc Học viện Khoa học và Công nghệ về việc ban hành Quy chế đào tạo trình độ thạc sĩ;

Căn cứ Quyết định số 1899/QĐ-HVKHCN ngày 15/11/2022 của Giám đốc Học viện Khoa học và Công nghệ về việc công nhận học viên cao học khóa 2022B - Đợt 2 năm 2022;

Căn cứ Quyết định số 226/QĐ-HVKHCN ngày 29/03/2024 của Giám đốc Học viện Khoa học và Công nghệ về việc công nhận đề tài và người hướng dẫn luận văn thạc sĩ;

Xét đề nghị của Trưởng khoa Khoa Công nghệ thông tin và Viễn thông, Trưởng phòng Đào tạo.

QUYẾT ĐỊNH:

Điều 1. Thành lập Hội đồng đánh giá luận văn thạc sĩ cho học viên Đỗ Minh Tuấn với đề tài: **Nghiên cứu kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam**

Ngành: Hệ thống thông tin

Mã số: 8 48 01 04

Danh sách thành viên Hội đồng đánh giá luận văn kèm theo Quyết định này.

Điều 2. Hội đồng có trách nhiệm đánh giá luận văn thạc sĩ theo đúng quy chế hiện hành của Bộ Giáo dục và Đào tạo, Học viện Khoa học và Công nghệ. Quyết định có hiệu lực tối đa 60 ngày kể từ ngày ký và phải đảm bảo thời hạn đào tạo theo quy định của Học viện. Hội đồng tự giải thể sau khi hoàn thành nhiệm vụ.

Điều 3. Trưởng phòng Tổ chức - Hành chính và Truyền thông, Trưởng phòng Đào tạo, Trưởng phòng Kế toán, Trưởng Khoa Công nghệ thông tin và Viễn thông, các thành viên có tên trong danh sách Hội đồng và học viên cao học có tên tại Điều 1 chịu trách nhiệm thi hành Quyết định này. *Ph*

Nơi nhận:

- Như Điều 3;
- Giám đốc HV (để b/c);
- Lưu hồ sơ học viên;
- Lưu: VT, ĐT, PQ.10.

KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC

TS. Trần Thị Phương Anh



DANH SÁCH HỘI ĐỒNG ĐÁNH GIÁ LUẬN VĂN THẠC SĨ

(Kèm theo Quyết định số 442/QĐ-HVKHCN ngày 30/09/2024
của Giám đốc Học viện Khoa học và Công nghệ)

Cho luận văn của học viên: Đỗ Minh Tuấn

Tên đề tài: Nghiên cứu kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam

Ngành: Hệ thống thông tin

Mã số: 8 48 01 04

Người hướng dẫn : TS. Vũ Văn Hiệu, Viện Công nghệ thông tin,
Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TT	Họ và tên, học hàm, học vị	Ngành	Cơ quan công tác	Trách nhiệm trong Hội đồng
1.	PGS.TS. Nguyễn Long Giang	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Chủ tịch
2.	PGS.TS. Bùi Thu Lâm	Hệ thống thông tin	Học viện Kỹ thuật Mật mã, Ban Cơ yếu Chính phủ	Phản biện 1
3.	TS. Nguyễn Mạnh Hùng	Hệ thống thông tin	Học viện Kỹ thuật Quân sự, Bộ Quốc phòng	Phản biện 2
4.	TS. Trần Đức Nghĩa	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Ủy viên - Thư ký
5.	PGS.TS. Nguyễn Việt Anh	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Ủy viên

Hội đồng gồm 05 thành viên./ *JM*

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT LUẬN VĂN THẠC SĨ

Họ và tên người nhận xét: Nguyễn Long Giang Học hàm, học vị: PGS.TS.
Chức danh trong Hội đồng: Chủ tịch Hội đồng
Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN
Họ và tên học viên: Đỗ Minh Tuấn
Tên đề tài: Nghiên cứu kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng
chẩn đoán một số bệnh ung thư da ở Việt Nam
Ngành: Hệ thống thông tin Mã số: 8 48 01 04

NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, tính thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận văn:

..... Nghiên cứu... phát hiện các mô hình... kết hợp CNN
đến... phân lớp... các bệnh ung thư da... nhận... từ... công việc
... từ... đến... nghiên cứu... ý... nghĩa... khoa học... thực tiễn

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình khoa học, luận văn đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo:

..... Đề tài không trùng lặp với các công trình khoa học,
luận văn đã công bố trong và ngoài nước
..... Tài liệu trích dẫn đầy đủ, rõ ràng và trung thực

3. Sự phù hợp giữa tên đề tài với nội dung nghiên cứu cũng như với chuyên ngành và mã số đào tạo:

..... Tên đề tài phù hợp với nội dung nghiên cứu và phù
hợp với chuyên ngành, mã số đào tạo

4. Độ tin cậy và tính hiện đại của phương pháp nghiên cứu đã sử dụng để hoàn thành luận văn:

..... Phương pháp nghiên cứu... kỹ thuật... về các
mô hình mạng kết hợp trên CNN đã

giải quyết bài toán phân lớp limb với ứng dụng
thực nghiệm, đánh giá trên các bộ dữ liệu thực tế
tư liệu dài, bài đánh giá dựa trên ứng dụng.....

5. Kết quả nghiên cứu của luận văn:

..... Tìm kiếm và thực hiện các mô hình CNN
liên tiếp, từ đó đánh giá ứng dụng dựa trên các ứng dụng mô
lập giải quyết bài toán thực nghiệm, đánh giá trên bộ
dữ liệu mẫu

6. Những hạn chế, thiếu sót của luận văn về nội dung, hình thức và các câu hỏi

..... Bổ sung mô hình kết hợp giữa mô hình (Stack)
cơ bản và ứng dụng của nó về kết hợp
..... Phân tích chi tiết hơn kết quả thực nghiệm, đánh
giá các mô hình, mô hình kết hợp
..... Phân tích đánh giá mô hình cần vẽ tranh bày, thêm một
bảng đánh giá trên tham khảo để bổ sung mục thêm
phần

7. Nếu tác giả chưa viết bài báo khoa học thì nội dung của luận văn có thể được viết thành các bài báo để gửi đăng trên tạp chí khoa học, sách chuyên ngành hoặc tuyển tập công trình hội nghị khoa học cấp quốc gia, quốc tế hay không?

..... Có thể viết báo khoa học trong lĩnh vực CNTT
và y học (chuyên ngành da liễu)

8. Kết luận chung (khẳng định mức độ đáp ứng các yêu cầu đối với một luận văn Thạc sĩ; luận văn có thể đưa ra bảo vệ để nhận học vị Thạc sĩ được hay không?):

..... Luận văn đáp ứng các yêu cầu đối với luận văn Th.S
..... Có thể đưa luận văn ra bảo vệ để nhận học vị
Th.S

Hà Nội, ngày 07 tháng 10 năm 2024

Người nhận xét
(Ký, ghi rõ họ tên)

PGS.TS. Nguyễn Long Giang

BẢN NHẬN XÉT LUẬN VĂN THẠC SĨ

Họ và tên người nhận xét: Bùi Thu Lâm. Học hàm, học vị: PGS.TS.

Chức danh trong Hội đồng: Phản biện 1.

Cơ quan công tác: Học viện Kỹ thuật Mật mã, Ban cơ yếu Chính phủ.

Họ và tên học viên: Đỗ Minh Tuấn.

Tên đề tài: Nghiên cứu một số kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam.

Ngành: Hệ thống thông tin.

Mã số: 8480104.

NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, thời sự và ý nghĩa khoa học:

Ung thư da, đặc biệt là Melanoma, là một trong những loại ung thư phổ biến và nguy hiểm. Việc chẩn đoán sớm có thể cải thiện cơ hội điều trị thành công, nhưng các phương pháp hiện tại đòi hỏi kinh nghiệm của bác sĩ chuyên khoa. Ứng dụng học sâu và mạng nơ-ron tích chập (CNN) vào việc hỗ trợ chẩn đoán tự động là một giải pháp rất có tiềm năng.

Nghiên cứu sử dụng các mô hình học sâu hiện đại để chẩn đoán ung thư da ở Việt Nam, nơi mà việc nghiên cứu về vấn đề này chưa được khai thác nhiều, mang lại ý nghĩa khoa học quan trọng.

2. Tính không trùng lặp, trung thực và đầy đủ trong trích dẫn:

Đề tài không trùng lặp với các nghiên cứu khác, đặc biệt trong bối cảnh Việt Nam chưa có nhiều nghiên cứu về chẩn đoán ung thư da bằng học sâu. Các trích dẫn tài liệu khoa học được trình bày phù hợp.

3. Phù hợp giữa tên đề tài, nội dung và chuyên ngành:

Tên đề tài và nội dung hoàn toàn phù hợp với chuyên ngành Hệ thống thông tin, mã số 8480104. Đề tài kết hợp giữa học sâu và ứng dụng thực tiễn trong y học, đúng với nội dung nghiên cứu của luận văn.

4. Phương pháp nghiên cứu và độ tin cậy:

Phương pháp nghiên cứu sử dụng các mạng CNN tiên tiến như GoogleNet, VGG16, ResNet, EfficientNet để phân loại các loại ung thư da từ hình ảnh y khoa. Kỹ thuật học chuyển giao (transfer learning) và mô hình tổ hợp (Stacked Ensemble) cũng được áp dụng, đảm bảo tính hiện đại và độ tin cậy.

5. Kết quả nghiên cứu:

Kết quả thực nghiệm trên tập dữ liệu thực tế từ bệnh viện cho thấy hiệu suất cao trong việc phát hiện và phân loại ung thư da, với các chỉ số như Accuracy, Precision, Recall, và AUC đạt mức tốt. Mô hình đề xuất giúp cải thiện khả năng tổng quát hóa và độ chính xác trong chẩn đoán.

6. Hạn chế, thiếu sót:

Hạn chế chính là chưa triển khai thực tế vào hệ thống y tế tại các bệnh viện để kiểm chứng hiệu quả mô hình trong môi trường thực tế.

Các mô hình học máy đơn giản thì không cần mô tả như trong mục 2.6

Làm rõ tại sao mô hình tổ hợp lại có kết quả tốt hơn.

7. Khả năng viết bài báo khoa học:

Nội dung hoàn toàn có thể viết thành bài báo khoa học về ứng dụng học sâu trong chẩn đoán ung thư da, có thể gửi đăng trên các tạp chí chuyên ngành y tế hoặc hệ thống thông tin.

8. Kết luận chung:

Luận văn đáp ứng yêu cầu về một nghiên cứu thạc sĩ. Luận văn đủ điều kiện bảo vệ

H.Đ.Ư.Đ., ngày 10 tháng 10 năm 2024

Người nhận xét
(Ký, ghi rõ họ tên)



Lưu ý:

- Nhận xét được làm thành 02 bản, có chữ ký của người nhận xét và gửi về phòng Đào tạo 02 ngày trước buổi bảo vệ.
- Địa chỉ liên hệ: CV. Phạm Thị Như Quỳnh phòng Đào tạo, Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội. ĐT02438689977- 0916467768

BẢN NHẬN XÉT LUẬN VĂN THẠC SĨ

Họ và tên người nhận xét: Nguyễn Mạnh Hùng. Học hàm, học vị: TS.

Chức danh trong Hội đồng: Phản biện 2.

Cơ quan công tác: Học viện Kỹ thuật Quân sự, Bộ Quốc phòng.

Họ và tên học viên: Đỗ Minh Tuấn.

Tên đề tài: Nghiên cứu một số kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam.

Ngành: Hệ thống thông tin.

Mã số: 8480104.

NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, tính thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận văn

Đề tài "Nghiên cứu kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam" có ý nghĩa khoa học và thực tiễn. Cụ thể: luận văn cung cấp cái nhìn cụ thể về ứng dụng của mạng nơ-ron tích chập trong chẩn đoán ung thư da, hướng tới đề xuất một mô hình hiệu quả cho bài toán này. Kết quả của luận văn có thể áp dụng để nâng cao kết quả chẩn đoán bệnh ung thư da ở Việt Nam.

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình khoa học, luận văn đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo

Đề tài và nội dung của luận văn không trùng lặp với các công trình khoa học, luận văn đã công bố ở trong và ngoài nước; Trích dẫn trung thực, rõ ràng và đầy đủ đến các tài liệu tham khảo.

3. Sự phù hợp giữa tên đề tài với nội dung nghiên cứu cũng như với chuyên ngành và mã số đào tạo

Tên đề tài phù hợp với nội dung nghiên cứu cũng như với chuyên ngành và mã số đào tạo.

4. Độ tin cậy và tính hiện đại của phương pháp nghiên cứu đã sử dụng để hoàn thành luận văn

Phương pháp nghiên cứu đảm bảo hiện đại và tin cậy. Cụ thể: luận văn đã nghiên cứu các phương pháp, mô hình học sâu, học kết hợp; thu thập bộ dữ liệu thực nghiệm VNCancer đảm bảo tin cậy.

Các mô hình được so sánh với nhau, đảm bảo kết quả thu được là khách quan và đáng tin cậy

5. Kết quả nghiên cứu của luận văn:

- Đề xuất mô hình Stacked Ensemble kết hợp các mạng CNN thành phần;

- Thực nghiệm đánh giá mô hình đề xuất. Mô hình đề xuất đã giúp cải thiện khả năng phân loại chính xác các loại tổn thương da và cho thấy tính tổng quát cao hơn trên các tập dữ liệu thử nghiệm.

6. Những hạn chế, thiếu sót của luận văn về nội dung, hình thức và câu hỏi:

Phát biểu bài toán đang đề ở mục 2.7.1 ở chương 2 phải đưa lên chương 1. Mục các nghiên cứu liên quan ở chương 1, cần bổ sung làm rõ hạn chế của những nghiên cứu này làm cơ sở cho việc đề xuất mô hình ở chương 2.

Cần chỉnh sửa lại chương 2, loại bỏ các nội dung không liên quan trực tiếp đến luận văn. Ví dụ như: bỏ mục 2.4 Học chuyển giao; mục 2.6 cần bỏ đi 4 mục con, chỉ giữ lại mục 2.6.5 và 2.6.6;

Cần nhất quán cách trình bày: (CNN 1, CNN 2, CNN 3) hay (CNN 0, CNN 1, CNN 2); Làm rõ các thành phần trong mô hình đề xuất (Hình 2.4: Kiến trúc mô hình mạng học sâu kết hợp, trang 37).

Đưa "3.1. Các chỉ số đánh giá hiệu năng cho phân loại đa lớp" lên chương 2.

Kết luận của các chương, kết luận của luận văn cần chỉnh sửa lại: đảm bảo ngắn gọn và làm rõ được các kết quả chính của chương, của luận văn.

Luận văn còn nhiều lỗi câu chữ, lỗi soạn thảo, đặc biệt là các công thức.

7. Nếu tác giả chưa viết bài báo khoa học thì nội dung của luận văn có thể được viết thành các bài báo để gửi đăng trên tạp chí khoa học, sách chuyên ngành hoặc tuyển tập công trình hội nghị khoa học cấp quốc gia, quốc tế hay không?

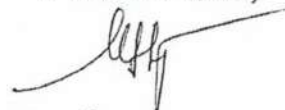
..... luận văn có thể hoàn thiện, viết thành bài báo.
.....
.....

8. Kết luận chung (khẳng định mức độ đáp ứng các yêu cầu đối với một luận văn Thạc sĩ; luận văn có thể đưa ra bảo vệ để nhận học vị Thạc sĩ được hay không?):

Nội dung của luận văn đáp ứng các yêu cầu đối với một luận văn Thạc sĩ; Đề nghị cơ sở đào tạo cho phép học viện được bảo vệ luận văn trước Hội đồng để nhận học vị Thạc sĩ.

Hà Nội, ngày 8 tháng 10 năm 2024

Người nhận xét
(Ký, ghi rõ họ tên)



Nguyễn Mạnh Hùng

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT LUẬN VĂN THẠC SĨ

Họ và tên người nhận xét: Nguyễn Việt Anh. Học hàm, học vị: PGS.TS.

Chức danh trong Hội đồng: Ủy viên.

Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN.

Họ và tên học viên: Đỗ Minh Tuấn.

Tên đề tài: Nghiên cứu một số kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam.

Ngành: Hệ thống thông tin.

Mã số: 8480104.

NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, tính thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận văn:

Ung thư da là một loại bệnh phổ biến ở Việt Nam và trên thế giới. Để đạt được cấp độ phát triển của đất nước

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình khoa học, luận văn đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo:

Thưa hiểu biết của người đọc, đề tài không có sự trùng lặp với các công bố trước đó

3. Sự phù hợp giữa tên đề tài với nội dung nghiên cứu cũng như với chuyên ngành và mã số đào tạo:

phù hợp

4. Độ tin cậy và tính hiện đại của phương pháp nghiên cứu đã sử dụng để hoàn thành luận văn:

Phương pháp nghiên cứu có tính hiện đại và đảm bảo độ tin cậy

5. Kết quả nghiên cứu của luận văn:

Các kết quả nghiên cứu về hình thức và nội dung của luận văn nghiên cứu

6. Những hạn chế, thiếu sót của luận văn về nội dung, hình thức và câu hỏi:

Nhiệm vụ nghiên cứu chưa được thực hiện đầy đủ, cần chỉnh sửa lại

Chưa làm rõ được tác dụng của kết hợp 3 mô hình CNN để đạt kết quả tốt?

Làm rõ hơn đóng góp cụ thể của học viên

7. Nếu tác giả chưa viết bài báo khoa học thì nội dung của luận văn có thể được viết thành các bài báo để gửi đăng trên tạp chí khoa học, sách chuyên ngành hoặc tuyển tập công trình hội nghị khoa học cấp quốc gia, quốc tế hay không?

Không

8. Kết luận chung (khẳng định mức độ đáp ứng các yêu cầu đối với một luận văn Thạc sĩ; luận văn có thể đưa ra bảo vệ để nhận học vị Thạc sĩ được hay không?):

Luận văn đáp ứng các yêu cầu đối với một luận văn Thạc sĩ, có thể đưa ra bảo vệ

....., ngày tháng năm 20...

Người nhận xét
(Ký, ghi rõ họ tên)

Nguyễn Văn Anh

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT LUẬN VĂN THẠC SĨ

Họ và tên người nhận xét: Trần Đức Nghĩa. Học hàm, học vị: TS.

Chức danh trong Hội đồng: Thư ký – Ủy viên.

Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN.

Họ và tên học viên: Đỗ Minh Tuấn.

Tên đề tài: Nghiên cứu một số kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam.

Ngành: Hệ thống thông tin.

Mã số: 8480104.

NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, tính thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận văn:

Cấp thiết, thời sự, ý nghĩa KHT

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình khoa học, luận văn đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo:

Không trùng lặp

3. Sự phù hợp giữa tên đề tài với nội dung nghiên cứu cũng như với chuyên ngành và mã số đào tạo:

Phù hợp

4. Độ tin cậy và tính hiện đại của phương pháp nghiên cứu đã sử dụng để hoàn thành luận văn:

Tin cậy

5. Kết quả nghiên cứu của luận văn:

Cần đặt đề mục hình và phần kết thúc rõ ràng.

6. Những hạn chế, thiếu sót của luận văn về nội dung, hình thức và câu hỏi:

Ra lời hành văn, đúng là.
Lời nói đúng góp của học viên.

7. Nếu tác giả chưa viết bài báo khoa học thì nội dung của luận văn có thể được viết thành các bài báo để gửi đăng trên tạp chí khoa học, sách chuyên ngành hoặc tuyển tập công trình hội nghị khoa học cấp quốc gia, quốc tế hay không?

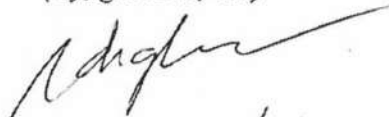
Không.

8. Kết luận chung (khẳng định mức độ đáp ứng các yêu cầu đối với một luận văn Thạc sĩ; luận văn có thể đưa ra bảo vệ để nhận học vị Thạc sĩ được hay không?):

Đáp ứng.
Có thể đưa ra.

....., ngày tháng năm 20...

Người nhận xét
(Ký, ghi rõ họ tên)


Trần Đức Nghĩa

Hà Nội, ngày 30 tháng 10 năm 2024

BIÊN BẢN HỌP HỘI ĐỒNG ĐÁNH GIÁ LUẬN VĂN THẠC SĨ

Thực hiện Quyết định số 1112/QĐ-HVKHCN ngày 30/09/2024 của Giám đốc Học viện Khoa học và Công nghệ về việc thành lập Hội đồng đánh giá luận văn thạc sĩ của học viên Đỗ Minh Tuấn.

Tên đề tài: **Nghiên cứu kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam**

Ngành: Hệ thống thông tin

Mã số: 8 48 01 04

Hôm nay, ngày 30/10/2024 Hội đồng đã họp tại Phòng 1510, Học viện Khoa học và Công nghệ vào lúc 14h00, Hội đồng gồm 05 thành viên:

- | | |
|------------------------------|-------------------|
| 1. PGS.TS. Nguyễn Long Giang | Chủ tịch hội đồng |
| 2. TS. Trần Đức Nghĩa | Thư ký - Ủy viên |
| 3. PGS.TS. Bùi Thu Lâm | Phản biện 1 |
| 4. TS. Nguyễn Mạnh Hùng | Phản biện 2 |
| 5. PGS.TS. Nguyễn Việt Anh | Ủy viên |

Thành viên vắng mặt: ... *không* (Phản biện hoặc ủy viên, đã có bản nhận xét đồng ý cho phép học viên được bảo vệ trước Hội đồng đánh giá luận văn thạc sĩ).

NỘI DUNG LÀM VIỆC

- Đại diện cơ sở đào tạo đọc quyết định thành lập Hội đồng đánh giá luận văn
- Chủ tịch Hội đồng, điều khiển phiên họp
- Thư ký HĐ, đọc lí lịch khoa học và bảng điểm của học viên
- Học viên trình bày luận văn trước Hội đồng
- Phản biện 1:

... *Tôi sao mà hội đồng họp 5.1.1. h.1.1.?*

... *Phản biện 2*

.....

- Phản biện 2:

..... Các' chốt sẽ làm về thời' nghiên' và' kết' quả' ?
..... Còn' anh' em' lớn' nhất' là' ? ... sau' thời' , cũng' thời'

7. Học viên trả lời:

..... Tiếp' theo' ... chốt' ... sau' theo' các' góp' ý'

8. Các thành viên HĐ và những người tham dự nêu câu hỏi

..... Anh' chốt' ... T.L.T.K. ... học' viên' có' thời' ... đặc' ... m' ?
..... Đang' học' ?
..... B' ... phân' ... tích' ... k. b.

10. Học viên trả lời

..... Tiếp' theo' ... chốt' ... sau' theo' các' góp' ý'

11. Hội đồng họp kín và cho điểm

- Hội đồng bầu ban kiểm phiếu gồm 3 thành viên:

Trưởng ban: ... P.G.S.T.S. Bùi Văn Lâm

Ủy viên: ... T.S. Nguyễn Mạnh Hùng

Ủy viên: ... T.S. Trần Thị Nguyễn

- Kết quả kiểm phiếu như sau:

Số phiếu phát ra: ... 5

Số phiếu thu về: ... 5

Tổng số điểm:



Điểm trung bình:

Điểm thưởng công trình công bố:.....

Tổng điểm đánh giá luận văn và thưởng công trình công bố:.....

- Kết luận của Hội đồng:

+ Luận văn đạt yêu cầu (đạt/không đạt yêu cầu)

+ Tính không trùng lặp nội dung và tên đề tài với các công trình công bố:

L.V. không trùng lặp.....

11. Chủ tịch Hội đồng, công bố kết quả, yêu cầu học viên chỉnh sửa luận văn với các nội dung sau:

Chỉnh sửa theo các ý kiến của H.Đ.

Lưu ý đúng góp ý của học viên.

Buổi họp đã kết thúc vào 16 giờ 00 phút ngày 30/10/2024.

Hà Nội, ngày 30 tháng 10 năm 2024

THƯ KÝ HỘI ĐỒNG

CHỦ TỊCH HỘI ĐỒNG

TS. Trần Đức Nghĩa

PGS.TS. Nguyễn Long Giang

XÁC NHẬN CỦA HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

**KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC**



Nguyễn Thị Trung

**BẢN GIẢI TRÌNH CHỈNH SỬA LUẬN VĂN
THEO KẾT LUẬN CỦA HỘI ĐỒNG ĐÁNH GIÁ LUẬN VĂN THẠC SĨ**

Họ tên học viên: Đỗ Minh Tuấn

Lớp: ITT2022B

Tên đề tài luận văn: Nghiên cứu kỹ thuật kết hợp các mạng nơ-ron tích chập và ứng dụng chẩn đoán một số bệnh ung thư da ở Việt Nam

Chuyên ngành: Hệ thống thông tin

Mã số: 8480104

Người hướng dẫn khoa học: TS. Vũ Văn Hiệu

Ngày bảo vệ luận văn: 30/10/2024

Căn cứ biên Bản họp Hội đồng đánh giá luận văn thạc sĩ, học viên đã chỉnh sửa luận văn như sau:



STT	Nội dung đề nghị bổ sung, chỉnh sửa	Nội dung đã bổ sung, chỉnh sửa
1	Tại sao mô hình tổng thể lại tốt hơn?	Đã trình bày phân tích mô hình tổng thể tốt hơn các mô hình của các Bộ học cơ sở Các trang chỉnh sửa: 79, 80
2	Chỉnh sửa chương 2	Đã chỉnh sửa chương 2: - Trích rút đặc trưng học sâu trong bài toán phân lớp ảnh - Mô hình đề xuất Các trang chỉnh sửa: 24 - 27, 35 - 44
3	Sửa lỗi chính tả, soạn thảo, công thức	Đã chỉnh sửa lỗi chính tả, soạn thảo và công thức Các trang chỉnh sửa: 33 - 38, 48 - 51

4	Quá nhiều tài liệu tham khảo	Đã giảm bớt những tài liệu tham khảo liên quan đến các khái niệm hoặc nghiên cứu cơ bản Các trang chỉnh sửa: 83 - 88
5	Giải thích rõ hơn về thử nghiệm và kết quả	Đã bổ sung nội dung giải thích rõ về thử nghiệm và kết quả Các trang chỉnh sửa: 52 - 56, 72
6	Bổ sung phân tích kết quả	Đã bổ sung nội dung phân tích kết quả Các trang chỉnh sửa: 74 - 78

Lưu ý: Trong trường hợp Hội đồng yêu cầu xin ý kiến của 02 phản biện sau bảo vệ, học viên cần xin chữ ký của 02 phản biện xác nhận.

Hà Nội, ngày 11 tháng 11 năm 2024

CHỦ TỊCH HỘI ĐỒNG

PGS. TS. Nguyễn Long Giang

NGƯỜI HƯỚNG DẪN

TS. Vũ Văn Hiệu

HỌC VIÊN

Đỗ Minh Tuấn

XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO
KT. GIÁM ĐỐC
CHỖ GIÁM ĐỐC



Nguyễn Thị Trung

