

BỘ GIÁO DỤC  
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC  
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

---



**Nguyễn Tú Anh**

**CHUẨN ĐOÁN BỆNH DA LIỄU QUA HÌNH ẢNH  
SỬ DỤNG MÔ HÌNH CỘNG TÁC CỦA PHÂN ĐOẠN VÀ PHÂN LỚP  
TRÊN BỘ DỮ LIỆU ISIC-2018**

**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



Nguyễn Tú Anh

**CHẨN ĐOÁN BỆNH DA LIỄU QUA HÌNH ẢNH  
SỬ DỤNG MÔ HÌNH CỘNG TÁC CỦA PHÂN ĐOẠN VÀ PHÂN LỚP  
TRÊN BỘ DỮ LIỆU ISIC-2018**

**LUẬN VĂN THẠC SĨ MÁY TÍNH**

**Ngành:** Hệ thống thông tin

**Mã số:** 8 48 01 04

NGƯỜI HƯỚNG DẪN KHOA HỌC:

PGS. TS. Nguyễn Việt Anh



Hà Nội - 2024

# LỜI CAM ĐOAN

Tôi xin cam đoan đề tài nghiên cứu trong luận văn này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Đồng thời, kết quả này chưa từng xuất hiện trong bất cứ một nghiên cứu nào. Các số liệu, kết quả nêu trong luận văn là trung thực nếu sai tôi hoàn chịu trách nhiệm trước pháp luật.

*Hà Nội, ngày 14 tháng 09 năm 2024*



**Nguyễn Tú Anh**

ntanhtm@gmail.com | 0339802255

# LỜI CẢM ƠN

Luận văn được hoàn thành tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Tác giả xin gửi lời cảm ơn đến người hướng dẫn, người giúp đỡ, đơn vị chuyên môn thực hiện luận văn, ban Lãnh đạo, phòng Đào tạo, các phòng chức năng của Học viện Khoa học và Công nghệ để luận văn được hoàn thành.

Tác giả xin chân thành cảm ơn và ghi nhận sự hỗ trợ và chỉ dạy tận tình của thầy PGS.TS. Nguyễn Việt Anh trong quá trình thực hiện luận văn thạc sĩ này. Những lời khuyên và chỉ dẫn từ các thầy đã giúp tác giả vượt qua những khó khăn trong quá trình học tập và phát triển kỹ năng nghiên cứu của mình, những kiến thức và kinh nghiệm của các thầy sẽ luôn là tài sản vô giá cho sự nghiệp nghiên cứu của tác giả trong giai đoạn tiếp theo.

Tác giả xin chân thành cảm ơn Ban lãnh đạo Viện Công nghệ thông tin, Học viện Khoa học và Công nghệ, Bộ phận Quản lý và các Phòng ban chức năng của Viện Công nghệ thông tin và Học viện Khoa học và Công nghệ đã hỗ trợ tác giả trong quá trình nghiên cứu và học tập tại Học viện. Tác giả xin chân thành cảm ơn PGS.TS. Nguyễn Long Giang, đã tạo điều kiện thuận lợi trong quá trình học tập và nghiên cứu của tác giả.

Thêm nữa, tác giả cũng gửi lời cảm ơn về những đóng góp và nhận xét quý báu của các cộng sự, đồng nghiệp và bạn bè trong suốt quá trình làm luận văn.

Cuối cùng, tác giả xin dành những lời cảm ơn tới các thành viên trong gia đình, sự khuyến khích và động viên của gia đình là động lực để tác giả hoàn thành luận văn này.

*Hà Nội, ngày 14 tháng 09 năm 2024*



**Nguyễn Tú Anh**

# MỤC LỤC

<b>Lời cam đoan</b>	<b>i</b>
<b>Lời cảm ơn</b>	<b>ii</b>
<b>Danh mục các ký hiệu, các chữ cái viết tắt</b>	<b>vii</b>
<b>Danh mục các hình vẽ, đồ thị</b>	<b>x</b>
<b>Danh mục các bảng</b>	<b>xi</b>
<b>Mở đầu</b>	<b>1</b>
1 Tính cấp thiết của đề tài . . . . .	1
2 Mục tiêu của Luận văn . . . . .	2
3 Phương pháp nghiên cứu . . . . .	5
4 Cơ sở khoa học và tính thực tiễn của đề tài . . . . .	6
5 Những đóng góp của luận văn . . . . .	7
6 Bố cục luận văn . . . . .	7
<b>1 TỔNG QUAN CÁC PHƯƠNG PHÁP CHẨN ĐOÁN BỆNH DA LIỄU VỚI HỌC SÂU</b>	<b>9</b>
1.1 Các phương pháp chẩn đoán bệnh da liễu . . . . .	9
1.1.1 Phương pháp truyền thống . . . . .	10
1.1.2 Chẩn đoán dựa trên hình ảnh . . . . .	11
1.1.3 Ứng dụng học máy và học sâu trong chẩn đoán da liễu . . . . .	12
1.2 Phân đoạn hình ảnh trong chẩn đoán bệnh da liễu . . . . .	13
1.2.1 Giới thiệu . . . . .	13
1.2.2 Phương pháp phân đoạn truyền thống . . . . .	14
1.2.3 Phương pháp phân đoạn dựa trên học máy . . . . .	14
1.2.4 Phương pháp phân đoạn dựa trên học sâu . . . . .	15
1.2.5 Các phương pháp tối ưu cho bài toán phân đoạn . . . . .	16
1.2.6 Thách thức và hướng phát triển . . . . .	17
1.3 Phân lớp hình ảnh trong chẩn đoán bệnh da liễu . . . . .	18

1.3.1	Giới thiệu . . . . .	18
1.3.2	Phương pháp truyền thống trong phân lớp hình ảnh . . . . .	18
1.3.3	Phân lớp hình ảnh dựa trên học sâu . . . . .	20
1.3.4	Các phương pháp tối ưu cho bài toán phân lớp . . . . .	20
1.4	Tổng quan về U-Net, ResNet và VGG . . . . .	22
1.4.1	U-Net . . . . .	22
1.4.2	ResNet và VGG . . . . .	28
1.5	Các nghiên cứu ứng dụng học sâu trong chẩn đoán bệnh da liễu . . . . .	34
<b>2</b>	<b>PHƯƠNG PHÁP CHẨN ĐOÁN BỆNH DA LIỄU ĐỀ XUẤT</b>	<b>38</b>
2.1	Cấu trúc và các thành phần của mô hình cộng tác phân lớp - phân đoạn . . . . .	38
2.1.1	Phân đoạn vùng tổn thương sử dụng mô hình biến thể của U-Net . . . . .	38
2.1.2	Phân lớp tổn thương sử dụng các mô hình phổ biến của ResNet và VGG . . . . .	41
2.1.3	Cộng tác của mô hình phân lớp và phân đoạn . . . . .	42
2.2	Phương pháp triển khai và tối ưu hóa mô hình cộng tác trong chẩn đoán bệnh da liễu . . . . .	44
2.2.1	Phương pháp huấn luyện mô hình cộng tác . . . . .	44
2.2.2	Các kỹ thuật được sử dụng trong phân đoạn hình ảnh . . . . .	46
2.2.3	Các kỹ thuật được sử dụng trong phân lớp hình ảnh . . . . .	48
2.3	Kết luận . . . . .	49
<b>3</b>	<b>KẾT QUẢ THỰC NGHIỆM</b>	<b>51</b>
3.1	Bộ dữ liệu sử dụng . . . . .	51
3.1.1	Tổng quan dữ liệu . . . . .	51
3.1.2	Thông kê mô tả và nhận xét dữ liệu . . . . .	53
3.2	Thiết lập thực nghiệm . . . . .	55
3.2.1	Tách bộ dữ liệu . . . . .	55
3.2.2	Cấu hình phần cứng và môi trường huấn luyện . . . . .	56
3.2.3	Cấu hình huấn luyện mô hình . . . . .	57
3.3	Kết quả thực nghiệm . . . . .	58
3.3.1	Kết quả phân đoạn hình ảnh . . . . .	58
3.3.2	Kết quả phân lớp hình ảnh . . . . .	62
3.3.3	Kết quả mô hình cộng tác phân đoạn - phân lớp . . . . .	68
3.4	Kết luận . . . . .	70
<b>4</b>	<b>KẾT LUẬN</b>	<b>72</b>
4.1	Kết luận . . . . .	72
4.2	Hướng phát triển . . . . .	73

4.2.1	Thu thập và xây dựng bộ dữ liệu da liễu thực tế tại Việt Nam . . . . .	73
4.2.2	Phát triển hệ thống AI chẩn đoán tự động bệnh lý da liễu . . . . .	73
4.2.3	Đánh giá và thử nghiệm lâm sàng . . . . .	74
4.3	Kết luận chung . . . . .	74
<b>DANH MỤC TÀI LIỆU THAM KHẢO</b>		<b>79</b>
<b>Appendices</b>		<b>1</b>
<b>PHỤ LỤC A Cơ sở lý thuyết</b>		<b>1</b>
A.1	Đại số tuyến tính . . . . .	1
A.1.1	Định nghĩa ma trận . . . . .	1
A.1.2	Phép cộng ma trận . . . . .	1
A.1.3	Phép nhân ma trận . . . . .	1
A.2	Giải tích . . . . .	2
A.2.1	Đạo hàm . . . . .	2
A.3	Xác suất . . . . .	3
A.3.1	Các phép toán xác suất . . . . .	3
A.3.2	Xác suất hậu nghiệm - Bayes . . . . .	4
A.4	Học máy . . . . .	4
A.4.1	Khái niệm . . . . .	4
A.4.2	Các phương pháp học . . . . .	5
A.4.3	Xây dựng sản phẩm ứng dụng học máy . . . . .	6
A.4.4	Các vấn đề trong học máy . . . . .	7
A.4.5	Phương pháp giải quyết . . . . .	8
A.4.6	Phương pháp đánh giá mô hình . . . . .	13
A.4.7	Giới thiệu các chỉ số mAP và LAMR . . . . .	18
A.4.8	Một số hàm tính lỗi phổ biến . . . . .	20
A.4.9	Một số hàm kích hoạt phổ biến . . . . .	21
A.4.10	Một số tham số huấn luyện - Model hyperparameter . . . . .	22
A.4.11	Một vài phương pháp tối ưu . . . . .	23
A.5	Mạng Nơ-ron nhân tạo . . . . .	25
A.5.1	Khái niệm . . . . .	25
A.6	Mạng nơ-ron tích chập . . . . .	28
A.6.1	Tầng tích chập - Convolutional layer . . . . .	28
A.6.2	Tầng tổng hợp - Pooling layer . . . . .	30
A.6.3	Tầng Fully-Connected (Fully-Connected layer) . . . . .	31

<b>PHỤ LỤC B Bộ dữ liệu HAM10000 và ISIC-2019</b>	<b>33</b>
B.1 Tổng quan	33
B.1.1 HAM10000	33
B.1.2 ISIC-2019	35
B.2 Thống kê mô tả và nhận xét dữ liệu	36
B.2.1 ISIC-2019	36
<b>PHỤ LỤC C WN Labeller - Công cụ quản lý dữ liệu hình ảnh và hỗ trợ chẩn đoán bệnh da liễu</b>	<b>43</b>
C.1 Quản lý dữ liệu hình ảnh	43
C.2 Gán nhãn hình ảnh	43
C.3 Tự động nhận diện vùng tổn thương da liễu bằng AI	44
C.4 Hiển thị và phân tích hình ảnh da liễu	46
C.5 Quản trị hệ thống	46
C.6 Kết luận	46



# DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ CÁI VIẾT TẮT

- **AI** - Artificial Intelligence (Trí tuệ nhân tạo)
- **CNN** - Convolutional Neural Network (Mạng nơ-ron tích chập)
- **U-Net** - Mô hình CNN cho phân đoạn hình ảnh y khoa
- **ResNet** - Residual Network (Mạng dư, một loại mạng nơ-ron sâu có kết nối tắt)
- **VGG** - Visual Geometry Group (Tên mô hình CNN VGG-16, VGG-19)
- **ISIC** - International Skin Imaging Collaboration (Tổ chức hợp tác quốc tế về hình ảnh da)
- **HAM10000** - Human Against Machine with 10000 training images (Bộ dữ liệu hình ảnh da liễu với 10.015 ảnh)
- **DPI** - Dots Per Inch (Số điểm ảnh trên mỗi inch, dùng để chỉ độ phân giải ảnh)
- **RGB** - Red Green Blue (Màu đỏ, xanh lá cây, xanh dương, hệ màu phổ biến trong hình ảnh kỹ thuật số)
- **BCC** - Basal Cell Carcinoma (Ung thư biểu mô tế bào đáy)
- **AK** - Actinic Keratoses (Dày sừng quang hóa)
- **MEL** - Melanoma (Ung thư da ác tính)
- **NV** - Nevus (Bớt hắc tố, nốt ruồi)
- **DF** - Dermatofibroma (U xơ da)
- **VASC** - Vascular Lesions (Tổn thương mạch máu)
- **BKL** - Benign Keratosis (Dày sừng lành tính)
- **CSV** - Comma-Separated Values (Tập văn bản phân tách bằng dấu phẩy)

- **GPU** - Graphics Processing Unit (Bộ xử lý đồ họa, thường dùng trong học sâu)
- **TP** - True Positive (Dương tính thực)
- **FP** - False Positive (Dương tính giả)
- **TN** - True Negative (Âm tính thực)
- **FN** - False Negative (Âm tính giả)
- **IoU** - Intersection over Union (Đo lường sự chồng lấn giữa kết quả phân đoạn và mặt nạ thực)
- **ROC** - Receiver Operating Characteristic (Đường cong ROC, biểu diễn độ nhạy và độ đặc hiệu)
- **AUC** - Area Under the Curve (Diện tích dưới đường cong ROC)
- **DNN** - Deep Neural Network (Mạng nơ-ron sâu)
- **CLAHE** - Contrast Limited Adaptive Histogram Equalization (Phương pháp nâng cao tương phản ảnh)
- **ReLU** - Rectified Linear Unit (Hàm kích hoạt trong mạng nơ-ron)
- **Adam** - Adaptive Moment Estimation (Thuật toán tối ưu học sâu)
- **SGD** - Stochastic Gradient Descent (Thuật toán giảm độ dốc ngẫu nhiên)
- **PCA** - Principal Component Analysis (Phân tích thành phần chính)
- **FN** - False Negative (Âm tính giả)

# DANH SÁCH HÌNH VẼ

1.1	Chẩn đoán lâm sàng bệnh da liễu. Nguồn: Medlatex . . . . .	10
1.2	Ảnh mô bệnh học dùng trong chẩn đoán bệnh da liễu trên xét nghiệm. Nguồn: KC-4.0-44/19-25 . . . . .	11
1.3	Ảnh dermoscopy dùng trong chẩn đoán bệnh da liễu. Nguồn: HAM10000 . . . . .	12
1.4	Phân đoạn hình ảnh da liễu . . . . .	13
1.5	Phân lớp hình ảnh bệnh da liễu với CNN. Nguồn: KC-4.0-44/19-25 . . . . .	18
1.6	Sự khác nhau giữa Object detection và Instance Segmentation . . . . .	23
1.7	Object detection, Semantic Segmentation và Instance Segmentation . . . . .	23
1.8	Các ứng dụng của bài toán Image Segmentation . . . . .	24
1.9	Kiến trúc mạng U-Net với hai đường dẫn mã hóa và giải mã (Nguồn: [15]) . . . . .	26
1.10	Transposed convolution (Nguồn: trong hình) . . . . .	26
1.11	Identity block (Nguồn: Wikipedia) . . . . .	30
1.12	ResNet-50 . . . . .	30
1.13	Mô hình phân lớp hình ảnh bệnh da liễu với CNN của <b>Kawahara và cộng sự</b> . . . . .	34
1.14	Mô hình phân lớp hình ảnh bệnh da liễu với CNN của <b>Esteva và cộng sự</b> . . . . .	35
1.15	Mô hình ResNet-152 . . . . .	35
1.16	Mô hình của Zhen và cộng sự . . . . .	36
2.1	Mô hình mạng U-Net sử dụng . . . . .	39
2.2	Phương pháp tổng quan . . . . .	42
2.3	Phương pháp tăng cường dữ liệu cho bài toán phân đoạn hình ảnh . . . . .	47
3.1	Task 1 của cuộc thi ISIC-2018 . . . . .	51
3.2	Một số ví dụ về các bệnh da liễu trong ISIC-2018 . . . . .	54
3.3	Biểu đồ chỉ số mAP của mô hình . . . . .	59
3.4	Trực quan hóa kết quả phân đoạn cho một số ảnh trong tập kiểm thử. . . . .	61
3.5	Ma trận nhầm lẫn của mô hình ResNet-101. . . . .	66
A.1	Quy trình máy học . . . . .	5
A.2	Phương pháp máy học . . . . .	6
A.3	Các bước để có một sản phẩm ứng dụng học máy . . . . .	7

A.4	Early stopping	8
A.5	K-Fold	9
A.6	Data augmentation	10
A.7	Drop-out	11
A.8	Các metrics đánh giá mô hình	13
A.9	Ví dụ về ma trận nhầm lẫn	14
A.10	Đường cong ROC	16
A.11	AUC	17
A.12	Các dự đoán được xếp hạng theo thứ tự tăng dần của điểm hồi quy logistic	17
A.13	Cách tính Dice Index	18
A.14	Hàm Sigmoid	21
A.15	Nơ-ron thần kinh	26
A.16	Nơ-ron thần kinh nhân tạo	26
A.17	Mạng lưới thần kinh nhân tạo	27
A.18	Convolutional Neural Network-CNN	28
A.19	Convolutional layer	29
A.20	Tích chập khối	30
A.21	Pooling layer	31
A.22	Tầng Fully-Connected	32
B.1	Bộ dữ liệu HAM10000	33
B.2	Một số ví dụ về các bệnh da liễu trong ISIC-2019	37
B.3	Bảng phân bố nhãn bộ ISIC-2019	38
B.4	Biểu đồ phân bố nhãn bộ ISIC-2019	38
B.5	Biểu đồ phân bố ảnh bệnh theo vị trí của bộ ISIC-2019	40
B.6	Biểu đồ phân bố ảnh bệnh theo vị trí của bộ ISIC-2019	42
C.1	Màn quản lý hình ảnh	44
C.2	Màn gán nhãn hình ảnh	44
C.3	Chức năng tự động nhận diện vùng tổn thương da	45
C.4	Quản lý bệnh án	47

# DANH SÁCH BẢNG

1	Số lượt bệnh nhân đến khám tại Bệnh viện Da liễu Trung ương giai đoạn 2016-2020 . . . . .	1
3.1	Kết quả tăng cường dữ liệu với mô hình ResNet50 . . . . .	62
3.2	Kết quả tăng cường dữ liệu với mô hình ResNet101 . . . . .	63
3.3	Kết quả tăng cường dữ liệu với mô hình ResNet152 . . . . .	63
3.4	Kết quả tăng cường dữ liệu với mô hình VGG16 . . . . .	63
3.5	Kết quả Dropout cho các mô hình ResNet và VGG16 (Các giá trị nổi bật được in đậm) . . . . .	64
3.6	Báo cáo phân loại của mô hình ResNet-101. . . . .	67
3.7	Báo cáo phân loại của ResNet101 khi không sử dụng phân đoạn. . . . .	69

# MỞ ĐẦU

## 1 Tính cấp thiết của đề tài

Trong những năm gần đây, tỷ lệ mắc các bệnh da liễu, đặc biệt là ung thư da, đã tăng lên đáng kể trên toàn cầu. Theo thống kê của Tổ chức Y tế Thế giới (WHO), ung thư da là một trong những loại ung thư phổ biến nhất và có tỷ lệ tử vong cao nếu không được phát hiện sớm. Năm 2006, báo cáo của WHO cho thấy bệnh da là bệnh đứng hàng thứ tư về gánh nặng bệnh tật toàn cầu, với hàng tỷ người mắc [1]. Tại nhiều quốc gia, việc phát hiện và chẩn đoán sớm các loại tổn thương da như u hắc tố ác tính (melanoma) và các bệnh lý da liễu khác đóng vai trò quan trọng trong việc cải thiện kết quả điều trị và giảm thiểu chi phí y tế.

Việc chẩn đoán chính xác các bệnh da liễu, đặc biệt là ung thư da, đòi hỏi sự tham gia của các chuyên gia y tế có kinh nghiệm và trang thiết bị hiện đại. Tuy nhiên, ở nhiều quốc gia, bao gồm cả Việt Nam, hệ thống y tế vẫn còn hạn chế về nguồn lực, dẫn đến việc chẩn đoán chưa đạt hiệu quả tối ưu. Đặc biệt, sự thiếu hụt các bác sĩ chuyên khoa da liễu và trang thiết bị y tế tiên tiến khiến cho việc chẩn đoán bệnh da liễu tại Việt Nam gặp nhiều thách thức.

Tại Bệnh viện Da liễu Trung ương, trung tâm chuyên ngành lớn nhất về bệnh da ở Miền Bắc số lượng bệnh nhân đến khám và điều trị tăng nhanh qua từng năm. Bảng 1 mô tả số lượt bệnh nhân mắc các bệnh vẩy nến, viêm da cơ địa và ung thư da đến khám và điều trị tại Bệnh viện Da liễu Trung ương từ 2016 - 2020 [2]:

STT	Tên bệnh	Tổng lượt khám theo năm				
		2016	2017	2018	2019	2020
1	Ung thư tế bào đáy	45	62	519	903	1107
2	Ung thư tế bào vảy	3	4	104	181	221
3	Ung thư tế bào hắc tố	9	10	43	58	32
4	Viêm da cơ địa	27480	21480	34836	32556	23100
5	Vẩy nến thường	4248	3768	6480	6380	7524
6	Vẩy nến thể mũ	376	510	613	562	594
7	Vẩy nến đỏ da toàn thân	312	252	324	408	180
8	Vẩy nến thể khớp	0	0	0	72	204
9	Vẩy nến không điển hình	0	12	12	12	0

Bảng 1: Số lượt bệnh nhân đến khám tại Bệnh viện Da liễu Trung ương giai đoạn 2016-2020

Như bảng trên cho thấy, số lượng ca mắc các bệnh về da tại bệnh viện đã tăng dần trong giai đoạn từ năm 2016 đến năm 2020. Mặc dù tỷ lệ tăng không quá đột biến, nhưng đây vẫn là một dấu hiệu đáng lo ngại về sự gia tăng của các bệnh lý da liễu nghiêm trọng, đặc biệt là trong bối cảnh nhận thức về nguy cơ ung thư da trong cộng đồng chưa cao và việc tiếp cận các phương pháp chẩn đoán hiện đại còn hạn chế.

Trong khi đó, sự phát triển nhanh chóng của các công nghệ trí tuệ nhân tạo, đặc biệt là học máy (machine learning) và học sâu (deep learning), đã mở ra những cơ hội mới cho lĩnh vực y tế, đặc biệt là trong chẩn đoán hình ảnh y khoa. Các phương pháp này có khả năng phân tích các đặc điểm phức tạp từ hình ảnh da liễu, giúp hỗ trợ bác sĩ trong việc phát hiện sớm và chính xác các loại bệnh da liễu, bao gồm cả ung thư da.

Luận văn này sử dụng hai bộ dữ liệu ISIC-2018<sup>1</sup>, được coi là bộ dữ liệu lớn và đa dạng nhất trong lĩnh vực da liễu, và đã trở thành tiêu chuẩn trong các nghiên cứu về ứng dụng học máy và học sâu cho chẩn đoán các bệnh da liễu. Việc khai thác hiệu quả những bộ dữ liệu này đòi hỏi các mô hình học sâu có khả năng xử lý và phân tích chính xác các đặc điểm phức tạp của hình ảnh da liễu, đảm bảo tính chính xác và khả năng tổng quát hóa cao.

Do đó, nghiên cứu và phát triển các phương pháp học máy và học sâu nhằm hỗ trợ chẩn đoán bệnh da liễu là vô cùng cấp thiết. Điều này không chỉ có ý nghĩa quan trọng đối với Việt Nam mà còn đóng góp vào sự phát triển chung của lĩnh vực y học toàn cầu. Việc ứng dụng các công nghệ hiện đại vào chẩn đoán bệnh không chỉ giúp giảm bớt gánh nặng cho các bác sĩ chuyên khoa mà còn cải thiện chất lượng chăm sóc sức khỏe, tăng cường khả năng phát hiện sớm và điều trị các bệnh lý da liễu nguy hiểm, từ đó giảm thiểu tỷ lệ tử vong do ung thư da.

## 2 Mục tiêu của luận văn

### Đặt vấn đề

Bệnh da liễu, đặc biệt là ung thư da, là một vấn đề y tế toàn cầu với tỷ lệ mắc bệnh ngày càng gia tăng. Việc chẩn đoán sớm và chính xác các bệnh lý da liễu có ý nghĩa quan trọng trong việc điều trị hiệu quả và giảm thiểu nguy cơ tử vong, nhất là đối với các loại ung thư da như melanoma. Tuy nhiên, quá trình chẩn đoán này đòi hỏi sự tham gia của các chuyên gia y tế giàu kinh nghiệm cùng với các thiết bị hiện đại. Tại các quốc gia đang phát triển như Việt Nam, các nguồn lực y tế còn hạn chế, khiến cho việc tiếp cận các dịch vụ chẩn đoán và điều trị chất lượng cao trở nên khó khăn hơn.

Với sự tiến bộ của trí tuệ nhân tạo (AI), đặc biệt là học máy (machine learning) và học sâu (deep learning), các hệ thống chẩn đoán tự động dựa trên hình ảnh y khoa đã mở ra những cơ hội mới để hỗ trợ các bác sĩ trong việc phát hiện sớm các dấu hiệu bệnh lý. Các mô hình

---

<sup>1</sup><https://challenge.isic-archive.com/data/#2018>

học sâu, khi được huấn luyện với các tập dữ liệu lớn và đa dạng như ISIC-2018, có khả năng phân tích các hình ảnh da liễu với độ chính xác cao, từ đó giúp tối ưu hóa quy trình chẩn đoán bệnh da liễu, đặc biệt là ung thư da.

Tuy nhiên, để đạt được hiệu quả tối đa, các mô hình cần được tối ưu hóa cho phù hợp với dữ liệu cụ thể và phải đáp ứng các tiêu chuẩn cao về độ chính xác, tốc độ xử lý, và khả năng giải thích kết quả. Do đó, nghiên cứu và phát triển các mô hình học sâu nhằm nâng cao hiệu quả chẩn đoán các bệnh da liễu trở thành một nhu cầu cấp thiết, đặc biệt trong bối cảnh số lượng ca bệnh da liễu đang gia tăng nhanh chóng tại Việt Nam và trên toàn thế giới.

## Thách thức trong chẩn đoán bệnh da liễu từ hình ảnh

Mặc dù có nhiều tiềm năng, việc chẩn đoán bệnh da liễu từ hình ảnh vẫn đối mặt với nhiều thách thức lớn. Một trong những thách thức chính là **tính phức tạp của hình ảnh da liễu**. Hình ảnh da có thể rất đa dạng, với nhiều yếu tố gây nhiễu như ánh sáng, màu da, hoặc sự xuất hiện của các loại bệnh lý khác nhau. Điều này đòi hỏi các mô hình học sâu phải có khả năng phân tích chính xác từng chi tiết nhỏ để đưa ra dự đoán chính xác.

Một khó khăn khác là **dữ liệu không cân bằng**. Trong các bộ dữ liệu hình ảnh y tế, số lượng hình ảnh của các bệnh hiếm gặp, chẳng hạn như ung thư da, thường ít hơn rất nhiều so với các loại bệnh thông thường hoặc lành tính. Điều này có thể dẫn đến hiện tượng mất cân bằng dữ liệu, làm cho mô hình học máy dễ bị thiên vị và không hoạt động tốt trên các lớp dữ liệu hiếm. Việc khắc phục vấn đề này là một thách thức lớn trong việc phát triển hệ thống chẩn đoán tự động.

Thêm vào đó, **chất lượng hình ảnh không đồng nhất** cũng gây ra khó khăn trong việc huấn luyện và đánh giá các mô hình. Hình ảnh da liễu có thể được chụp dưới nhiều điều kiện khác nhau, với độ phân giải, ánh sáng, và chất lượng khác nhau. Điều này làm cho mô hình học sâu khó có thể học được các đặc trưng chung, từ đó ảnh hưởng đến khả năng tổng quát hóa và áp dụng của mô hình trên các dữ liệu mới.

## Đối tượng nghiên cứu

Đối tượng nghiên cứu của luận văn là việc ứng dụng các mô hình học sâu vào bài toán chẩn đoán bệnh da liễu, với trọng tâm là sử dụng và kết hợp hai thành phần chính: **phân đoạn vùng tổn thương da** và **phân lớp các loại tổn thương**. Cụ thể, luận văn nghiên cứu cách các mô hình tiên tiến như U-Net (cho phân đoạn) và ResNet, VGG (cho phân lớp) có thể được tích hợp trong một hệ thống cộng tác, nhằm nâng cao độ chính xác và hiệu quả chẩn đoán trên bộ dữ liệu chuẩn ISIC-2018.

Bên cạnh đó, luận văn cũng tập trung vào việc xử lý các thách thức đặc thù của dữ liệu da liễu, bao gồm:

- **Sự đa dạng:** Hình dạng, kích thước, và màu sắc của các tổn thương rất đa dạng.



- **Sự mất cân bằng:** Một số loại tổn thương xuất hiện với tần suất rất thấp, gây khó khăn trong quá trình huấn luyện.
- **Nhiều nền:** Ảnh hưởng của nền không liên quan và điều kiện ánh sáng không đồng đều trong hình ảnh.

Mục tiêu của nghiên cứu là phát triển một phương pháp cộng tác giữa phân đoạn và phân lớp, không chỉ cải thiện khả năng xác định chính xác vùng tổn thương mà còn giúp phân loại bệnh lý da với độ chính xác cao. Qua đó, luận văn hướng tới tạo nền tảng cho việc ứng dụng các hệ thống hỗ trợ chẩn đoán y khoa trong thực tế.

## Mục tiêu nghiên cứu

Mục tiêu của luận văn là phát triển một hệ thống học sâu hiệu quả để hỗ trợ chẩn đoán bệnh da liễu, tập trung vào việc cải thiện độ chính xác và hiệu suất của mô hình thông qua sự kết hợp cộng tác giữa hai thành phần chính: **phân đoạn vùng tổn thương** và **phân lớp các loại tổn thương da**. Hệ thống được xây dựng và đánh giá trên bộ dữ liệu chuẩn ISIC-2018, nhằm đáp ứng các yêu cầu đặc thù trong lĩnh vực y tế.

Cụ thể, luận văn hướng đến các mục tiêu chi tiết sau:

### 1. Xây dựng mô hình phân đoạn:

- Sử dụng U-Net và các biến thể của nó để xác định chính xác vùng tổn thương trong ảnh da liễu.
- Đảm bảo việc tách biệt rõ ràng giữa vùng tổn thương và vùng da bình thường hoặc nền.

### 2. Xây dựng mô hình phân lớp: Ứng dụng các mô hình học sâu tiên tiến như ResNet và VGG để phân loại vùng tổn thương đã được xác định thành các loại bệnh lý da liễu cụ thể.

### 3. Tích hợp mô hình cộng tác phân đoạn - phân lớp:

- Kết hợp hai thành phần phân đoạn và phân lớp trong một hệ thống cộng tác.
- Tận dụng tối đa ưu điểm của từng thành phần để cải thiện hiệu suất chẩn đoán toàn diện.

### 4. Đánh giá và tối ưu hóa mô hình:

- Đánh giá hiệu quả của hệ thống trên bộ dữ liệu ISIC-2018, sử dụng các chỉ số đo lường như Dice Similarity Coefficient (DSC), Accuracy, và F1-Score.
- Tối ưu hóa mô hình để đảm bảo tính chính xác và khả năng tổng quát hóa cao, phù hợp với dữ liệu thực tế.

Mục tiêu cuối cùng của nghiên cứu là xây dựng một phương pháp có tính ứng dụng cao, có thể hỗ trợ bác sĩ trong việc chẩn đoán bệnh da liễu, từ đó góp phần nâng cao hiệu quả điều trị và cải thiện chất lượng chăm sóc y tế.

### Phạm vi nghiên cứu

- **Bệnh lý nghiên cứu:** Luận văn tập trung vào việc chẩn đoán và phân loại các loại tổn thương da khác nhau có trong bộ dữ liệu ISIC-2018. Các loại tổn thương này bao gồm nhiều dạng bệnh lý da liễu khác nhau, từ các tổn thương lành tính đến các dạng tổn thương ác tính, với mục tiêu phân lớp chính xác các bệnh lý da liễu dựa trên hình ảnh.
- **Kỹ thuật áp dụng:** Luận văn sử dụng các phương pháp học máy và học sâu, đặc biệt là các mô hình mạng nơ-ron tích chập (CNN), nhằm phát triển hệ thống phân loại tự động các loại tổn thương da. Các kỹ thuật như tiền xử lý dữ liệu, huấn luyện mô hình và tối ưu hóa quá trình học tập cũng được áp dụng để nâng cao hiệu suất phân loại.
- **Bộ dữ liệu:** Luận văn sử dụng bộ dữ liệu ISIC-2018, là những tập dữ liệu tiêu chuẩn lớn với hàng chục nghìn hình ảnh da liễu được gán nhãn bởi các chuyên gia. Các bộ dữ liệu này được sử dụng để huấn luyện, đánh giá và thử nghiệm các mô hình học sâu trong nghiên cứu.

## 3. Phương pháp nghiên cứu

Luận văn sử dụng phương pháp nghiên cứu thực nghiệm với trọng tâm là triển khai, đánh giá, và tối ưu hóa các mô hình học sâu trong bài toán chẩn đoán bệnh da liễu. Phương pháp nghiên cứu được xây dựng dựa trên quy trình cụ thể như sau:

### 1. Phân tích và chuẩn bị dữ liệu:

- Bộ dữ liệu ISIC-2018, bao gồm các ảnh da liễu và nhãn tương ứng, được sử dụng làm cơ sở cho nghiên cứu.
- Dữ liệu được tiền xử lý thông qua các bước như chuẩn hóa (*normalization*), thay đổi kích thước (*resize*), và áp dụng các kỹ thuật tăng cường dữ liệu (*data augmentation*) để cải thiện tính đa dạng và khả năng tổng quát hóa của mô hình.

### 2. Xây dựng các mô hình phân đoạn và phân lớp:

- **Mô hình phân đoạn:** U-Net và các biến thể của nó được triển khai để xác định vùng tổn thương trong ảnh da liễu. Mô hình này được huấn luyện và đánh giá dựa trên các chỉ số như Dice Similarity Coefficient (DSC) và Intersection over Union (IoU).

- **Mô hình phân lớp:** Các mạng nơ-ron tích chập (CNN) tiên tiến như ResNet và VGG được áp dụng để phân loại các vùng tổn thương đã được phân đoạn thành các loại bệnh lý da liễu.

### 3. Kết hợp mô hình phân đoạn và phân lớp:

- Mô hình phân đoạn và phân lớp được tích hợp thành một hệ thống cộng tác, trong đó kết quả của mô hình phân đoạn được sử dụng làm đầu vào cho mô hình phân lớp.
- Phương pháp này được thiết kế để tối ưu hóa hiệu suất của toàn bộ hệ thống.

### 4. Huấn luyện và đánh giá mô hình:

- Các mô hình được huấn luyện trên tập huấn luyện, kiểm tra trên tập kiểm tra, và đánh giá trên tập kiểm định của bộ dữ liệu ISIC-2018.
- Các chỉ số đánh giá chính bao gồm: độ chính xác (*accuracy*), Dice Similarity Coefficient (DSC), Intersection over Union (IoU), F1-Score, Precision, và Recall.

### 5. Phân tích và tối ưu hóa:

- Các kết quả thực nghiệm được phân tích để xác định ưu điểm và hạn chế của từng thành phần mô hình.
- Luận văn sử dụng các phương pháp tối ưu hóa, chẳng hạn như tinh chỉnh siêu tham số (*hyperparameter tuning*) và tăng cường dữ liệu, để cải thiện hiệu suất mô hình.

Phương pháp nghiên cứu này đảm bảo rằng hệ thống được xây dựng không chỉ có hiệu quả cao trong điều kiện thực nghiệm mà còn phù hợp để ứng dụng thực tế trong việc hỗ trợ chẩn đoán bệnh da liễu.

## 4 Cơ sở khoa học và tính thực tiễn của đề tài

Cơ sở khoa học của đề tài dựa trên các nghiên cứu hiện đại về học sâu và trí tuệ nhân tạo, đặc biệt là mạng nơ-ron tích chập (CNN), trong lĩnh vực phân tích hình ảnh y khoa. CNN đã chứng minh hiệu quả vượt trội trong việc tự động phát hiện và phân loại hình ảnh da liễu với độ chính xác cao. Ngoài ra, các nghiên cứu liên quan đến phương pháp học chuyển giao (transfer learning) và các kỹ thuật tiền xử lý dữ liệu cũng đóng vai trò quan trọng trong việc tối ưu hóa mô hình.

Tính thực tiễn của đề tài thể hiện ở việc ứng dụng hệ thống chẩn đoán tự động để hỗ trợ bác sĩ da liễu trong việc chẩn đoán bệnh, giúp giảm tải cho hệ thống y tế, đặc biệt tại các khu

vực thiếu chuyên gia. Hệ thống này có thể giúp cung cấp dịch vụ chăm sóc sức khỏe từ xa, nâng cao khả năng phát hiện sớm các bệnh da liễu nguy hiểm như ung thư da, từ đó giảm thiểu rủi ro và nâng cao chất lượng chăm sóc sức khỏe cho người dân.

## 5 Những đóng góp của luận văn

Luận văn này đóng góp vào lĩnh vực chẩn đoán bệnh da liễu dựa trên hình ảnh thông qua các nội dung sau:

- Đề xuất một phương pháp ứng dụng mô hình mạng nơ-ron tích chập (CNN) để phân loại và nhận diện các loại tổn thương da khác nhau từ bộ dữ liệu ISIC-2018, giúp nâng cao độ chính xác trong chẩn đoán các bệnh lý da liễu.
- Thực nghiệm và đánh giá hệ thống chẩn đoán tự động trên bộ dữ liệu hình ảnh da liễu tiêu chuẩn ISIC-2018, cung cấp kết quả có giá trị tham khảo cho các nghiên cứu tương lai và ứng dụng thực tiễn.
- Đóng góp giải pháp tối ưu hóa quy trình tiền xử lý dữ liệu hình ảnh da liễu, cải thiện chất lượng hình ảnh đầu vào và nâng cao hiệu quả của các mô hình phân loại.
- Xây dựng hệ thống hỗ trợ chẩn đoán có khả năng ứng dụng trong thực tiễn, hỗ trợ việc chẩn đoán từ xa và giảm tải cho các cơ sở y tế trong việc điều trị và theo dõi bệnh nhân.

## 6 Bố cục luận văn

Luận văn bao gồm phần mở đầu và năm chương, được tổ chức như sau:

- *"Mở đầu"*: Trình bày tổng quan về vấn đề nghiên cứu, nêu rõ tính cấp thiết và ý nghĩa của đề tài trong thực tiễn và khoa học. Phần này cũng định hướng cho toàn bộ luận văn, giới thiệu mục tiêu, đối tượng và phương pháp nghiên cứu.
- *Chương 1 "TỔNG QUAN CÁC PHƯƠNG PHÁP CHẨN ĐOÁN BỆNH DA LIỄU VỚI HỌC SÂU"*: Cung cấp cái nhìn tổng quan về các phương pháp chẩn đoán bệnh da liễu, với trọng tâm là các kỹ thuật phân tích hình ảnh. Tác giả đi sâu vào hai phương pháp chủ yếu: phân đoạn và phân lớp hình ảnh da liễu, đồng thời tóm lược các nghiên cứu tiêu biểu trong lĩnh vực này, qua đó xây dựng nền tảng lý thuyết cho nghiên cứu.
- *Chương 2 "PHƯƠNG PHÁP CHẨN ĐOÁN BỆNH DA LIỄU ĐỀ XUẤT"*: Trình bày chi tiết về mô hình chẩn đoán tự động được đề xuất, kết hợp hai bước chính: phân đoạn và phân lớp hình ảnh. Mô hình UNet được sử dụng để phân đoạn vùng tổn thương da. Sau đó, các mô hình ResNet50, ResNet101, ResNet152 và VGG16 được áp dụng cho

việc phân lớp các vùng tổn thương. Phần này cũng mô tả quy trình huấn luyện và các cải tiến trong kiến trúc mô hình nhằm tối ưu hóa hiệu suất.

- *Chương 3 "KẾT QUẢ VÀ THẢO LUẬN"*: Trình bày các thử nghiệm và kết quả đánh giá mô hình trên bộ dữ liệu ISIC-2018. Kết quả cho thấy các mô hình ResNet50, ResNet101, ResNet152, VGG16 và UNet đạt hiệu suất cao trong việc phân đoạn và phân loại tổn thương da. Các chỉ số như độ chính xác, độ nhạy, độ bao phủ và F1-score được sử dụng để đánh giá. Mô hình cũng được so sánh với các phương pháp khác nhằm kiểm chứng tính ưu việt.
- *Chương 4 "KẾT LUẬN"*: Tóm tắt những kết quả chính của nghiên cứu, khẳng định tính hiệu quả của các mô hình trong việc chẩn đoán bệnh da liễu qua hình ảnh. Chương này cũng thảo luận về các hạn chế của nghiên cứu và đưa ra những hướng phát triển trong tương lai, bao gồm việc thử nghiệm trên các bộ dữ liệu khác và mở rộng ứng dụng trong thực tiễn.

# Chương 1

## TỔNG QUAN CÁC PHƯƠNG PHÁP CHẨN ĐOÁN BỆNH DA LIỄU VỚI HỌC SÂU

Sự phát triển mạnh mẽ của công nghệ xử lý hình ảnh và trí tuệ nhân tạo (AI) trong những năm gần đây đã mang lại nhiều tiến bộ quan trọng trong việc hỗ trợ chẩn đoán bệnh da liễu. Đặc biệt, các hệ thống dựa trên học máy (machine learning) và học sâu (deep learning) đã trở thành những công cụ hữu ích giúp các bác sĩ có thể phát hiện và chẩn đoán bệnh chính xác hơn, nhanh chóng hơn so với phương pháp truyền thống. Trong chương này, chúng tôi sẽ trình bày tổng quan về các nghiên cứu liên quan đến chẩn đoán bệnh da liễu qua hình ảnh, từ đó cung cấp cái nhìn toàn diện về các phương pháp và kỹ thuật hiện có. Chương này cũng sẽ tập trung phân tích các phương pháp phân đoạn và phân lớp hình ảnh da liễu, hai kỹ thuật quan trọng trong các hệ thống chẩn đoán tự động.

### 1.1 Các phương pháp chẩn đoán bệnh da liễu

Chẩn đoán bệnh da liễu là một trong những thách thức lớn trong y học lâm sàng, do sự đa dạng và phức tạp của các bệnh lý về da. Việc chẩn đoán không chính xác có thể dẫn đến sai lầm trong điều trị, gây hậu quả nghiêm trọng cho bệnh nhân. Do đó, nghiên cứu và phát triển các phương pháp chẩn đoán chính xác và hiệu quả là vô cùng cần thiết. Trong những thập kỷ gần đây, nhiều phương pháp chẩn đoán bệnh da liễu đã được đề xuất và triển khai, từ các phương pháp truyền thống dựa trên quan sát của bác sĩ cho đến các kỹ thuật hiện đại sử dụng học máy (Machine Learning - ML) và học sâu (Deep Learning - DL). Phần này sẽ tập trung phân tích các phương pháp chẩn đoán chính, bao gồm phương pháp truyền thống, các phương pháp dựa trên hình ảnh và ứng dụng của ML và DL trong chẩn đoán da liễu.

### 1.1.1 Phương pháp truyền thống

Các phương pháp truyền thống trong chẩn đoán bệnh da liễu dựa chủ yếu vào quan sát lâm sàng và kinh nghiệm của bác sĩ. Bác sĩ sử dụng mắt thường để quan sát các đặc điểm về màu sắc, hình dạng, kích thước và kết cấu của tổn thương trên da. Các kỹ thuật hỗ trợ như sinh thiết da (skin biopsy) hoặc kính hiển vi ánh sáng cũng được sử dụng để cung cấp thông tin chi tiết hơn về cấu trúc mô da và giúp phân loại chính xác các bệnh lý.

Mặc dù các phương pháp này đã được sử dụng rộng rãi trong nhiều năm, chúng vẫn tồn tại một số hạn chế. Khả năng chẩn đoán phụ thuộc nhiều vào kinh nghiệm của bác sĩ, và một số bệnh lý có triệu chứng tương tự nhau, làm tăng nguy cơ chẩn đoán sai. Ngoài ra, việc sử dụng các kỹ thuật xâm lấn như sinh thiết có thể gây đau đớn và mất thời gian.

#### *Chẩn đoán lâm sàng*

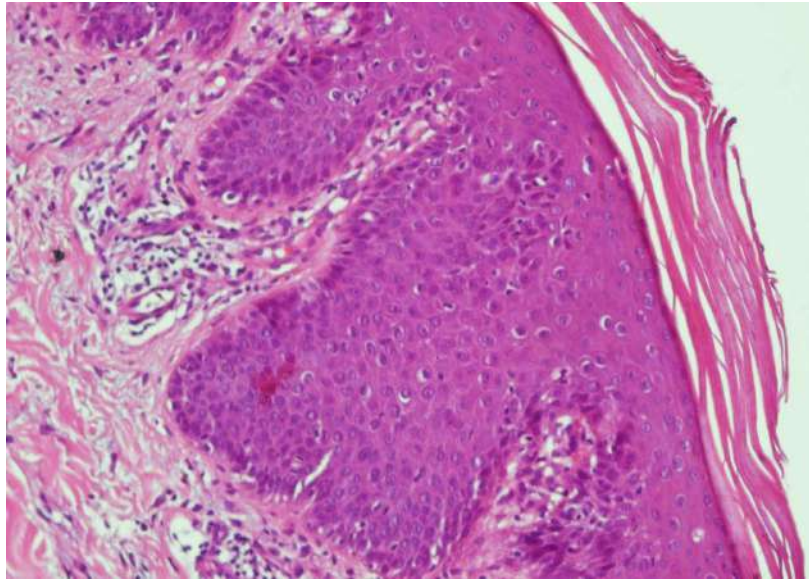


Hình 1.1: Chẩn đoán lâm sàng bệnh da liễu. Nguồn: Medlatex

Chẩn đoán lâm sàng là phương pháp dựa trên kinh nghiệm và chuyên môn của các bác sĩ da liễu. Các bác sĩ sử dụng khả năng quan sát, kết hợp với các triệu chứng lâm sàng và lịch sử bệnh của bệnh nhân để đưa ra chẩn đoán. Phương pháp này đòi hỏi thời gian, kỹ năng và kiến thức sâu rộng về các bệnh lý da liễu. Tuy nhiên, việc chẩn đoán bằng mắt thường có thể gặp khó khăn trong các trường hợp bệnh phức tạp hoặc khi các biểu hiện trên da không rõ ràng. Vì vậy, các công cụ hỗ trợ chẩn đoán dựa trên công nghệ hình ảnh đã dần trở thành sự bổ sung quan trọng cho phương pháp này.

### ***Chẩn đoán dựa trên xét nghiệm***

Trong nhiều trường hợp, để có chẩn đoán chính xác hơn, bác sĩ cần thực hiện thêm các xét nghiệm y tế, chẳng hạn như sinh thiết hoặc phân tích mô học (Hình 1.2 mô tả một hình ảnh mô bệnh dùng trong xét nghiệm). Sinh thiết da liên quan đến việc lấy mẫu nhỏ của da bị tổn thương để phân tích dưới kính hiển vi, giúp phát hiện các bất thường ở cấp độ tế bào và mô học. Đây là phương pháp chuẩn xác nhưng xâm lấn và đòi hỏi thời gian để có kết quả, khiến cho việc sử dụng nó trong các chẩn đoán nhanh là không thực tế.



Hình 1.2: Ảnh mô bệnh học dùng trong chẩn đoán bệnh da liễu trên xét nghiệm. Nguồn: KC-4.0-44/19-25

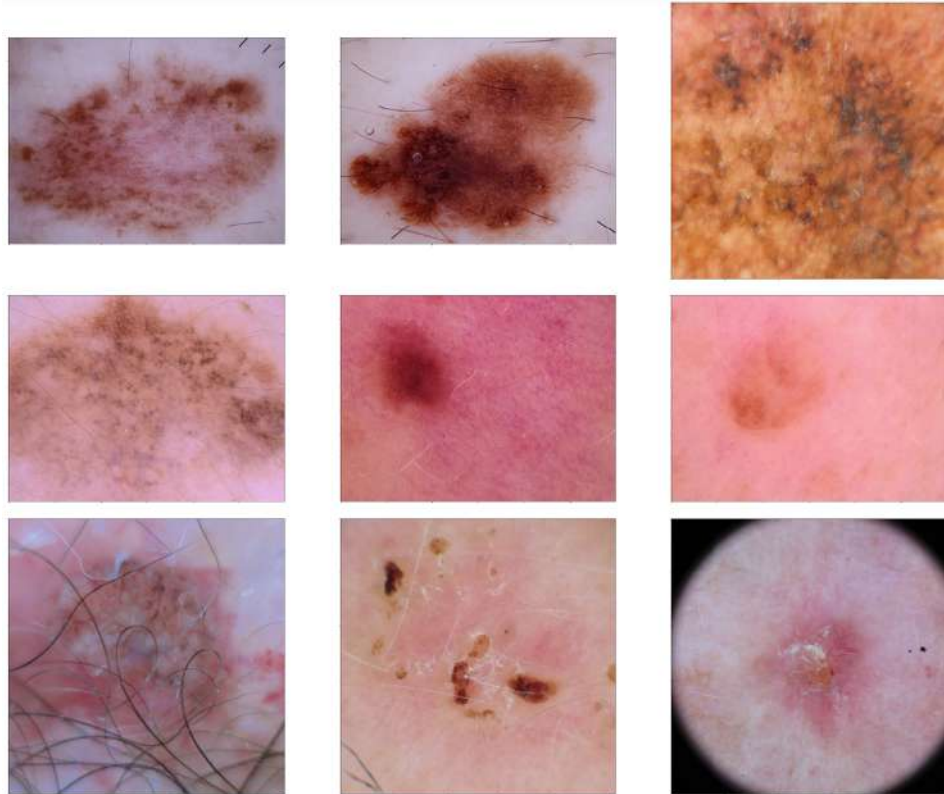
### **1.1.2 Chẩn đoán dựa trên hình ảnh**

Với sự phát triển của công nghệ hình ảnh y học, các phương pháp chẩn đoán dựa trên hình ảnh đã trở thành một công cụ quan trọng trong việc phát hiện và phân loại các bệnh da liễu. Các công nghệ như kính hiển vi phản quang laser (confocal laser microscopy), siêu âm da, và Dermoscopy đã mang lại khả năng quan sát các lớp da sâu hơn mà không cần can thiệp xâm lấn.

Dermoscopy (hình 1.3) là một trong những phương pháp phổ biến nhất, đặc biệt hữu ích trong chẩn đoán ung thư da như u hắc tố (melanoma). Bằng cách sử dụng các công cụ hình ảnh độ phân giải cao, bác sĩ có thể quan sát được các cấu trúc dưới bề mặt da, từ đó xác định được các đặc điểm bất thường như sự phân bố không đều của sắc tố, các đường viền không rõ ràng, hay những thay đổi về mạch máu.

Tuy nhiên, các phương pháp này vẫn cần sự can thiệp và đánh giá của con người, do đó vẫn tồn tại rủi ro chẩn đoán sai nếu bác sĩ thiếu kinh nghiệm hoặc kỹ năng trong việc phân tích hình ảnh.





Hình 1.3: Ảnh dermoscopy dùng trong chẩn đoán bệnh da liễu. Nguồn: HAM10000

### 1.1.3 Ứng dụng học máy và học sâu trong chẩn đoán da liễu

Với sự phát triển mạnh mẽ của công nghệ thông tin và trí tuệ nhân tạo (AI), các phương pháp học máy (ML) và học sâu (DL) đã dần trở thành xu hướng trong việc hỗ trợ chẩn đoán bệnh da liễu, đặc biệt là dựa trên hình ảnh da. Các mô hình AI có thể học từ dữ liệu lớn, từ đó tự động phát hiện các mẫu hình, tính năng quan trọng và đưa ra dự đoán với độ chính xác cao.

#### *Học máy trong chẩn đoán da liễu*

Các phương pháp học máy truyền thống, chẳng hạn như máy vector hỗ trợ (SVM), k-láng giềng gần nhất (k-NN), hay mạng nơ-ron nhân tạo (ANN), đã được sử dụng để phát triển các hệ thống hỗ trợ chẩn đoán dựa trên dữ liệu hình ảnh da. SVM thường được sử dụng do khả năng phân loại tốt trong các bài toán phi tuyến. Kỹ thuật học máy này đặc biệt hiệu quả khi kết hợp với các phương pháp trích xuất đặc trưng (feature extraction) từ hình ảnh da như phân đoạn hình ảnh (image segmentation) và phân tích kết cấu.

#### *Học sâu trong chẩn đoán da liễu*

Gần đây, học sâu (Deep Learning) đã tạo ra bước đột phá trong chẩn đoán hình ảnh y học nói chung và da liễu nói riêng. Mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) là một trong những phương pháp phổ biến nhất trong việc phân tích hình ảnh da. CNN có

khả năng tự động trích xuất các đặc trưng từ hình ảnh mà không cần can thiệp của con người, giúp cải thiện đáng kể độ chính xác trong việc phát hiện các bệnh lý.

Nhiều nghiên cứu đã chứng minh rằng các mô hình CNN có thể đạt hiệu suất ngang ngửa, thậm chí vượt trội so với bác sĩ da liễu trong việc phát hiện các tổn thương da nguy hiểm như u hắc tố. Ví dụ, nghiên cứu trên bộ dữ liệu ISIC-2018 đã cho thấy khả năng của các mô hình CNN trong phân loại chính xác các loại tổn thương da dựa trên hình ảnh, đạt được độ chính xác cao hơn so với các phương pháp học máy truyền thống.

### *Thách thức và hướng phát triển*

Mặc dù các phương pháp học sâu mang lại kết quả đầy hứa hẹn, vẫn tồn tại một số thách thức. Việc thu thập và xử lý dữ liệu lớn về hình ảnh da có thể gặp khó khăn do sự khác biệt về thiết bị, điều kiện ánh sáng và độ phân giải của hình ảnh. Ngoài ra, các mô hình học sâu thường đòi hỏi tài nguyên tính toán lớn và dễ bị ảnh hưởng bởi hiện tượng overfitting nếu dữ liệu không đủ đa dạng.

Trong tương lai, các nghiên cứu cần tập trung vào việc cải thiện khả năng tổng quát hóa của các mô hình AI, phát triển các phương pháp kết hợp giữa học sâu và kiến thức chuyên môn của bác sĩ, và tăng cường tính minh bạch và giải thích được (explainability) của các mô hình AI.

## **1.2 Phân đoạn hình ảnh trong chẩn đoán bệnh da liễu**

### **1.2.1 Giới thiệu**



(a) Hình ảnh chụp da



(b) Hình ảnh phân đoạn da tổn thương

Hình 1.4: Phân đoạn hình ảnh da liễu

Phân đoạn hình ảnh (*image segmentation*) là một bước quan trọng trong phân tích và chẩn đoán hình ảnh y học, đặc biệt là trong lĩnh vực da liễu [3]. Mục tiêu chính của phân đoạn hình ảnh là tách các vùng quan trọng, chẳng hạn như các tổn thương da, khỏi các vùng không liên quan nhằm tạo điều kiện thuận lợi cho việc chẩn đoán và điều trị (hình 1.4 minh họa). Phân đoạn giúp chia một hình ảnh thành các phần nhỏ hơn (vùng hoặc đối tượng) có cùng các đặc điểm như màu sắc hoặc kết cấu, từ đó giúp xác định rõ ràng các khu vực tổn thương da để phân tích chi tiết hơn. Điều này rất hữu ích cho việc chẩn đoán các bệnh như ung thư da, viêm da hoặc các bệnh lý khác mà không cần can thiệp xâm lấn. Trong bối cảnh này, phân đoạn chính xác vùng da bị tổn thương đóng vai trò rất quan trọng trong việc phát hiện và phân loại bệnh [4]. Sự phát triển của các phương pháp học máy (ML) và học sâu (DL) đã thúc đẩy sự phát triển của các kỹ thuật phân đoạn tự động, giúp nâng cao đáng kể độ chính xác và hiệu quả.

## 1.2.2 Phương pháp phân đoạn truyền thống

Các phương pháp phân đoạn truyền thống thường dựa trên các kỹ thuật xử lý hình ảnh như ngưỡng hóa (*thresholding*) [5], phát hiện cạnh (*edge detection*) [6], và phân vùng (*region-based methods*) [7].

- **Ngưỡng hóa (*Thresholding*):** Phương pháp này phân tách các vùng sáng và tối dựa trên ngưỡng cố định hoặc động. Tuy nhiên, việc áp dụng ngưỡng hóa gặp khó khăn khi hình ảnh có độ tương phản thấp hoặc có nhiễu cao [5].
- **Phát hiện cạnh (*Edge detection*):** Sử dụng các kỹ thuật như Canny hoặc Sobel để phát hiện các đường biên của tổn thương da [6]. Tuy nhiên, phương pháp này không hiệu quả khi các vùng biên không rõ ràng.
- **Phương pháp dựa trên vùng (*Region-based methods*):** Sử dụng các kỹ thuật như *Region Growing* hoặc *Watershed* để phân chia ảnh thành các vùng đồng nhất về màu sắc hoặc kết cấu [7]. Những phương pháp này thường không chính xác khi hình ảnh có sự biến đổi phức tạp về màu sắc.

Mặc dù các phương pháp truyền thống dễ hiểu và dễ thực hiện, chúng thường gặp khó khăn khi xử lý các hình ảnh phức tạp với sự biến đổi đa dạng của tổn thương da.

## 1.2.3 Phương pháp phân đoạn dựa trên học máy

### 1.2.3.1 Phương pháp dựa trên trích xuất đặc trưng

Các phương pháp học máy truyền thống như *máy vector hỗ trợ (SVM)* [3], *k-láng giềng gần nhất (k-NN)* [8], và *cây quyết định (Decision Trees)* [9] thường được sử dụng kết hợp với trích

xuất đặc trưng như *Histogram of Oriented Gradients (HOG)* [10] và *Local Binary Patterns (LBP)* [11].

### 1.2.3.2 Học máy thống kê

Các phương pháp như *Random Forests* [12] hoặc *Markov Random Fields* [13] dựa trên mô hình thống kê cũng đã được áp dụng trong phân đoạn da. Các phương pháp này tận dụng thông tin không gian giữa các điểm ảnh để tăng cường tính chính xác, đặc biệt trong các trường hợp biên giới mờ giữa vùng tổn thương và da.

## 1.2.4 Phương pháp phân đoạn dựa trên học sâu

### 1.2.4.1 Mạng nơ-ron tích chập (CNN)

Mạng nơ-ron tích chập (*Convolutional Neural Networks - CNN*) đã tạo ra đột phá lớn trong phân đoạn hình ảnh, đặc biệt là hình ảnh y học. CNN có khả năng tự động trích xuất các đặc trưng quan trọng từ hình ảnh mà không cần sự can thiệp thủ công [14].

### 1.2.4.2 Kiến trúc U-Net

U-Net là một kiến trúc học sâu được thiết kế đặc biệt cho phân đoạn hình ảnh y học [15]. Với cấu trúc đối xứng, bao gồm các lớp mã hóa và giải mã, U-Net có khả năng học cả thông tin ngữ cảnh và chi tiết. Các nghiên cứu đã chỉ ra rằng U-Net và các biến thể của nó, chẳng hạn như *Attention U-Net* [16] và *Residual U-Net* [17], đạt hiệu suất vượt trội so với các phương pháp truyền thống.

Mô hình U-Net là một trong những kiến trúc học sâu được sử dụng rộng rãi nhất cho phân đoạn hình ảnh y tế. U-Net có thiết kế hình chữ U, với hai phần chính:

- **Encoder (bên trái):** Nhận diện các đặc trưng của hình ảnh và giảm dần kích thước không gian (spatial resolution).
- **Decoder (bên phải):** Tăng kích thước không gian trở lại và dự đoán phân đoạn chi tiết cho từng điểm ảnh.

U-Net đặc biệt hiệu quả trong các bài toán phân đoạn y tế nhờ khả năng xử lý tốt các đặc trưng phức tạp của ảnh da liễu và xác định chính xác các vùng tổn thương da.

### 1.2.4.3 Các phương pháp học sâu khác

Ngoài U-Net, các mô hình như *Fully Convolutional Networks (FCNs)* [18], *SegNet* [19], và *DeepLab* [20] cũng đã được áp dụng thành công trong phân đoạn tổn thương da. Các mô hình này có khả năng xử lý hiệu quả các biến thể về kích thước và hình dạng của tổn thương.

## 1.2.5 Các phương pháp tối ưu cho bài toán phân đoạn

### 1.2.5.1 Thuật toán tối ưu Adam

Trong quá trình huấn luyện mô hình, thuật toán tối ưu *Adam* (Adaptive Moment Estimation) [21] là một phương pháp tối ưu dựa trên gradient có khả năng điều chỉnh tốc độ học (learning rate) một cách tự động dựa trên các ước lượng trung bình bậc nhất và bậc hai của gradient. Điều này giúp Adam hội tụ nhanh chóng và ổn định, đặc biệt hữu ích trong các bài toán có dữ liệu phức tạp và kích thước lớn như phân đoạn hình ảnh da liễu.

Adam được chọn vì những ưu điểm nổi bật sau:

- **Điều chỉnh tốc độ học** một cách tự động cho từng tham số, giúp tối ưu hóa quá trình học mà không cần thiết lập tốc độ học cố định.
- **Kết hợp giữa các phương pháp tối ưu trước đó** như *Momentum* và *RMSProp*, từ đó giúp cân bằng giữa sự ổn định và tốc độ hội tụ của mô hình.
- **Khả năng thích ứng tốt** trong các bài toán có dữ liệu lớn và phức tạp, tránh hiện tượng gradient vanishing và exploding, đặc biệt trong các mô hình mạng nơ-ron sâu.

### 1.2.5.2 Hàm mất mát Dice Loss trong phương pháp huấn luyện

Hàm mất mát **Dice Loss** [22] là một phần quan trọng trong phương pháp huấn luyện mô hình (minh hoạ hình A.13), chi tiết xem tại phức lục A.4.6.9), đặc biệt trong bài toán phân đoạn tổn thương da. Dice Loss được thiết kế để đo lường mức độ chồng lấp giữa vùng dự đoán (*predicted mask*) và vùng thực tế (*ground truth mask*). Trong bài toán phân đoạn y tế, nơi yêu cầu sự chính xác cao về hình dạng và kích thước vùng tổn thương, Dice Loss vượt trội hơn các hàm mất mát truyền thống như *binary cross-entropy*.

**Công thức Dice Loss** Dice Loss được định nghĩa dựa trên chỉ số Dice Similarity Coefficient (DSC), một thước đo sự tương đồng giữa hai tập hợp  $A$  và  $B$ . Công thức của DSC như sau:

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|}$$

Trong ngữ cảnh phân đoạn ảnh, công thức này được viết lại như sau:

$$DSC = \frac{2 \times \sum_i p_i g_i}{\sum_i p_i + \sum_i g_i}$$

Với:

- $p_i$ : Giá trị dự đoán (*probability*) tại pixel thứ  $i$ .
- $g_i$ : Giá trị thực tế (*ground truth*) tại pixel thứ  $i$ .

Hàm mất mát Dice Loss được định nghĩa là  $1 - DSC$ :

$$Dice\ Loss = 1 - \frac{2 \times \sum_i p_i g_i}{\sum_i p_i + \sum_i g_i}$$

### Ưu điểm của Dice Loss trong huấn luyện

1. **Xử lý dữ liệu không cân bằng:** Trong bài toán phân đoạn, số lượng pixel thuộc vùng tổn thương thường nhỏ hơn nhiều so với vùng không tổn thương. Dice Loss tập trung tối đa vào sự chồng lấp giữa vùng dự đoán và vùng thực tế, do đó giảm thiểu tác động của sự mất cân bằng dữ liệu.
2. **Nhấn mạnh vào hình dạng và kích thước:** Dice Loss ưu tiên các dự đoán chính xác về biên (*boundary*) và hình dạng của vùng tổn thương, giúp mô hình U-Net đạt hiệu suất cao hơn trong việc tái hiện chi tiết cấu trúc của tổn thương da.
3. **Tương thích với U-Net:** Với kiến trúc đối xứng và các kết nối tắt của U-Net, Dice Loss tương thích hoàn hảo để tận dụng các đặc trưng chi tiết từ các tầng mã hóa và giải mã, cải thiện kết quả phân đoạn toàn diện.

**Kết hợp Dice Loss với Binary Cross-Entropy (BCE)** Trong nghiên cứu này, luận văn sử dụng hàm mất mát kết hợp giữa Dice Loss và Binary Cross-Entropy (BCE). Hàm mất mát kết hợp này có dạng:

$$Loss = \alpha \times BCE + \beta \times Dice\ Loss$$

Với:

- $\alpha$  và  $\beta$ : Hệ số cân bằng giữa hai thành phần, thường được chọn sao cho tổng bằng 1 (ví dụ,  $\alpha = 0.5, \beta = 0.5$ ).
- BCE đo lường mức độ chính xác tại từng pixel, trong khi Dice Loss tập trung vào toàn bộ hình dạng và kích thước vùng tổn thương, tạo ra sự cân bằng trong huấn luyện.

#### 1.2.5.3 Lịch trình học (*Learning Rate Schedule*)

Trong quá trình huấn luyện, luận văn sử dụng một lịch trình học giảm dần (*learning rate schedule*) để điều chỉnh tốc độ học theo từng epoch. Cụ thể, tốc độ học sẽ giảm đi một nửa nếu hàm mất mát trên tập kiểm tra không giảm sau một số epoch nhất định. Điều này giúp mô hình hội tụ tốt hơn và tránh rơi vào các cực tiểu cục bộ (*local minima*).

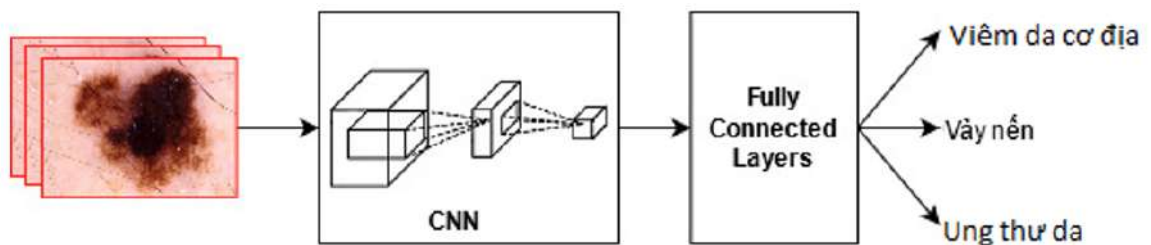
#### 1.2.6 Thách thức và hướng phát triển

Mặc dù các phương pháp học sâu đã đạt được nhiều thành công, vẫn còn tồn tại một số thách thức:

- **Độ đa dạng của dữ liệu:** Hình ảnh da có thể thay đổi theo điều kiện ánh sáng, màu da và thiết bị chụp, làm giảm khả năng tổng quát hóa của mô hình.
- **Thiếu dữ liệu gán nhãn:** Học sâu yêu cầu lượng lớn dữ liệu được gán nhãn chính xác, điều này khó khăn và tốn kém trong lĩnh vực da liễu.
- **Giải thích được của mô hình:** Các mô hình học sâu, đặc biệt trong lĩnh vực y học, cần phải minh bạch và giải thích được để tạo niềm tin cho các chuyên gia y tế.

## 1.3 Phân lớp hình ảnh trong chẩn đoán bệnh da liễu

### 1.3.1 Giới thiệu



Hình 1.5: Phân lớp hình ảnh bệnh da liễu với CNN. Nguồn: KC-4.0-44/19-25

Phân lớp hình ảnh là một bước quan trọng trong quá trình chẩn đoán bệnh da liễu, đặc biệt là khi sử dụng các phương pháp hiện đại dựa trên trí tuệ nhân tạo (AI), học máy (Machine Learning - ML) và học sâu (Deep Learning - DL). Phân lớp hình ảnh đóng vai trò xác định loại tổn thương da cụ thể từ hình ảnh y tế, giúp hỗ trợ các bác sĩ trong việc chẩn đoán các bệnh như u hắc tố (melanoma), ung thư biểu mô tế bào đáy (basal cell carcinoma), và các loại tổn thương lành tính khác.

Sự đa dạng về hình thái, màu sắc và cấu trúc của các tổn thương da khiến việc phân lớp chính xác trở thành một thách thức. Trong nhiều thập kỷ qua, các phương pháp phân lớp dựa trên kỹ thuật xử lý hình ảnh truyền thống đã được sử dụng rộng rãi. Tuy nhiên, sự ra đời của học máy và học sâu đã mở ra những cải tiến vượt bậc, cho phép tự động hóa quá trình phân lớp và đạt được độ chính xác cao hơn đáng kể [3], [4].

Phần này sẽ cung cấp một cái nhìn tổng quan về các phương pháp phân lớp hình ảnh trong chẩn đoán bệnh da liễu, từ các phương pháp truyền thống cho đến các phương pháp hiện đại dựa trên học sâu.

### 1.3.2 Phương pháp truyền thống trong phân lớp hình ảnh

Trước khi học máy và học sâu trở thành công cụ phổ biến, phân lớp hình ảnh chủ yếu dựa vào việc trích xuất các đặc trưng thủ công từ hình ảnh và áp dụng các thuật toán phân loại



truyền thống.

### ***Trích xuất đặc trưng***

Các phương pháp trích xuất đặc trưng thủ công đã được phát triển để chuyển đổi hình ảnh thô thành một tập hợp các đặc trưng có ý nghĩa, từ đó giúp phân loại các đối tượng trong hình ảnh. Các đặc trưng phổ biến được sử dụng trong phân lớp hình ảnh da liễu bao gồm:

*Histogram of Oriented Gradients (HOG)* là một phương pháp trích xuất đặc trưng phổ biến, được sử dụng để phát hiện các đường biên và cấu trúc trong hình ảnh dựa trên hướng của gradient [10]. Phương pháp này giúp trích xuất các đặc trưng về hình dạng và kết cấu của các tổn thương da, hỗ trợ quá trình phân lớp.

*Local Binary Patterns (LBP)* tập trung vào việc mã hóa thông tin kết cấu bằng cách phân tích sự thay đổi cường độ sáng giữa các điểm ảnh lân cận [11]. LBP đặc biệt hiệu quả trong việc phân loại các tổn thương có kết cấu phức tạp, như các tổn thương ung thư da.

Những đặc trưng này sau đó được đưa vào các thuật toán phân lớp truyền thống như máy vector hỗ trợ (SVM), k-láng giềng gần nhất (k-NN) và cây quyết định (Decision Trees) để phân loại các tổn thương da.

### ***Các thuật toán phân lớp truyền thống***

Máy vector hỗ trợ (SVM) là một trong những thuật toán phân lớp phổ biến nhất trong xử lý hình ảnh y học. Nó hoạt động tốt trong việc xử lý các không gian đa chiều, đặc biệt khi kết hợp với các hàm kernel để giải quyết các vấn đề phân loại phi tuyến [3]. Trong chẩn đoán bệnh da liễu, SVM đã được sử dụng thành công để phân loại các tổn thương da thành các loại lành tính và ác tính dựa trên các đặc trưng đã trích xuất.

Phương pháp k-láng giềng gần nhất (k-NN) phân loại các điểm ảnh dựa trên nhãn của các điểm lân cận gần nhất trong không gian đặc trưng [8]. Mặc dù dễ hiểu và dễ thực hiện, k-NN thường gặp khó khăn với các bộ dữ liệu lớn, đòi hỏi hiệu năng tính toán cao và khả năng phân lớp không tốt khi có nhiều nhiễu trong dữ liệu.

Cây quyết định (Decision Trees) phân loại các điểm ảnh bằng cách chia nhỏ không gian đặc trưng thành các vùng có quy tắc dựa trên các thuộc tính quan trọng nhất [9]. Phương pháp này có ưu điểm là trực quan và dễ hiểu, nhưng dễ dẫn đến tình trạng quá khớp (overfitting) nếu không có biện pháp điều chỉnh phù hợp.

Mặc dù các phương pháp truyền thống đã đóng góp đáng kể vào quá trình phát triển phân lớp hình ảnh, chúng thường gặp khó khăn trong việc xử lý các bộ dữ liệu lớn và phức tạp, nơi mà các phương pháp học sâu hiện đại có thể vượt trội hơn.



### 1.3.3 Phân lớp hình ảnh dựa trên học sâu

Sự phát triển của học sâu, đặc biệt là mạng nơ-ron tích chập (Convolutional Neural Networks - CNN), đã mở ra một kỷ nguyên mới trong phân lớp hình ảnh y học. Các mô hình học sâu có khả năng tự động học các đặc trưng từ hình ảnh mà không cần can thiệp thủ công, giúp cải thiện đáng kể độ chính xác trong việc phân loại các tổn thương da [4].

#### *Mạng nơ-ron tích chập (CNN)*

CNN là nền tảng của các mô hình phân lớp hình ảnh hiện đại. CNN hoạt động thông qua các lớp tích chập, trong đó các đặc trưng như cạnh, góc, hình dạng được tự động học từ hình ảnh. Một số đặc trưng quan trọng được học ở các lớp đầu tiên, trong khi các đặc trưng phức tạp hơn được học ở các lớp sâu hơn [14].

Các nghiên cứu gần đây đã chứng minh rằng CNN có thể đạt được độ chính xác cao hơn so với các phương pháp truyền thống trong phân loại các tổn thương da. Ví dụ, nghiên cứu của Esteva và cộng sự (2017) đã chỉ ra rằng CNN có thể phân loại các loại ung thư da, bao gồm u hắc tố, với độ chính xác ngang ngửa với các bác sĩ da liễu [23].

#### *Kiến trúc CNN nâng cao*

Các kiến trúc CNN hiện đại như ResNet (Residual Networks) và DenseNet (Densely Connected Convolutional Networks) đã được phát triển để giải quyết các vấn đề về độ sâu của mạng và tình trạng gradient suy giảm [24], [25]. ResNet sử dụng các kết nối bỏ qua (skip connections) để giúp mạng học tốt hơn khi độ sâu tăng lên, còn DenseNet sử dụng các kết nối dày đặc giữa các lớp, giúp tăng cường khả năng truyền thông tin và cải thiện độ chính xác trong phân lớp.

#### *Các mô hình kết hợp*

Ngoài các kiến trúc CNN, các mô hình kết hợp như Attention Mechanism đã được ứng dụng để giúp mô hình tập trung vào các vùng quan trọng trong hình ảnh, chẳng hạn như vùng tổn thương da có nguy cơ cao [16]. Kỹ thuật Transfer Learning cũng được sử dụng rộng rãi để tận dụng các mô hình đã huấn luyện trên tập dữ liệu lớn và áp dụng cho các bài toán cụ thể như phân lớp hình ảnh da liễu [26].

### 1.3.4 Các phương pháp tối ưu cho bài toán phân lớp

#### *1.3.4.1 Tối ưu hóa Adam*

Mô hình ResNet trong bài toán phân loại ảnh y tế được biên dịch (compile) bằng thuật toán tối ưu hóa Adam [27], một trong những phương pháp tối ưu hóa được sử dụng phổ biến và được đánh giá cao trong cộng đồng học sâu. Adam (Adaptive Moment Estimation) kết hợp

giữa hai phương pháp là AdaGrad và RMSProp, tận dụng lợi thế của cả hai để điều chỉnh tốc độ học (learning rate) trong quá trình huấn luyện.

Adam được thiết kế để đạt được sự cân bằng giữa việc học nhanh chóng và khả năng hội tụ ổn định. Thuật toán này tính toán các giá trị trung bình và trung bình bình phương của gradient theo thời gian, từ đó tự động điều chỉnh tốc độ học cho mỗi tham số của mô hình. Điều này giúp Adam thích ứng tốt với các bài toán có dữ liệu đa dạng và phức tạp, đặc biệt là trong các mạng nơ-ron sâu như ResNet.

Công thức của Adam sử dụng hai tham số chính:

- $\beta_1$  (thường là 0.9): Hệ số quán tính của moment đầu tiên (mean).
- $\beta_2$  (thường là 0.999): Hệ số quán tính của moment thứ hai (uncentered variance).

Adam giúp mô hình hội tụ nhanh hơn so với các phương pháp tối ưu hóa khác như SGD (Stochastic Gradient Descent), đồng thời có khả năng tránh khỏi các điểm kém tối ưu trong không gian tham số.

#### 1.3.4.2 Hàm mất mát *Categorical Crossentropy*

Trong bài toán phân loại nhiều lớp, hàm mất mát Categorical Crossentropy [28] là lựa chọn phù hợp nhất. Hàm mất mát này đo lường sự khác biệt giữa phân phối xác suất mà mô hình dự đoán và phân phối xác suất mục tiêu. Cụ thể, hàm mất mát này được định nghĩa như sau:

$$Loss = - \sum_{i=1}^N y_i \log(\hat{y}_i)$$

Trong đó:

- $y_i$  là nhãn thực tế của lớp thứ  $i$ .
- $\hat{y}_i$  là xác suất mô hình dự đoán cho lớp thứ  $i$ .
- $N$  là số lượng lớp.

Categorical Crossentropy đặc biệt hữu ích trong các bài toán mà mỗi mẫu đầu vào chỉ thuộc về một lớp duy nhất trong số nhiều lớp (one-hot encoding). Hàm mất mát này sẽ trừng phạt mạnh các dự đoán sai, giúp mô hình học cách phân biệt giữa các lớp khác nhau một cách hiệu quả.

#### *Kỹ thuật chuẩn hoá dữ liệu*

**Preprocessing\_function** là hàm tiền xử lý đặc biệt. Trong quá trình huấn luyện mô hình ResNet50, luận văn đã sử dụng hàm `preprocess_input`<sup>1</sup> từ thư viện Keras, đảm bảo chuẩn

<sup>1</sup>`tf.keras.applications.resnet50.preprocess_input`

hóa dữ liệu đầu vào trước khi đưa vào mô hình. Hàm này chuẩn hóa ảnh theo cách tương thích với cách mà ResNet50 được huấn luyện trên tập dữ liệu ImageNet, bao gồm các bước sau:

- Chuẩn hóa giá trị điểm ảnh (Pixel Scaling): Giá trị điểm ảnh từ phạm vi  $[0, 255]$  được chuẩn hóa về khoảng  $[-1, 1]$ , giúp mô hình hội tụ nhanh hơn trong quá trình huấn luyện.
- Trừ giá trị trung bình kênh màu (Channel-wise Mean Subtraction): Trừ đi giá trị trung bình của từng kênh màu RGB (đỏ, lục, lam) nhằm giảm thiểu sự khác biệt về cường độ ánh sáng và màu sắc giữa các hình ảnh, giúp mô hình tập trung vào các đặc trưng hình dạng và cấu trúc của tổn thương da.

## 1.4 Tổng quan về U-Net, ResNet và VGG

### 1.4.1 U-Net

U-Net là một kiến trúc mạng nơ-ron tích chập (CNN) được phát triển đặc biệt để giải quyết bài toán phân đoạn ảnh (semantic segmentation). Mô hình này đặc biệt hữu ích trong các ứng dụng y tế, như phân đoạn hình ảnh y khoa, nhờ khả năng học được các đặc trưng đa cấp độ của ảnh và tái tạo lại ảnh với độ phân giải cao.

#### 1.4.1.1 Bài toán phân đoạn hình ảnh

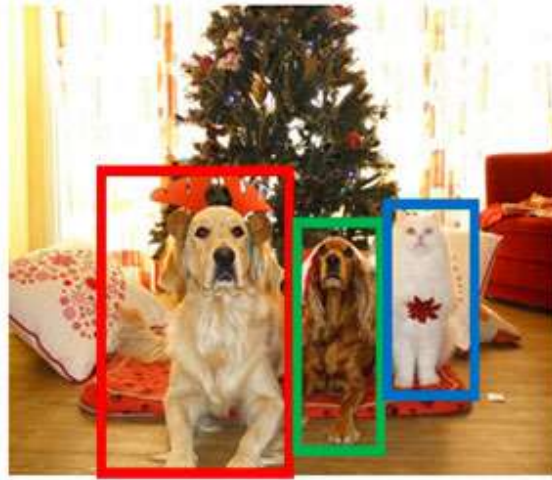
Trong lĩnh vực xử lý ảnh, có nhiều bài toán liên quan đến việc nhận diện và phân loại các đối tượng trong ảnh. Một trong những bài toán quan trọng là *object detection* (nhận diện đối tượng), nơi mô hình đi tìm các *bounding box* (hộp bao) quanh các đối tượng và sau đó phân loại chúng. Tuy nhiên, việc sử dụng các *bounding box* không thể biểu thị chính xác hình dạng của đối tượng. Các vùng trong *bounding box* có thể chứa các đối tượng khác hoặc nhiễu, không liên quan đến đối tượng chính, như ví dụ minh họa trên hình 1.6.

Do đó, *image segmentation* ra đời nhằm khắc phục vấn đề này. Bài toán phân đoạn hình ảnh (image segmentation) tập trung vào việc phân chia ảnh thành các vùng khác nhau, mỗi vùng đại diện cho một đối tượng hoặc một phần của đối tượng. Điều này cho phép hệ thống xác định chính xác hình dạng của đối tượng, thay vì chỉ đơn thuần là đóng khung như trong bài toán *object detection*. Hình 1.6 dưới đây minh họa sự khác biệt giữa hai bài toán này.

Bài toán image segmentation được chia ra làm 2 loại:

- **Semantic segmentation:** Thực hiện segment với từng lớp khác nhau, ví dụ: tất cả người là 1 lớp, tất cả ô tô là 1 lớp.

## Object Detection



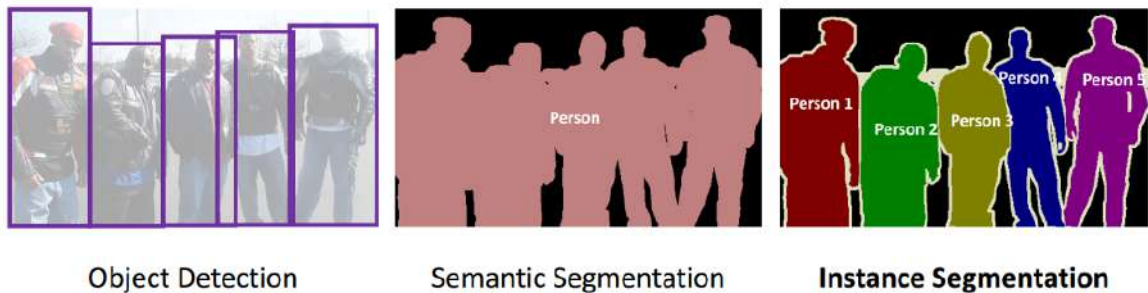
## Instance Segmentation



Hình 1.6: Sự khác nhau giữa Object detection và Instance Segmentation

- **Instance segmentation:** Thực hiện segment với từng đối tượng trong một lớp. Ví dụ có 3 người trong ảnh thì sẽ có 3 vùng segment khác nhau cho mỗi người.

Sự khác nhau rõ ràng được thể hiện qua hình 1.7.



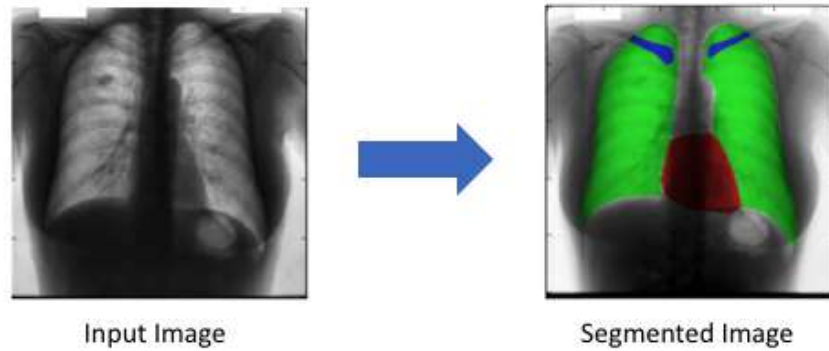
Hình 1.7: Object detection, Semantic Segmentation và Instance Segmentation

Tùy thuộc từng bài toán mà ta sử dụng kiểu segmentation nào.

Một số ứng dụng của bài toán Image segmentation:

- Ô tô tự lái: Segmentation dùng để xác định đường, các xe ô tô, người đi bộ,... để hỗ trợ cho ô tô tự lái
- Chẩn đoán trong y học: Segmentation được ứng dụng rất nhiều trong y học để hỗ trợ việc chẩn đoán bệnh. Ví dụ phân tích ảnh X-quang (Hình 1.8a).
- Xác định nông sản trong Nông nghiệp: Những kỹ thuật nhìn hình ảnh tiên tiến như vậy

cho nông nghiệp có thể làm giảm việc giám sát nông nghiệp theo cách thủ công, nâng cao hiệu năng của các Robot nông nghiệp 1.8b.



(a) Chẩn đoán trong y học



(b) Nông nghiệp

Hình 1.8: Các ứng dụng của bài toán Image Segmentation

#### 1.4.1.2 Kiến trúc mạng U-Net

Mạng U-Net [15] là một trong những kiến trúc tiên tiến được phát triển bởi Olaf Ronneberger và các cộng sự, dành riêng cho nhiệm vụ phân đoạn ảnh y học (*Bio-Medical Image Segmentation*). Đây là một mạng *fully convolutional network (FCN)* được thiết kế với khả năng học và phân đoạn chính xác các đối tượng trong ảnh với độ phân giải cao. Mạng U-Net đã chứng tỏ hiệu quả vượt trội của nó trong nhiều bài toán phân đoạn y học nhờ kiến trúc đặc biệt với hai



đường dẫn: đường dẫn mã hóa (*encoder*) và đường dẫn giải mã (*decoder*), tạo nên một cấu trúc hình chữ U.

Kiến trúc U-Net bao gồm hai thành phần chính:

- Đường dẫn co hay còn gọi là bộ mã hóa (*contracting path* hoặc *encoder*), có nhiệm vụ nắm bắt bối cảnh toàn cục của hình ảnh bằng cách giảm dần kích thước không gian của dữ liệu thông qua các phép tích chập (*convolutional layers*) và các lớp gộp cực đại (*max-pooling layers*).
- Đường dẫn mở rộng hay còn gọi là bộ giải mã (*expansive path* hoặc *decoder*), có nhiệm vụ tái tạo kích thước không gian của hình ảnh gốc và thực hiện bản địa hóa chính xác các đối tượng cần phân đoạn thông qua các phép chập chuyển vị (*transposed convolution*) để thực hiện quá trình tăng mẫu (*upsampling*).

Kiến trúc tổng thể của U-Net được mô tả như trong Hình 1.9. Đường dẫn mã hóa của mạng bao gồm một loạt các lớp tích chập với kích thước hạt nhân  $3 \times 3$ , được theo sau bởi lớp *ReLU* và lớp gộp cực đại  $2 \times 2$  để giảm kích thước không gian của ảnh. Mỗi bước giảm kích thước đều tăng gấp đôi số lượng bộ lọc, giúp mạng học được các đặc trưng phức tạp hơn ở mỗi lớp.

Một yếu tố quan trọng trong kiến trúc U-Net là các kết nối tắt (*skip connections*) giữa các tầng tương ứng trong bộ mã hóa và bộ giải mã. Các kết nối này có vai trò chuyển trực tiếp các đặc trưng từ đường dẫn mã hóa sang đường dẫn giải mã, cho phép mô hình sử dụng cả thông tin toàn cục (bối cảnh) và thông tin chi tiết (biên của đối tượng), giúp cải thiện độ chính xác trong phân đoạn hình ảnh. Mỗi kết nối tắt này sao chép và nối các đặc trưng của một lớp trong bộ mã hóa với lớp tương ứng trong bộ giải mã.

Trong kiến trúc của U-Net, ngoài các tầng tích chập và gộp cực đại thông thường, còn có sự xuất hiện của các tầng *Transposed Convolution*, hay còn được gọi là *Deconvolution* hoặc *Up-Convolution*. Mục đích chính của *Transposed Convolution* là thực hiện quá trình tăng mẫu (*upsampling*), mở rộng ma trận đầu vào để tái tạo kích thước không gian lớn hơn. Điều này giúp mạng tái lập các chi tiết bị mất trong quá trình giảm mẫu ở bộ mã hóa và khôi phục kích thước ban đầu của ảnh.

Trong khi các phép tích chập thông thường được sử dụng để thu nhỏ kích thước ảnh và học các đặc trưng từ không gian nhỏ hơn, *Transposed Convolution* thực hiện quá trình ngược lại, tăng kích thước không gian của ma trận đặc trưng, như được mô tả trong Hình 1.10. Tầng này có vai trò rất quan trọng trong bộ giải mã, giúp mạng tái cấu trúc hình ảnh và thực hiện phân đoạn ở cấp độ chi tiết cao.

Khác với các phép tích chập thông thường, nơi mà các ma trận đầu vào được thu gọn thông qua việc nhân các bộ lọc, phép *Transposed Convolution* mở rộng các ma trận đặc trưng bằng cách chèn thêm các giá trị 0 giữa các điểm dữ liệu. Điều này giúp mô hình học lại các chi tiết nhỏ bị mất trong quá trình giảm mẫu ở đường dẫn mã hóa. Phương pháp này

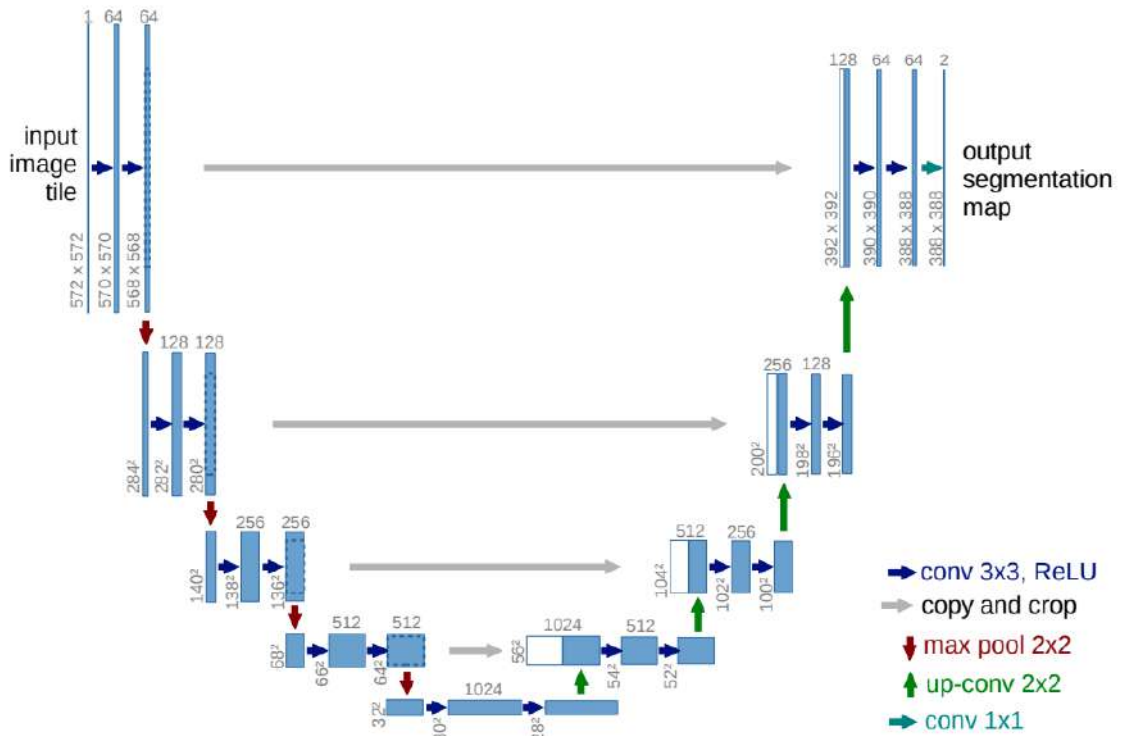
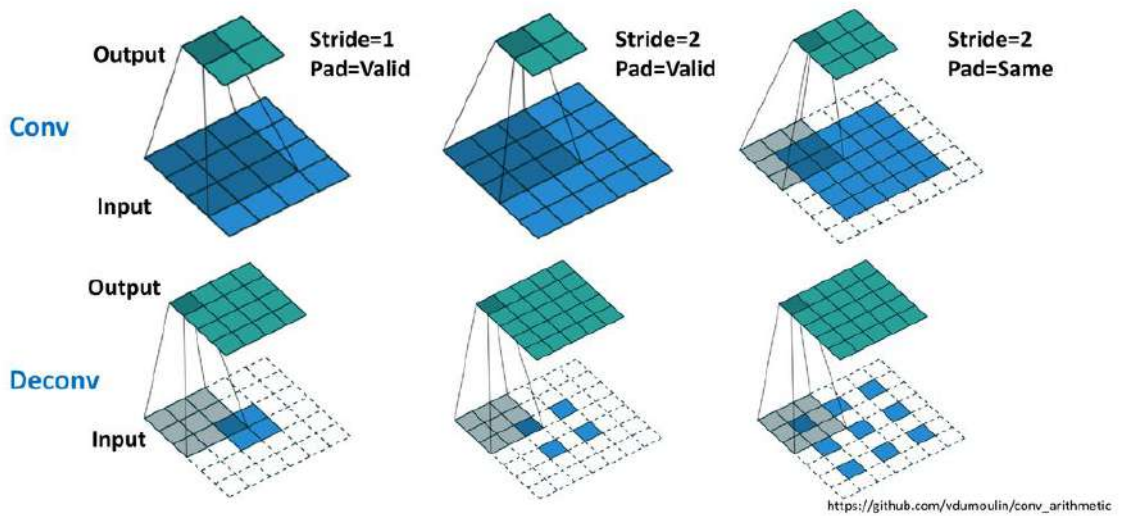


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Hình 1.9: Kiến trúc mạng U-Net với hai đường dẫn mã hóa và giải mã (Nguồn: [15])

### Transposed Convolution (Deconvolution)



Hình 1.10: Transposed convolution (Nguồn: trong hình)

cũng được sử dụng rộng rãi trong các mô hình sinh, chẳng hạn như *Generative Adversarial*

*Networks (GANs)* <sup>2</sup>.

#### **1.4.1.3 U-Net trong các bài toán phân đoạn hình ảnh y học**

Mạng U-Net đã chứng minh hiệu quả đặc biệt của nó trong các bài toán phân đoạn ảnh y học, nhờ khả năng học các đặc trưng đa cấp độ và bản địa hóa chính xác các đối tượng cần phân đoạn. Mô hình này không chỉ vượt trội trong các bài toán về hình ảnh y học như phân đoạn mô bệnh học, mà còn được áp dụng trong nhiều lĩnh vực khác như phân đoạn hình ảnh vệ tinh và ảnh viễn thám.

Khả năng đặc biệt của U-Net là có thể chấp nhận các hình ảnh có kích thước bất kỳ do nó không sử dụng các lớp kết nối đầy đủ (*fully connected layers*), điều này làm cho mạng trở nên linh hoạt trong các bài toán thực tế với các ảnh có độ phân giải không đồng nhất.

Nhờ vào thiết kế đối xứng và khả năng học sâu các đặc trưng ở nhiều cấp độ, U-Net đã trở thành một chuẩn mực trong các bài toán phân đoạn ảnh y học và tiếp tục được phát triển thêm với nhiều cải tiến khác, như *Attention U-Net* [16] và *Residual U-Net*, nhằm cải thiện hiệu suất trong các bài toán phức tạp hơn.

#### **1.4.1.4 Ưu điểm của U-Net và biến thể đối với bài toán chẩn đoán bệnh da liễu**

U-Net đã khẳng định được vị thế của mình trong các bài toán phân đoạn ảnh y tế nhờ khả năng xử lý hiệu quả những dữ liệu phức tạp đòi hỏi độ chính xác cao. Đối với bài toán chẩn đoán bệnh da liễu, mô hình này mang lại nhiều lợi thế đáng kể, giúp tối ưu hóa việc phát hiện và phân đoạn vùng tổn thương từ các hình ảnh da liễu.

Một trong những ưu điểm nổi bật của U-Net là khả năng bảo toàn thông tin chi tiết trong ảnh. Với các kết nối tắt (*skip connections*) giữa các tầng mã hóa và giải mã, mô hình không chỉ tập trung vào các đặc trưng trừu tượng ở cấp độ cao mà còn duy trì được các đặc trưng không gian chi tiết của vùng tổn thương. Điều này đặc biệt quan trọng trong chẩn đoán bệnh da liễu, nơi các tổn thương thường có hình dạng phức tạp, kích thước đa dạng và cần phân đoạn chính xác đến từng pixel.

Biến thể U-Net được sử dụng trong nghiên cứu này còn tăng cường hiệu quả thông qua việc mở rộng độ sâu của kiến trúc. Việc sử dụng 5 tầng mã hóa và giải mã giúp mô hình học được cả các đặc trưng tổng quát của tổn thương, chẳng hạn như hình dạng tổng thể, và các chi tiết nhỏ như đường viền hay kết cấu bề mặt. Điều này cải thiện đáng kể khả năng phân đoạn, đặc biệt đối với các hình ảnh có nền phức tạp hoặc tổn thương nhỏ.

Ngoài ra, sự linh hoạt trong việc xử lý dữ liệu đầu vào dưới dạng ảnh RGB cho phép mô hình tận dụng thông tin màu sắc, vốn là yếu tố quan trọng trong việc phân biệt tổn thương với các vùng da bình thường. Thông qua đầu ra dạng ảnh nhị phân, mô hình cung cấp các dự đoán có độ chính xác cao về vị trí và hình dạng vùng tổn thương, hỗ trợ trực tiếp cho các

<sup>2</sup>[https://en.wikipedia.org/wiki/Generative\\_adversarial\\_network](https://en.wikipedia.org/wiki/Generative_adversarial_network)



bước phân tích và chẩn đoán sau này.

Hơn nữa, nhờ sự tích hợp với các công cụ và thư viện mạnh mẽ như Keras và TensorFlow, biến thể U-Net không chỉ dễ dàng triển khai mà còn tối ưu hóa được hiệu suất trên các bộ dữ liệu lớn như ISIC-2018. Khả năng này không chỉ giúp giảm thời gian huấn luyện mà còn đảm bảo mô hình có thể áp dụng hiệu quả trong thực tế, ngay cả khi đối mặt với các bộ dữ liệu có tính phức tạp cao.

Tóm lại, U-Net và biến thể của nó đã chứng minh khả năng vượt trội trong bài toán phân đoạn tổn thương da. Sự kết hợp giữa kiến trúc đối xứng, khả năng học sâu các đặc trưng và tính linh hoạt trong xử lý dữ liệu đã giúp mô hình này trở thành công cụ hiệu quả trong việc hỗ trợ chẩn đoán bệnh da liễu.

## 1.4.2 ResNet và VGG

### 1.4.2.1 Tổng quan

Như đã được trình bày chi tiết tại mục 1.4.2.2, ResNet (Residual Network) và VGG (Visual Geometry Group) là hai trong số các kiến trúc mạng nơ-ron sâu tiên tiến, được sử dụng rộng rãi trong các bài toán thị giác máy tính, bao gồm phân loại và phân đoạn ảnh. Ở đây, luận văn chỉ tập trung tóm lược những đặc điểm chính của hai mô hình này, đồng thời làm rõ vai trò của chúng trong việc hỗ trợ chẩn đoán bệnh da liễu.

ResNet được phát triển để giải quyết vấn đề suy giảm gradient (vanishing gradient) khi độ sâu của mạng tăng lên. Kiến trúc này sử dụng các khối dư (residual blocks), cho phép luồng thông tin được truyền trực tiếp qua các tầng nhờ các kết nối tắt (skip connections). Khả năng này không chỉ giúp ResNet học được các đặc trưng phức tạp từ dữ liệu mà còn đảm bảo hiệu quả huấn luyện ngay cả khi mạng trở nên rất sâu. Trong bài toán chẩn đoán bệnh da liễu, ResNet mang lại lợi thế vượt trội nhờ khả năng học tốt các đặc trưng tổng quát và trích xuất các mẫu hình quan trọng từ hình ảnh tổn thương.

VGG, với cấu trúc đơn giản và trực quan hơn, tập trung vào việc sử dụng các tầng tích chập nhỏ ( $3 \times 3$ ) liên tiếp để xây dựng các đặc trưng sâu. Dù không có các kết nối tắt như ResNet, VGG vẫn nổi bật nhờ tính nhất quán trong thiết kế và khả năng hoạt động ổn định trên nhiều loại dữ liệu. Đặc biệt, các đặc trưng do VGG học được có tính tổng quát cao, giúp ích rất nhiều khi áp dụng vào các bài toán cụ thể như chẩn đoán bệnh da liễu thông qua việc phân loại tổn thương.

Cả ResNet và VGG đều đã được chứng minh là hiệu quả trên các tập dữ liệu lớn và phức tạp như ImageNet<sup>3</sup>. Sự ổn định và khả năng mở rộng của chúng khiến hai mô hình này trở thành lựa chọn lý tưởng để ứng dụng vào bài toán phân loại tổn thương da. Trong nghiên cứu này, luận văn tiếp tục khai thác sức mạnh của ResNet và VGG như những mô hình nền tảng,

<sup>3</sup><https://www.image-net.org/>

kết hợp với các cải tiến và tối ưu hóa để đạt hiệu quả cao nhất. Những chi tiết triển khai cụ thể của các mô hình này sẽ được trình bày trong phần tiếp theo.

### 1.4.2.2 ResNet

ResNet-Residual Network <sup>4</sup> được giới thiệu đến công chúng vào năm 2015 và thậm chí đã giành được vị trí thứ 1 trong cuộc thi ILSVRC 2015 với tỉ lệ lỗi top 5 chỉ 3.57%. Không những thế nó còn đứng vị trí đầu tiên trong cuộc thi ILSVRC and COCO 2015 với ImageNet Detection, ImageNet localization, Coco detection và Coco segmentation. Hiện tại thì có rất nhiều biến thể của kiến trúc ResNet với số lớp khác nhau như ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152,... Với tên là ResNet theo sau là một số chỉ kiến trúc ResNet với số lớp nhất định.

Mạng ResNet (R) là một mạng CNN được thiết kế để làm việc với hàng trăm hoặc hàng nghìn lớp chập. Một vấn đề xảy ra khi xây dựng mạng CNN với nhiều lớp chập sẽ xảy ra hiện tượng Vanishing Gradient dẫn tới quá trình học tập không tốt.

### Vanishing Gradient

Trước hết thì Backpropagation Algorithm là một kỹ thuật thường được sử dụng trong quá trình training. Ý tưởng chung của thuật toán là sẽ đi từ output layer đến input layer và tính toán gradient của cost function tương ứng cho từng parameter (weight) của mạng. Gradient Descent sau đó được sử dụng để cập nhật các parameter đó.

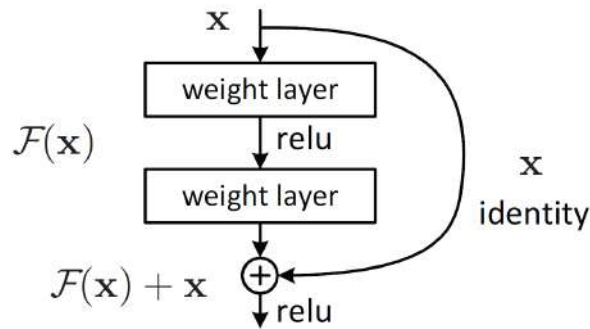
Toàn bộ quá trình trên sẽ được lặp đi lặp lại cho tới khi mà các parameter của network được hội tụ. Thông thường chúng ta sẽ có một hyperparameter (số Epoch - số lần mà training set được duyệt qua một lần và weights được cập nhật) định nghĩa cho số lượng vòng lặp để thực hiện quá trình này. Nếu số lượng vòng lặp quá nhỏ thì ta gặp phải trường hợp mạng có thể sẽ không cho ra kết quả tốt và ngược lại thời gian training sẽ lâu nếu số lượng vòng lặp quá lớn.

Tuy nhiên, trong thực tế Gradients thường sẽ có giá trị nhỏ dần khi đi xuống các layer thấp hơn. Dẫn đến kết quả là các cập nhật thực hiện bởi Gradients Descent không làm thay đổi nhiều weights của các layer đó và làm chúng không thể hội tụ và mạng sẽ không thu được kết quả tốt. Hiện tượng như vậy gọi là **Vanishing Gradients**.

### Kiến trúc mạng

Để khắc phục vấn đề này, Microsoft đã giới thiệu một framework học dư (deep residual learning framework). Thay vì hy vọng rằng mỗi vài lớp xếp chồng lên nhau sẽ phù hợp trực tiếp với một ánh xạ cơ bản mong muốn, họ để các lớp này phù hợp với một ánh xạ dư. Công thức của  $F(x) + x$  có thể được thực hiện bởi các mạng nơ-ron truyền thẳng với các kết nối

<sup>4</sup><https://neurohive.io/en/popular-networks/Resnet/>

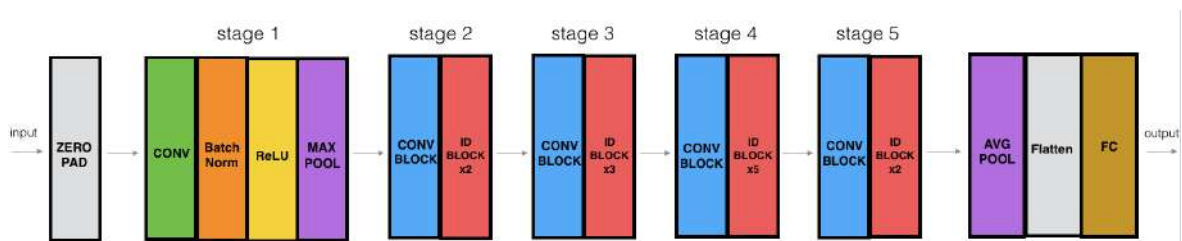


Hình 1.11: Identity block (Nguồn: Wikipedia)

tất (hình 1.11). Kết nối lối tắt là những kết nối bỏ qua một hoặc nhiều lớp được thể hiện bởi đường cong trong trên. Các kết nối lối tắt thực hiện identity mapping và đầu ra của chúng được thêm vào đầu ra của các lớp xếp chồng. Bằng cách sử dụng mạng dư, có nhiều vấn đề có thể được giải quyết như:

- ResNets rất dễ tối ưu hóa, nhưng các mạng "đơn giản" (chỉ đơn giản là xếp chồng các lớp) cho thấy lỗi đào tạo cao hơn khi độ sâu tăng lên.
- ResNets có thể dễ dàng đạt được độ chính xác từ độ sâu tăng lên đáng kể, tạo ra kết quả tốt hơn các mạng trước đó.

Hình 1.12 mô tả chi tiết kiến trúc mạng nơ ron ResNet-50<sup>5</sup>:



Hình 1.12: ResNet-50

Kiến trúc *ResNet50* là một trong những biến thể nổi tiếng của mạng *Residual Networks* (*ResNet*), được phát triển với mục đích giải quyết vấn đề gradient suy giảm trong việc huấn luyện các mô hình sâu. Điểm đặc trưng của *ResNet50* là sự kết hợp giữa các khối tích chập truyền thống và các khối đặc biệt gọi là *Residual Block*. Kiến trúc này gồm 50 lớp tích chập và sử dụng các khối *Identity Block* và *Convolutional Block* để giúp mô hình học hiệu quả các đặc trưng mà không làm mất thông tin trong các tầng sâu. Nội dung của Hình 1.12 mô tả chi tiết các giai đoạn của mạng *ResNet50* với các thành phần và kích thước cụ thể như sau:

<sup>5</sup>[https://github.com/priya-dwivedi/Deep-Learning/blob/master/resnet\\_keras/Residual\\_Networks\\_yourself.ipynb](https://github.com/priya-dwivedi/Deep-Learning/blob/master/resnet_keras/Residual_Networks_yourself.ipynb)

### ***Zero-Padding***

Trước khi bước vào giai đoạn đầu tiên của mạng, đầu vào sẽ được đệm thêm các giá trị 0 (*Zero-padding*) với kích thước bộ đệm là  $(3 \times 3)$ . Điều này giúp bảo toàn thông tin biên của ảnh trong quá trình áp dụng các phép tích chập.

### ***Giai đoạn 1: Lớp chuyển đổi và Pooling***

Trong giai đoạn này, mục tiêu là giảm kích thước không gian của ảnh và trích xuất các đặc trưng thô ban đầu. Giai đoạn này bao gồm các bước sau:

- Lớp tích chập 2D (*Convolution*) sử dụng 64 bộ lọc với kích thước  $(7 \times 7)$  và sải chân (*stride*) là  $(2 \times 2)$ . Lớp này có tên gọi là *chuyển đổi*, giúp giảm kích thước ảnh đầu vào và trích xuất các đặc trưng từ không gian lớn.
- Sau đó, *Batch Normalization (BatchNorm)* được áp dụng trên trục kênh của đầu vào nhằm ổn định quá trình huấn luyện và tăng tốc độ hội tụ của mô hình.
- Tiếp theo, lớp gộp cực đại (*MaxPooling*) được sử dụng với cửa sổ  $(3 \times 3)$  và sải chân  $(2 \times 2)$ . Lớp *MaxPooling* giúp giảm kích thước không gian của đặc trưng và chọn ra các giá trị lớn nhất trong cửa sổ tích chập.

### ***Giai đoạn 2: Các khối Residual đầu tiên***

Giai đoạn này bắt đầu với một *Convolutional Block* theo sau bởi hai *Identity Block*. Mục tiêu chính là trích xuất các đặc trưng chi tiết hơn với độ phân giải cao hơn. Cụ thể:

- *Convolutional Block*: Sử dụng ba bộ lọc với kích thước lần lượt là  $[64, 64, 256]$ . Kích thước hạt nhân (*kernel size*) của các bộ lọc là  $3 \times 3$ , và sải chân (*stride*) của lớp tích chập là  $(1 \times 1)$ . Khối đầu tiên này có tên là khối "a".
- *Identity Block*: Hai khối *Identity Block* tiếp theo, sử dụng ba bộ lọc có kích thước tương tự  $[64, 64, 256]$ , với kích thước hạt nhân  $3 \times 3$  và không thay đổi kích thước đầu vào. Các khối này có tên là khối "b" và "c", và chúng giúp giữ nguyên kích thước không gian và tăng cường khả năng học các đặc trưng phức tạp của mạng.

### ***Giai đoạn 3: Khối Convolutional và Identity mở rộng***

Ở giai đoạn này, số lượng bộ lọc được tăng lên nhằm giúp mạng học các đặc trưng chi tiết hơn từ ảnh:

- *Convolutional Block*: Sử dụng ba bộ lọc với kích thước  $[128, 128, 512]$ , kích thước hạt nhân là  $3 \times 3$ , và sải chân  $(2 \times 2)$  nhằm giảm tiếp kích thước không gian. Khối này được đặt tên là khối "a".

- *Identity Block*: Ba khối *Identity Block* tiếp theo, cũng sử dụng các bộ lọc có kích thước [128, 128, 512], với kích thước hạt nhân  $3 \times 3$  và giữ nguyên kích thước không gian qua các khối "b", "c", và "d".

#### ***Giai đoạn 4: Tăng cường trích xuất đặc trưng***

Ở giai đoạn này, mạng tăng cường trích xuất các đặc trưng phức tạp hơn và tiếp tục giảm kích thước không gian:

- *Convolutional Block*: Ba bộ lọc với kích thước [256, 256, 1024], kích thước hạt nhân là  $3 \times 3$ , sải chân ( $2 \times 2$ ). Khối này có tên là khối "a".
- *Identity Block*: Năm *Identity Block* tiếp theo với kích thước bộ lọc [256, 256, 1024] và hạt nhân  $3 \times 3$ , giữ nguyên kích thước đầu vào trong các khối "b", "c", "d", "e", và "f".

#### ***Giai đoạn 5: Trích xuất sâu các đặc trưng cuối cùng***

Giai đoạn cuối cùng trước khi thực hiện quá trình tổng hợp và phân loại, mạng tiếp tục tăng cường trích xuất đặc trưng với các bộ lọc lớn hơn:

- *Convolutional Block*: Ba bộ lọc với kích thước [512, 512, 2048], kích thước hạt nhân là  $3 \times 3$ , và sải chân ( $2 \times 2$ ). Khối này được đặt tên là khối "a".
- *Identity Block*: Hai khối *Identity Block* với các bộ lọc [512, 512, 2048], kích thước hạt nhân  $3 \times 3$ , và được đặt tên là "b" và "c".

#### ***Giai đoạn cuối cùng: Tổng hợp và phân loại***

Sau khi đã trích xuất các đặc trưng quan trọng thông qua các giai đoạn trên, mạng thực hiện các bước cuối cùng để tổng hợp và phân loại:

- *Average Pooling 2D*: Mạng sử dụng một lớp gộp trung bình (*Average Pooling*) với cửa sổ có kích thước  $2 \times 2$ , có tên là *avg\_pool*. Lớp này giúp giảm kích thước của các đặc trưng cuối cùng mà vẫn bảo toàn thông tin quan trọng.
- *Flatten*: Sau đó, đầu ra của lớp gộp trung bình được làm phẳng thành một vector đơn, mà không cần bất kỳ tham số hoặc tên nào.
- *Fully Connected Layer*: Cuối cùng, lớp kết nối đầy đủ (*Dense Layer*) được sử dụng để giảm số chiều của vector xuống số lớp mong muốn, kết hợp với hàm kích hoạt *softmax* để tạo ra xác suất phân loại cho từng lớp.

Như vậy, mạng *ResNet50* bao gồm nhiều giai đoạn với sự kết hợp của các khối *Convolutional Block* và *Identity Block*, giúp học các đặc trưng phức tạp từ hình ảnh ở nhiều cấp độ khác nhau. Các đặc điểm nổi bật của *ResNet50* như khả năng học các đặc trưng đa cấp độ, cùng với việc khắc phục vấn đề gradient suy giảm nhờ *Residual Block*, đã giúp mạng này trở thành một trong những kiến trúc phổ biến nhất trong nhiều bài toán nhận dạng và phân loại hình ảnh phức tạp.

*ResNet* nổi bật với khả năng xử lý hiệu quả các mạng nơ-ron sâu thông qua cơ chế kết nối tắt (*skip connections*). Cơ chế này không chỉ giúp giảm thiểu vấn đề suy giảm gradient (*vanishing gradient*) mà còn cho phép mạng học được các đặc trưng sâu mà không làm mất thông tin quan trọng ở các tầng trước đó. Điều này đặc biệt hữu ích trong bài toán chẩn đoán bệnh da liễu, khi các tổn thương có thể rất phức tạp, với các mẫu hình, kết cấu và màu sắc đa dạng.

Các phiên bản *ResNet* sâu hơn, như *ResNet101* và *ResNet152*, cho phép mạng học được các đặc trưng phức tạp hơn, từ đó cải thiện độ chính xác trong việc nhận diện và phân loại các loại bệnh lý da khác nhau.

#### 1.4.2.3 Các mô hình họ *ResNet*

**ResNet50** *ResNet50* được sử dụng như một mô hình cơ sở. Với 50 lớp, *ResNet50* nổi bật nhờ các kết nối tắt (*skip connections*), cho phép duy trì gradient ổn định trong suốt quá trình huấn luyện. Điều này giúp mô hình học được các đặc trưng sâu quan trọng từ ảnh tổn thương da, ngay cả trong điều kiện dữ liệu có sự phức tạp về màu sắc và hình thái.

**ResNet101** Ngoài *ResNet50*, *ResNet101* được triển khai để tận dụng độ sâu lớn hơn với 101 lớp. So với *ResNet50*, *ResNet101* có khả năng học tốt hơn các đặc trưng phức tạp, từ đó cải thiện hiệu suất trong việc phân loại các tổn thương khó nhận biết hoặc có hình dạng bất thường. Đặc biệt, mô hình này thể hiện hiệu quả cao khi làm việc với các tổn thương da có kích thước và cấu trúc đa dạng.

**ResNet152** Trong số các mô hình *ResNet*, *ResNet152* là kiến trúc sâu nhất được sử dụng trong luận văn. Với 152 lớp, *ResNet152* mang lại khả năng học chi tiết các đặc trưng ở mức độ cao nhất, giúp tăng độ chính xác khi phân loại các loại tổn thương da khác nhau. Mặc dù độ sâu lớn có thể dẫn đến thời gian huấn luyện lâu hơn, hiệu quả phân loại vượt trội của *ResNet152* đã chứng minh tính hữu ích của nó trong các bài toán đòi hỏi độ chính xác cao như chẩn đoán bệnh da liễu.

#### 1.4.2.4 Mô hình *VGG16*

Bên cạnh các mô hình *ResNet*, một mô hình với 16 lớp tích chập, được biết đến với kiến trúc đơn giản và hiệu quả. Mặc dù không sở hữu các kết nối tắt như *ResNet*, *VGG16* vẫn giữ được

tính mạnh mẽ trong việc trích xuất các đặc trưng từ ảnh tổn thương da. Khi kết hợp với các kỹ thuật tăng cường dữ liệu (*Data Augmentation*), VGG16 đã chứng minh khả năng ổn định và hiệu quả, đặc biệt trong việc xử lý các bộ dữ liệu có kích thước hạn chế.

Mặc dù có thiết kế đơn giản hơn so với ResNet, VGG lại chứng minh sự hiệu quả nhờ tính nhất quán trong kiến trúc. Việc sử dụng các tầng tích chập nhỏ ( $3 \times 3$ ) liên tiếp cho phép VGG học được các đặc trưng hình học và kết cấu với độ chính xác cao.

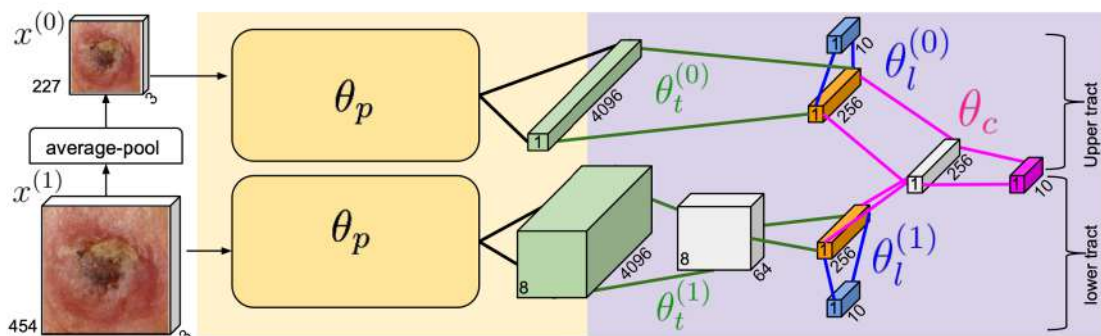
Đặc biệt, kiến trúc VGG dễ dàng được tinh chỉnh và kết hợp với các kỹ thuật tăng cường dữ liệu (*Data Augmentation*), giúp cải thiện hiệu suất trong trường hợp dữ liệu bị giới hạn hoặc có sự khác biệt về chất lượng ảnh. Trong bối cảnh bài toán chẩn đoán bệnh da liễu, điều này rất quan trọng để đảm bảo khả năng tổng quát hóa khi mô hình phải xử lý các hình ảnh tổn thương đến từ nhiều nguồn khác nhau.

#### 1.4.2.5 Lợi ích chung của ResNet và VGG

Cả ResNet và VGG đều được thiết kế để hoạt động hiệu quả trên các bộ dữ liệu lớn và phức tạp, chẳng hạn như ImageNet, và do đó, chúng cũng dễ dàng chuyển giao cho các bài toán chuyên biệt như chẩn đoán bệnh da liễu.

- **ResNet:** Mang lại lợi thế trong việc học các đặc trưng sâu và trừu tượng.
- **VGG:** Phát huy hiệu quả nhờ sự đơn giản và ổn định trong huấn luyện.

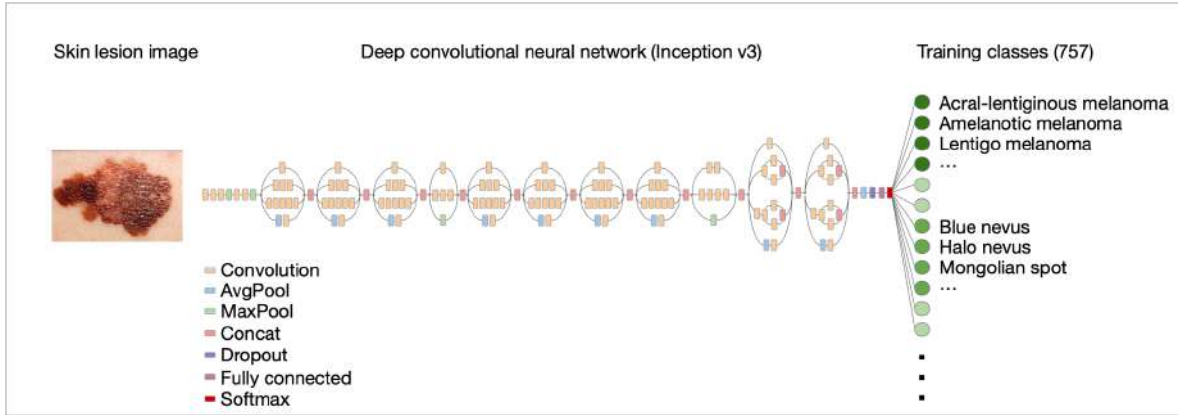
## 1.5 Các nghiên cứu ứng dụng học sâu trong chẩn đoán bệnh da liễu



Hình 1.13: Mô hình phân lớp hình ảnh bệnh da liễu với CNN của **Kawahara và cộng sự**

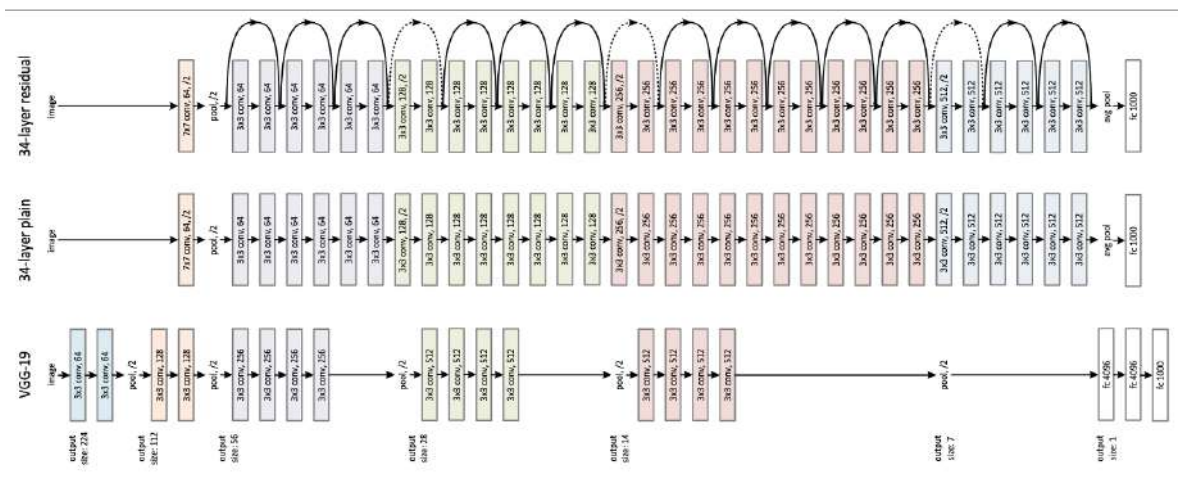
Một trong những nghiên cứu tiên phong về ứng dụng học sâu trong chẩn đoán bệnh da liễu là công trình của **Kawahara và cộng sự (2016)** [29]. Nhóm nghiên cứu đã sử dụng mạng CNN để trích xuất đặc trưng của hình ảnh da ở nhiều tỷ lệ khác nhau, tương tự như phương

pháp Overfeat (hình 1.13). CNN đã được huấn luyện và thử nghiệm trên một tập dữ liệu hình ảnh da liễu lớn, bao gồm 10 loại bệnh khác nhau. Kết quả nghiên cứu cho thấy, phương pháp này đã đạt được độ chính xác vượt trội so với các phương pháp truyền thống trước đó, cho thấy tiềm năng của CNN trong việc phân loại các bệnh da liễu phức tạp.



Hình 1.14: Mô hình phân lớp hình ảnh bệnh da liễu với CNN của **Esteva và cộng sự**

Tiếp nối thành công đó, **Esteva và cộng sự (2017)** [30] đã thực hiện một nghiên cứu có ảnh hưởng lớn, khi áp dụng một mô hình CNN cải tiến mang tên Google Inception V3 (hình 1.14). Nhóm nghiên cứu sử dụng một tập dữ liệu khổng lồ gồm 129.450 hình ảnh da để huấn luyện mô hình. Điểm nổi bật của nghiên cứu này là mô hình đã đạt được độ chính xác tương đương với các bác sĩ da liễu trong phân loại các loại u ác tính và u lành tính (đặc biệt là giữa melanoma và các tổn thương lành tính). Mô hình đạt chỉ số AUC (Area Under the Curve) trên 91%, cho thấy sự vượt trội trong việc phân loại và chẩn đoán ung thư da, một trong những loại bệnh da liễu nguy hiểm nhất.



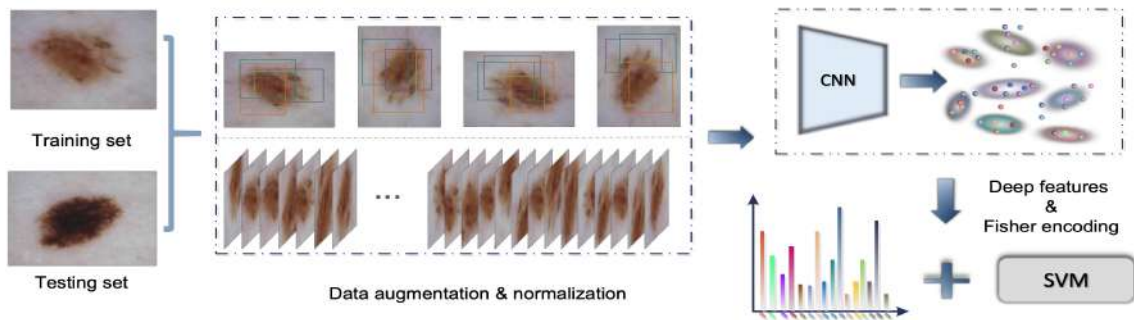
Hình 1.15: Mô hình ResNet-152

Một nghiên cứu khác được thực hiện bởi **Han và cộng sự (2018)** [31], trong đó nhóm nghiên cứu đã áp dụng phương pháp học chuyển giao (transfer learning) trên mạng Microsoft ResNet-152 (hình 1.15) để phân loại 12 loại ung thư da khác nhau. Nghiên cứu này cho thấy



rằng độ chính xác của mô hình có thể bị ảnh hưởng bởi sự khác biệt về sắc tộc của bệnh nhân, đây là một phát hiện quan trọng trong bối cảnh các hệ thống AI cần được tối ưu hóa cho nhiều nhóm đối tượng khác nhau. Điều này cũng gợi mở hướng đi mới trong việc phát triển các mô hình chẩn đoán da liễu có thể tùy chỉnh và phù hợp với các đặc điểm dân tộc học của bệnh nhân.

**Brinker và cộng sự (2019) [32]** tiếp tục củng cố tiềm năng của CNN trong chẩn đoán bệnh da liễu khi nhóm nghiên cứu sử dụng phương pháp học chuyển giao trên mạng ResNet-50 để phân loại u ác tính và nevi (nốt ruồi lành tính). Họ đã thu thập 4.204 hình ảnh sinh thiết và yêu cầu 144 bác sĩ da liễu đánh giá để so sánh độ chính xác của CNN với các chuyên gia. Kết quả cho thấy, CNN đạt độ nhạy 82,3% và độ đặc hiệu 77,9%, vượt trội hơn so với các bác sĩ da liễu cơ sở và chuyên gia. Nghiên cứu này không chỉ chứng minh khả năng vượt trội của AI trong chẩn đoán u ác tính mà còn mở ra tiềm năng ứng dụng các hệ thống AI trong hỗ trợ quyết định lâm sàng.



Hình 1.16: Mô hình của Zhen và cộng sự

Bên cạnh đó, **Zhen và cộng sự (2017) [33]** đã có những cải tiến sáng tạo trong việc kết hợp CNN với các phương pháp khác như Fisher Vector Encoding và SVM (hình 1.16). Thay vì sử dụng toàn bộ hình ảnh để làm đầu vào cho CNN, nhóm nghiên cứu đã trích xuất các vùng nhỏ trên ảnh nhằm giải quyết vấn đề thiếu hụt dữ liệu. Nghiên cứu đã đạt được độ chính xác 83,09% trên tập dữ liệu ISBI 2016, bao gồm 1.279 hình ảnh da liễu. Kết quả này cho thấy sự linh hoạt của CNN khi được kết hợp với các phương pháp tiền xử lý dữ liệu và các thuật toán máy học khác.

Ngoài ra, **Menegola và cộng sự [34]** đã thực hiện nghiên cứu phân loại ung thư da thành ba nhóm chính: ác tính, lành tính và ung thư biểu mô tế bào đáy. Kết quả cho thấy, việc chẩn đoán ung thư tế bào đáy (basal cell carcinoma) dễ dàng hơn so với các loại u ác tính khác như melanoma. Nghiên cứu này nhấn mạnh rằng, việc chẩn đoán chính xác các loại tổn thương khác nhau đòi hỏi sự phát triển của các mô hình chuyên biệt hóa cho từng loại bệnh.

Một nghiên cứu khác vào năm 2018 đã áp dụng CNN trong việc phân loại bệnh **vảy nến** với chỉ số AUC đạt 0,981, cao hơn so với 25 bác sĩ da liễu Trung Quốc [35]. Tuy nhiên, nghiên cứu này cũng chỉ ra rằng việc chẩn đoán vảy nến ở các vùng da đầu và móng tay còn nhiều hạn chế do dữ liệu không đồng nhất và chất lượng thấp. Điều này cho thấy vẫn còn

nhiều thách thức trong việc phát triển các hệ thống AI có thể xử lý hiệu quả mọi loại dữ liệu đa chiều phức tạp.

## Chương 2

# PHƯƠNG PHÁP CHẨN ĐOÁN BỆNH DA LIỄU ĐỀ XUẤT

Chương này trình bày phương pháp chẩn đoán bệnh da liễu bằng mô hình cộng tác giữa phân lớp và phân đoạn. Nội dung bao gồm việc áp dụng các mô hình tiên tiến để phân đoạn vùng tổn thương, phân lớp tổn thương, và cơ chế kết hợp giữa hai thành phần này nhằm tối ưu hóa hiệu quả chẩn đoán.

### 2.1 Cấu trúc và các thành phần của mô hình cộng tác phân lớp - phân đoạn

Phần này trình bày cấu trúc và các thành phần chính của mô hình cộng tác giữa phân lớp và phân đoạn. Hai thành phần quan trọng của hệ thống bao gồm mô hình phân đoạn, có nhiệm vụ xác định chính xác vùng tổn thương trên ảnh da, và mô hình phân lớp, chịu trách nhiệm phân loại tổn thương dựa trên các đặc trưng đã được trích xuất. Ngoài ra, phương pháp cộng tác giữa hai mô hình cũng được đề cập nhằm tối ưu hóa hiệu quả chẩn đoán. Mục tiêu của phần này là làm rõ từng thành phần của hệ thống và cách chúng được phối hợp để giải quyết bài toán chẩn đoán bệnh da liễu.

#### 2.1.1 Phân đoạn vùng tổn thương sử dụng mô hình biến thể của U-Net

##### 2.1.1.1 Cấu trúc U-Net

Như đã trình bày trong phần 1.4.1, U-Net là một kiến trúc mạng nơ-ron tích chập được thiết kế đặc biệt cho các bài toán phân đoạn ảnh, với ưu điểm nổi bật trong việc xử lý các vấn đề liên quan đến y sinh học và hình ảnh y tế. Trong phần này, chúng tôi sẽ tập trung nhấn mạnh các đặc điểm cấu trúc chính của U-Net, làm cơ sở cho các biến thể được áp dụng trong bài toán chẩn đoán bệnh da liễu.

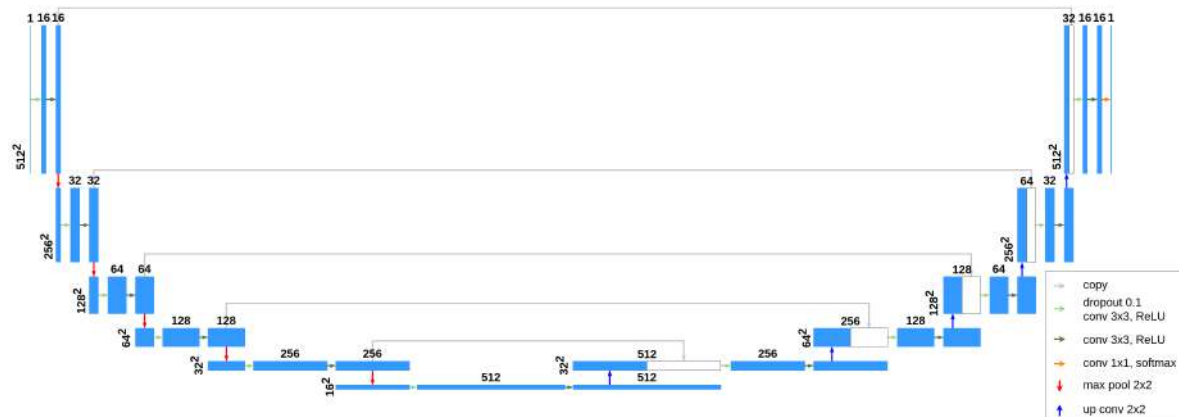
Kiến trúc U-Net bao gồm hai thành phần chính: **bộ mã hóa (encoder)** và **bộ giải mã**

(**decoder**), được kết nối bởi một cấu trúc hình chữ U thông qua các kết nối tắt (*skip connections*). Bộ mã hóa sử dụng các tầng tích chập và pooling để giảm kích thước không gian của ảnh đầu vào, đồng thời trích xuất các đặc trưng quan trọng. Ngược lại, bộ giải mã thực hiện việc mở rộng không gian của ảnh, tái dựng lại các chi tiết cần thiết để đạt được phân đoạn chính xác. Các kết nối tắt cho phép truyền tải thông tin chi tiết từ các tầng tương ứng trong bộ mã hóa sang bộ giải mã, giúp bảo toàn thông tin không gian quan trọng.

Đặc biệt, trong bài toán chẩn đoán bệnh da liễu, cấu trúc U-Net gốc đã được chứng minh là rất hiệu quả trong việc phân đoạn vùng tổn thương. Tuy nhiên, để phù hợp hơn với tính chất phức tạp của bộ dữ liệu ISIC-2018, các biến thể của U-Net, được trình bày trong 2.1.1.2, đã được áp dụng để cải thiện độ chính xác và khả năng tổng quát hóa. Việc sử dụng các biến thể này cho phép mô hình xử lý tốt hơn các tổn thương da với hình thái đa dạng và điều kiện ánh sáng khác nhau.

Phần tiếp theo sẽ tập trung vào mô tả các biến thể U-Net đã được triển khai, cùng những thay đổi cụ thể nhằm tối ưu hóa hiệu quả của mô hình.

### 2.1.1.2 Biến thể của U-Net được sử dụng



Hình 2.1: Mô hình mạng U-Net sử dụng

Trong nghiên cứu này, để giải quyết bài toán phân đoạn tổn thương da, mô hình U-Net gốc đã được điều chỉnh và triển khai với các đặc điểm phù hợp, nhằm tối ưu hóa hiệu suất phân đoạn trên bộ dữ liệu ISIC-2018. Các thay đổi và đặc điểm nổi bật của biến thể U-Net được sử dụng như sau (hình 2.1):

#### 1. Đầu vào và đầu ra:

- **Đầu vào:** Là các ảnh da liễu RGB có kích thước  $n \times n \times 3$ , nơi  $n$  (chiều rộng và chiều cao) có thể thay đổi tùy theo cấu hình dữ liệu. Kích thước này đảm bảo rằng mô hình có thể xử lý được cả các ảnh lớn lẫn các ảnh đã được nén hoặc cắt xén.

- *Đầu ra*: Là ảnh nhị phân với kích thước  $n \times n \times 1$ , trong đó mỗi điểm ảnh thể hiện xác suất thuộc vùng tổn thương. Vùng sáng trong ảnh nhị phân đại diện cho các khu vực có khả năng cao là tổn thương da.

## 2. Độ sâu của kiến trúc:

- Mô hình được triển khai với độ sâu 5, nghĩa là số tầng giảm kích thước (*max pooling*) tối đa là 5 lần, và tương ứng số tầng mở rộng (*up-convolution*) cũng là 5.
- Kiến trúc đối xứng này giúp mô hình nắm bắt được các đặc trưng ở cả cấp độ thấp (chi tiết bề mặt) và cấp độ cao (cấu trúc tổng thể) của tổn thương da. Qua thử nghiệm, độ sâu này mang lại hiệu suất tốt hơn các phiên bản U-Net có độ sâu nhỏ hơn.

## 3. Kiến trúc đối xứng và kết nối tắt:

- Kiến trúc hình chữ U đối xứng, với các kết nối tắt (*skip connections*), đảm bảo rằng thông tin không gian quan trọng từ các tầng mã hóa (encoder) được truyền thẳng tới các tầng giải mã (decoder).
- Điều này rất cần thiết để mô hình phân đoạn chính xác các chi tiết nhỏ trong tổn thương da.

## 4. Công nghệ và thư viện sử dụng:

- Biến thể U-Net này được triển khai bằng thư viện **Keras**<sup>1</sup> trên nền **TensorFlow**, giúp tối ưu hóa hiệu năng và thuận tiện trong việc mở rộng hoặc tùy chỉnh.
- Cấu trúc mô hình được thiết kế cho tác vụ *Semantic Segmentation*, tức là mỗi điểm ảnh trong ảnh đầu vào đều được gán nhãn tương ứng (thuộc hoặc không thuộc vùng tổn thương).

## 5. Tối ưu hóa và điều chỉnh đặc thù:

- Việc sử dụng đầu vào RGB thay vì ảnh đơn kênh (*grayscale*) cho phép mô hình tận dụng thông tin màu sắc và kết cấu, vốn có ý nghĩa quan trọng trong việc phân biệt tổn thương da với vùng da bình thường.
- Đầu ra dạng xác suất (ảnh nhị phân) giúp dễ dàng áp dụng các ngưỡng khác nhau để tinh chỉnh kết quả phân đoạn tùy theo yêu cầu.

Biến thể U-Net được sử dụng trong nghiên cứu này không chỉ kế thừa các ưu điểm từ mô hình gốc, mà còn được tinh chỉnh để phù hợp với tính chất phức tạp của dữ liệu ISIC-2018,

<sup>1</sup>[https://www.tensorflow.org/api\\_docs/python/tf/keras](https://www.tensorflow.org/api_docs/python/tf/keras)

nơi các tổn thương da có hình dạng và kích thước đa dạng, cùng điều kiện ánh sáng và màu sắc không đồng nhất. Những thay đổi này giúp mô hình đạt được hiệu quả vượt trội trong việc phân đoạn chính xác các vùng tổn thương.

### 2.1.2 Phân lớp tổn thương sử dụng các mô hình phổ biến của ResNet và VGG

Trong nghiên cứu này, ResNet và VGG được sử dụng với các đặc điểm chính:

- **Trọng số tiền huấn luyện (Pre-trained weights):** Trọng số của mô hình được khởi tạo từ bộ trọng số tiền huấn luyện trên tập dữ liệu ImageNet.
- **Kiến trúc cơ bản:** Loại bỏ phần *fully connected* phía trên (`include_top=False`) để thay thế bằng các lớp tùy chỉnh.
- **Kỹ thuật Global Average Pooling (GAP):** Đầu ra của mạng cơ sở được làm phẳng thông qua kỹ thuật GAP, giúp giảm số lượng tham số, tránh overfitting và duy trì các đặc trưng quan trọng nhất.

Để phù hợp với bài toán phân lớp hình ảnh, mô hình được bổ sung các lớp phía trên như sau:

- **Lớp Dropout:** Hai lớp Dropout với tỷ lệ 45% được thêm vào để giảm thiểu hiện tượng overfitting, đặc biệt khi làm việc với tập dữ liệu có kích thước hạn chế.
- **Lớp Dense (512):** Lớp *fully connected* Sử dụng hàm kích hoạt ReLU để học các đặc trưng phi tuyến từ đầu ra của mạng cơ sở.
- **Lớp đầu ra (Output layer):** Sử dụng hàm kích hoạt Softmax với số lượng neuron tương ứng với số lớp (NCLASS), đảm bảo đầu ra là một phân phối xác suất.

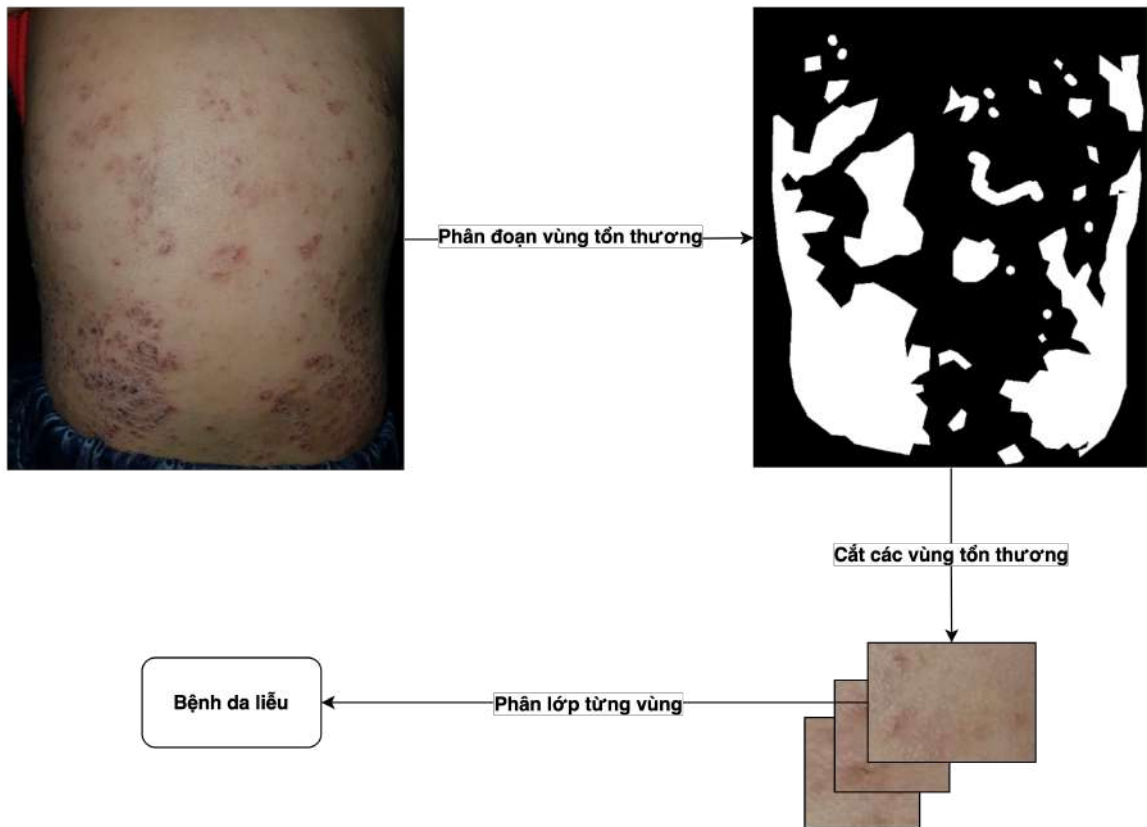
Mô hình được huấn luyện với các chiến lược sau:

- **Đóng băng trọng số (Freezing weights):** Trong một số thực nghiệm, các lớp của mạng cơ sở ResNet50V2 được đóng băng để chỉ huấn luyện các lớp phía trên. Sau đó, mô hình được *fine-tuned* bằng cách cho phép tất cả các lớp trở nên khả huấn luyện.
- **Bộ tối ưu hoá (Optimizer):** Adam được sử dụng với tốc độ học (*learning rate*) ban đầu là 0.001, đảm bảo sự cân bằng giữa tốc độ hội tụ và tránh hiện tượng nhảy quá mức trong không gian tham số.
- **Hàm mất mát (Loss function):** Cross-entropy loss được áp dụng, phù hợp với bài toán phân lớp nhiều lớp.

## 2.1.3 Cộng tác của mô hình phân lớp và phân đoạn

### 2.1.3.1 Phương pháp cộng tác

Phương pháp cộng tác giữa mô hình phân đoạn và phân lớp trong nghiên cứu này được triển khai qua hai bước chính (hình 2.2), mỗi bước tận dụng sức mạnh của một mô hình học sâu cụ thể nhằm tối ưu hóa hiệu quả chẩn đoán bệnh da liễu.



Hình 2.2: Phương pháp tổng quan

**Bước 1: Phân đoạn vùng tổn thương da bằng U-Net** Quá trình bắt đầu với việc sử dụng mô hình U-Net và các biến thể của nó để phân đoạn hình ảnh da, xác định chính xác vùng bị tổn thương. U-Net, với kiến trúc hình chữ U đối xứng và cơ chế kết nối tắt (*skip connections*), được tối ưu cho các bài toán phân đoạn hình ảnh y tế, cho phép tái hiện rõ ràng cả các đặc trưng tổng quát lẫn chi tiết của vùng tổn thương.

Kết quả của bước này là một bản đồ nhị phân, trong đó các vùng sáng biểu thị những khu vực có khả năng cao là tổn thương. Việc xác định vùng tổn thương trước khi phân lớp giúp giảm nhiễu từ các khu vực không liên quan, chẳng hạn như vùng da bình thường, nền hoặc các yếu tố môi trường trong hình ảnh.

**Bước 2: Phân lớp các vùng tổn thương đã được phân đoạn** Sau khi vùng tổn thương được phân đoạn, các vùng này được sử dụng làm đầu vào cho bước phân lớp, áp dụng các mô

hình mạng nơ-ron tích chập mạnh mẽ như ResNet-50, ResNet-101, ResNet-152 và VGG-16. Các mô hình này chịu trách nhiệm phân loại vùng tổn thương thành các loại bệnh lý da cụ thể.

### 2.1.3.2 Ưu điểm của mô hình cộng tác trong bài toán bệnh da liễu

Mô hình cộng tác giữa phân đoạn và phân lớp mang lại nhiều ưu điểm vượt trội trong bài toán chẩn đoán bệnh da liễu, nhờ vào sự kết hợp chặt chẽ giữa việc xác định chính xác vùng tổn thương và phân loại hiệu quả các loại bệnh lý. Cách tiếp cận này giải quyết đồng thời các thách thức đặc thù của dữ liệu da liễu, từ hình dạng tổn thương đa dạng đến ảnh hưởng của nhiễu nền.

**Khả năng tập trung vào vùng tổn thương da** Một trong những lợi ích lớn nhất của mô hình cộng tác là khả năng tập trung vào vùng tổn thương da. Thay vì xử lý toàn bộ hình ảnh, mô hình phân đoạn (U-Net) trước tiên xác định chính xác các khu vực bị ảnh hưởng, loại bỏ các vùng da bình thường hoặc các yếu tố không liên quan như nền và ánh sáng. Điều này không chỉ làm giảm nhiễu mà còn cải thiện độ chính xác của bước phân lớp tiếp theo. Nhờ vậy, các mô hình phân lớp như ResNet và VGG có thể tập trung vào việc trích xuất các đặc trưng quan trọng từ chính vùng tổn thương, tối ưu hóa hiệu suất chẩn đoán.

**Khai thác ưu thế của từng loại mạng học sâu** Mô hình cộng tác khai thác được ưu thế của từng loại mạng học sâu. U-Net, với kiến trúc đặc trưng cho các bài toán phân đoạn, đảm bảo rằng vùng tổn thương được xác định rõ ràng, ngay cả khi tổn thương có hình dạng phức tạp hoặc kích thước nhỏ. Trong khi đó, các mô hình phân lớp như ResNet và VGG, nhờ khả năng học sâu và trích xuất đặc trưng mạnh mẽ, giúp phân loại chính xác tổn thương da thành các loại bệnh lý cụ thể. Sự kết hợp này tạo ra một hệ thống toàn diện, phát huy thế mạnh riêng của từng mô hình.

**Tăng tính tổng quát hóa và hiệu quả xử lý dữ liệu** Hơn nữa, mô hình cộng tác giúp tăng tính tổng quát hóa và hiệu quả xử lý dữ liệu. Việc tách biệt hai bước xử lý (phân đoạn và phân lớp) làm giảm sự phụ thuộc vào việc mô hình phân lớp phải xử lý toàn bộ hình ảnh, vốn có thể chứa nhiễu. Thay vào đó, việc chỉ tập trung vào vùng tổn thương đã được phân đoạn làm tăng hiệu quả tính toán, giảm thời gian huấn luyện, đồng thời cải thiện hiệu suất trên các tập dữ liệu lớn và phức tạp như ISIC-2018.

**Phù hợp với bài toán chẩn đoán bệnh lý y học** Cuối cùng, cách tiếp cận này đặc biệt phù hợp với các bài toán chẩn đoán bệnh lý y học, nơi sự chính xác là ưu tiên hàng đầu. Khả năng cộng tác giữa phân đoạn và phân lớp không chỉ đảm bảo rằng các mô hình hoạt động hiệu quả mà còn cung cấp các kết quả rõ ràng và có thể giải thích được, điều này rất quan trọng



trong môi trường y khoa. Việc tách biệt các bước phân đoạn và phân lớp cũng giúp dễ dàng phát hiện và cải thiện những sai sót tiềm tàng trong từng giai đoạn.

**Kết luận** Nhìn chung, mô hình cộng tác trong bài toán bệnh da liễu mang lại sự cân bằng giữa độ chính xác, hiệu suất và khả năng tổng quát hóa, tạo nền tảng vững chắc cho việc phát triển các hệ thống chẩn đoán tự động trong tương lai.

## 2.2 Phương pháp triển khai và tối ưu hóa mô hình cộng tác trong chẩn đoán bệnh da liễu

Phần này tập trung vào việc triển khai thực tế mô hình cộng tác giữa phân lớp và phân đoạn trong chẩn đoán bệnh da liễu. Nội dung bao gồm các phương pháp huấn luyện mô hình, quy trình đánh giá hiệu quả, và các kỹ thuật tối ưu hóa được áp dụng nhằm cải thiện độ chính xác và hiệu năng. Qua đó, phần này không chỉ minh họa quy trình thực hiện mà còn cung cấp các giải pháp kỹ thuật để đảm bảo mô hình đạt được hiệu quả cao nhất trên bộ dữ liệu ISIC-2018.

### 2.2.1 Phương pháp huấn luyện mô hình cộng tác

Để huấn luyện mô hình cộng tác giữa phân đoạn và phân lớp trong bài toán chẩn đoán bệnh da liễu, luận văn áp dụng quy trình gồm hai giai đoạn chính: huấn luyện riêng biệt cho mô hình phân đoạn và phân lớp, sau đó kết hợp hai thành phần này thành một hệ thống hoàn chỉnh để tối ưu hóa toàn bộ quá trình chẩn đoán.

**Giai đoạn 1: Huấn luyện mô hình phân đoạn (U-Net và các biến thể)** Giai đoạn đầu tiên tập trung vào việc huấn luyện mô hình phân đoạn U-Net để xác định chính xác vùng tổn thương trên ảnh da liễu.

- **Dữ liệu đầu vào:** Các ảnh da liễu từ bộ dữ liệu ISIC-2018, đã được tiền xử lý, được sử dụng làm đầu vào. Tương ứng với mỗi ảnh là một nhãn phân đoạn (*segmentation mask*) dạng nhị phân, trong đó mỗi pixel biểu thị việc nó thuộc hoặc không thuộc vùng tổn thương.
- **Hàm mất mát:** Mô hình sử dụng hàm mất mát kết hợp giữa *binary cross-entropy* và *dice loss* (chi tiết xem tại phụ lục A.4.6.9), nhằm cân bằng giữa độ chính xác pixel và sự tương đồng về hình dạng giữa nhãn dự đoán và nhãn thực.
- **Tối ưu hóa:** Bộ tối ưu hóa Adam được sử dụng với tốc độ học (*learning rate*) được điều chỉnh thích hợp qua từng epoch.

- **Huấn luyện:** Mô hình được huấn luyện trong một số lượng epoch cố định, với dữ liệu được tăng cường bằng các kỹ thuật như xoay, lật, hoặc thay đổi độ sáng để cải thiện khả năng tổng quát hóa.

Kết quả từ mô hình phân đoạn là các ảnh nhị phân xác định vùng tổn thương, được sử dụng làm đầu vào cho giai đoạn tiếp theo.

**Giai đoạn 2: Huấn luyện mô hình phân lớp (ResNet và VGG)** Giai đoạn thứ hai huấn luyện các mô hình phân lớp trên các vùng tổn thương đã được phân đoạn từ giai đoạn trước.

- **Dữ liệu đầu vào:** Ảnh vùng tổn thương được cắt từ kết quả phân đoạn của U-Net, sau đó được chuẩn hóa và định cỡ (*resize*) để phù hợp với yêu cầu đầu vào của mô hình phân lớp (ResNet hoặc VGG).
- **Hàm mất mát:** *Cross-entropy loss* được sử dụng để đo lường độ sai lệch giữa nhãn dự đoán và nhãn thực tế, do bài toán phân lớp có bản chất đa lớp (*multi-class*).
- **Tối ưu hóa:** Bộ tối ưu hóa Adam hoặc SGD được áp dụng, với các kỹ thuật như giảm tốc độ học (*learning rate decay*) khi mô hình đạt hiệu suất cao hơn trên tập kiểm tra.
- **Huấn luyện:** Dữ liệu được tăng cường bằng các kỹ thuật như thay đổi tỷ lệ hoặc độ phân giải, để mô hình phân lớp học được các đặc trưng đa dạng từ vùng tổn thương.

**Kết hợp mô hình phân đoạn và phân lớp** Sau khi huấn luyện riêng lẻ, hai mô hình được kết hợp thành một quy trình cộng tác hoàn chỉnh. Mô hình phân đoạn U-Net hoạt động như bước tiền xử lý, xác định vùng tổn thương, và các vùng này được đưa vào mô hình phân lớp để dự đoán loại bệnh lý.

- **Chuỗi công việc:** Mô hình phân đoạn chạy trước, kết quả của nó được xử lý để làm đầu vào cho mô hình phân lớp.
- **Tối ưu hóa quy trình:** Trong giai đoạn này, các thông số của mô hình phân đoạn thường được giữ nguyên (đông cứng), chỉ tập trung tinh chỉnh mô hình phân lớp để đạt hiệu suất tối đa.

**Đánh giá và tinh chỉnh** Quy trình huấn luyện được giám sát bằng cách sử dụng các tập dữ liệu huấn luyện, kiểm tra, và kiểm định (*training, validation, testing*). Sau mỗi epoch, hiệu suất của hệ thống được đánh giá dựa trên các chỉ số như độ chính xác (*accuracy*), *F1-score*, và *Intersection over Union* (IoU) để đảm bảo sự nhất quán và độ tin cậy của mô hình.

**Kết luận** Phương pháp huấn luyện này không chỉ tận dụng được ưu điểm của từng mô hình mà còn đảm bảo sự phối hợp hiệu quả giữa các thành phần, giúp cải thiện độ chính xác và hiệu quả trong bài toán chẩn đoán bệnh da liễu.

## 2.2.2 Các kỹ thuật được sử dụng trong phân đoạn hình ảnh

### 2.2.2.1 Kỹ thuật đào tạo

Luận văn áp dụng một quy trình đào tạo chung cho tất cả các mô hình, được mô tả trong thuật toán dưới đây:

#### Thuật toán: Quy trình đào tạo mô hình

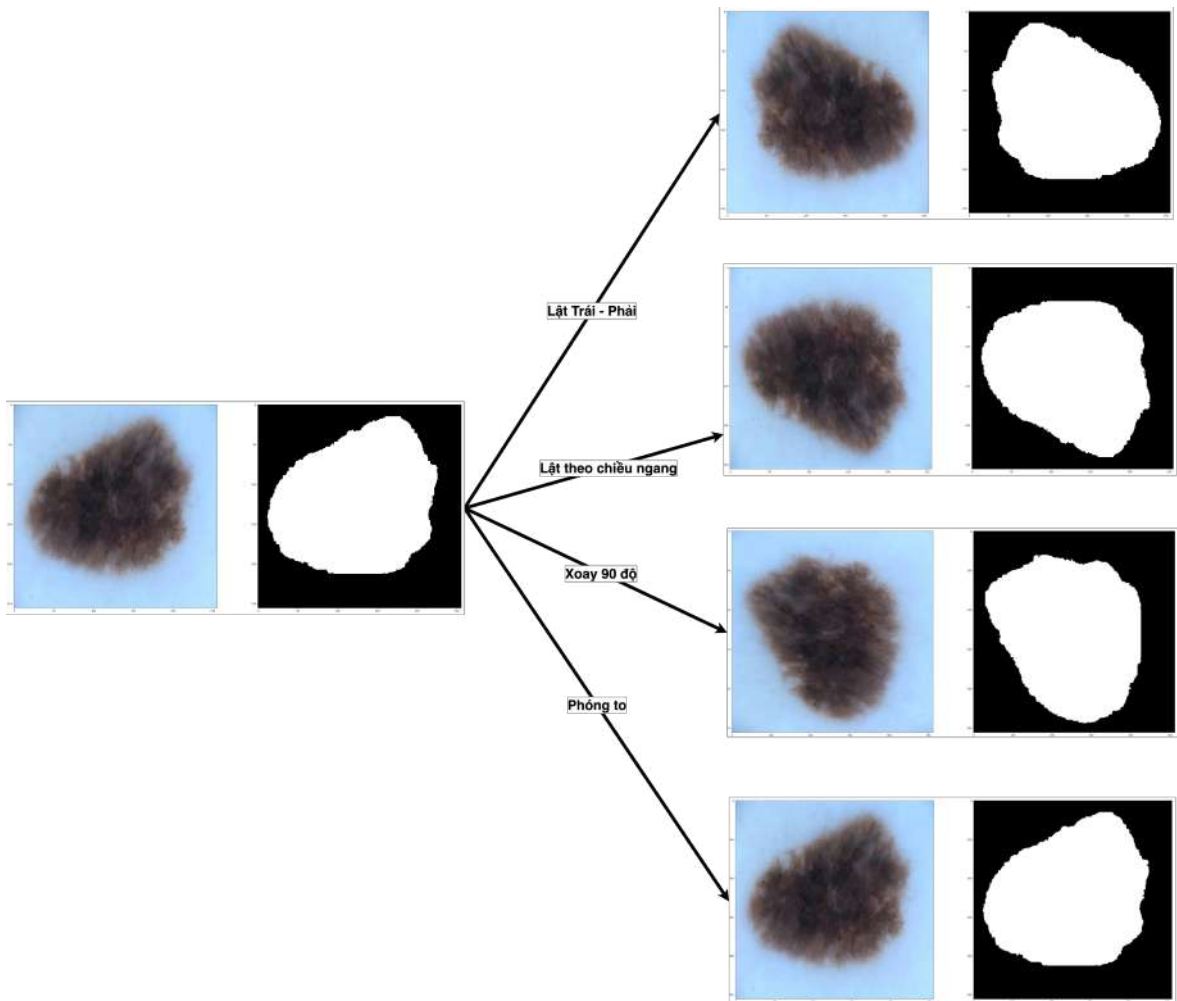
- **Khởi tạo:**  $e = 0$
- **Lặp trong khi**  $e < 2$ :
  - **Khởi tạo:**  $lr\_start = 1e-4$ ,  $lr\_end = 1e-8$
  - **Lặp trong khi**  $lr\_start > lr\_end$ :
    1. Nếu đã có trọng số (*weights*) được lưu từ trước, tải các trọng số đó vào mô hình.
    2. Sử dụng hàm *loss* phù hợp cho bài toán, với *learning rate* ban đầu là  $lr\_start$ .
    3. Huấn luyện mô hình với số lượng *epoch* là 500.
    4. Nếu sau 10 *epoch* mà giá trị *val\_loss* không giảm, áp dụng **early stopping**.
    5. Lưu lại trọng số nếu giá trị *val\_loss* hiện tại nhỏ hơn giá trị trọng số đã lưu trước đó (hoặc lưu lại ngay nếu chưa có trọng số nào được lưu).
    6. Giảm  $lr\_start$  theo hệ số 10:  $lr\_start = lr\_start/10$ .
  - Tăng  $e$  thêm 1:  $e = e + 1$ .

Mục tiêu của kỹ thuật này là đào sâu vào mô hình với learning rate giảm dần, giúp tối ưu hóa quá trình học của mô hình và tránh tình trạng overfitting.

### 2.2.2.2 Kỹ thuật tăng cường dữ liệu

Trong nghiên cứu này, các kỹ thuật tăng cường dữ liệu được sử dụng để cải thiện khả năng tổng quát hóa của mô hình học sâu, đồng thời tăng cường sự đa dạng của tập dữ liệu hình ảnh. Các phương pháp này bao gồm các phép biến đổi hình học như lật ảnh, xoay ảnh và thu nhỏ vùng quan trọng được minh họa tại hình 2.3. Cụ thể, các kỹ thuật được triển khai như sau:

- **Lật ảnh theo chiều ngang (Horizontal Flip):** Kỹ thuật này lật ảnh và mặt nạ tương ứng theo chiều ngang (trái qua phải). Phép biến đổi này giúp tạo ra các biến thể đối xứng, tăng khả năng nhận diện của mô hình đối với các hình dạng xuất hiện ở các hướng đối xứng khác nhau. Trong mã nguồn, phép lật được thực hiện bằng hàm `np.fliplr`.
- **Lật ảnh theo chiều dọc (Vertical Flip):** Tương tự như lật ngang, kỹ thuật này lật ngược ảnh và mặt nạ theo chiều dọc (trên xuống dưới). Phương pháp này bổ sung thêm



Hình 2.3: Phương pháp tăng cường dữ liệu cho bài toán phân đoạn hình ảnh

các biến thể theo chiều dọc, hỗ trợ mô hình nhận diện tốt hơn đối tượng trong các góc nhìn khác nhau. Hàm `np.flipud` được sử dụng để thực hiện phép biến đổi này.

- **Xoay ảnh 90 độ (Rotation by 90 Degrees):** Ảnh và mặt nạ được xoay một góc 90 độ ngược chiều kim đồng hồ. Phép xoay này đặc biệt hữu ích khi đặc điểm của đối tượng có thể xuất hiện ở nhiều góc độ xoay khác nhau. Hàm `np.rot90` được sử dụng để xoay ảnh và mặt nạ.
- **Thu nhỏ vùng quan trọng trong ảnh (Downsizing ROI):** Phương pháp này cắt vùng quan trọng trong ảnh, được xác định bởi mặt nạ chứa đối tượng mục tiêu, sau đó phóng to vùng này trở lại kích thước ban đầu. Vùng được chọn dựa trên hộp giới hạn (*bounding box*) tính toán từ các pixel trong mặt nạ có giá trị khác 0. Sau khi cắt, hình ảnh và mặt nạ được nội suy tuyến tính bằng hàm `cv2.resize`. Kỹ thuật này giúp tăng cường sự chú ý của mô hình vào các khu vực có chứa đối tượng quan trọng.
- **Áp dụng ngẫu nhiên (Random Application):** Các phép biến đổi trên được áp dụng với xác suất xác định trước (ví dụ: 50% cho các phép lật và xoay, 100% cho phép

thu nhỏ). Thứ tự thực hiện các phép biến đổi có thể được xáo trộn để tăng sự đa dạng của tập dữ liệu mở rộng. Quá trình này được triển khai với các hàm ngẫu nhiên `np.random.rand` và `np.random.shuffle`.

Các kỹ thuật tăng cường dữ liệu không chỉ mở rộng kích thước tập dữ liệu mà còn giúp mô hình học sâu trở nên mạnh mẽ hơn, đặc biệt trong việc xử lý các biến dạng hình học và các biến thể góc nhìn. Điều này rất quan trọng trong chẩn đoán bệnh da liễu, nơi đặc điểm tổn thương có thể xuất hiện ở nhiều vị trí và hướng khác nhau. Kết hợp các kỹ thuật này, tập dữ liệu huấn luyện được mở rộng và đa dạng hơn, giúp mô hình U-Net học được các đặc trưng phong phú hơn từ hình ảnh và giảm thiểu nguy cơ overfitting.

## 2.2.3 Các kỹ thuật được sử dụng trong phân lớp hình ảnh

### 2.2.3.1 Thuật toán huấn luyện

Các mô hình phân lớp được huấn luyện bằng thuật toán *Adam* [21] và hàm mất mát *categorical cross-entropy*. luận văn sử dụng kỹ thuật học chuyển giao (transfer learning), trong đó các mô hình được tiền huấn luyện trên tập dữ liệu ImageNet trước khi tinh chỉnh (fine-tuning) trên dữ liệu da liễu ISIC. Kỹ thuật dropout và batch normalization cũng được áp dụng để giảm thiểu hiện tượng overfitting và cải thiện hiệu suất.

### 2.2.3.2 Fine-tuning

Mô hình ResNet50 được tải sẵn trọng số từ tập dữ liệu ImageNet, sau đó được tinh chỉnh (fine-tuning) cho bài toán phân loại ảnh tổn thương da từ dữ liệu ISIC 2018 Challenge. Các lớp của mô hình ResNet50 ban đầu được cố định (frozen) để tránh việc làm mất đi các trọng số đã học từ ImageNet, sau đó các lớp mới được thêm vào để thực hiện nhiệm vụ phân loại cụ thể:

- GlobalAveragePooling2D: Lớp này chuyển đổi đầu ra của ResNet50 thành một vector đặc trưng duy nhất bằng cách tính trung bình các giá trị trên toàn bộ bản đồ đặc trưng.
- Dense (256 units, activation='relu'): Lớp này có 256 đơn vị với hàm kích hoạt ReLU [36], giúp học các đặc trưng phức tạp hơn từ vector đặc trưng.
- Dense (7 units, activation='softmax'): Lớp cuối cùng là lớp sử dụng hàm Softmax [37] với 7 đơn vị, tương ứng với 7 loại tổn thương da cần phân loại.

### 2.2.3.3 Kỹ thuật tăng cường dữ liệu cho bài toán phân lớp

Để tăng cường tính đa dạng của dữ liệu huấn luyện trong bài toán phân lớp, các kỹ thuật làm giàu dữ liệu được áp dụng nhằm cải thiện khả năng tổng quát hóa của mô hình học sâu được

đề cập tại A.4.5.3. Các phương pháp này bao gồm các phép biến đổi hình học cơ bản, cụ thể như sau:

- **Lật ảnh ngang (Horizontal Flip):** Ảnh đầu vào được lật ngẫu nhiên theo chiều ngang. Phép lật này tạo ra các biến thể đối xứng của hình ảnh, giúp mô hình nhận diện tốt hơn các đối tượng ở các hướng khác nhau.
- **Lật ảnh dọc (Vertical Flip):** Tương tự như lật ngang, ảnh được lật ngẫu nhiên theo chiều dọc. Kỹ thuật này bổ sung các biến thể đối xứng theo chiều dọc, phù hợp cho các trường hợp đối tượng có tính chất định hướng.
- **Xoay ảnh 90 độ (Rotation by 90 Degrees):** Để tăng cường khả năng nhận diện các hình dạng xuất hiện ở nhiều góc độ khác nhau, ảnh được xoay ngẫu nhiên một góc 90 độ ngược chiều kim đồng hồ.
- **Áp dụng ngẫu nhiên các phép biến đổi:** Các phép biến đổi được áp dụng theo thứ tự xáo trộn để đảm bảo sự đa dạng cao nhất trong dữ liệu huấn luyện. Thứ tự thực thi được ngẫu nhiên hóa, và mỗi ảnh chỉ áp dụng một số phép biến đổi nhất định dựa trên xác suất cấu hình trước.

Nhờ các kỹ thuật trên, tập dữ liệu phân lớp không chỉ được mở rộng về kích thước mà còn tăng cường tính phong phú trong cấu trúc và phân bố của hình ảnh, hỗ trợ mô hình học sâu học tập tốt hơn và tránh hiện tượng overfitting.

## 2.3 Kết luận

Chương này đã trình bày chi tiết về mô hình cộng tác giữa phân đoạn và phân lớp, một phương pháp tiếp cận hiệu quả trong bài toán chẩn đoán bệnh da liễu. Trước tiên, các thành phần chính của mô hình được phân tích, bao gồm mô hình phân đoạn U-Net và các biến thể của nó, cùng với các mô hình phân lớp mạnh mẽ như ResNet và VGG. Qua đó, luận văn đã chỉ ra cách mỗi thành phần đóng vai trò quan trọng trong việc cải thiện hiệu quả chẩn đoán: U-Net đảm nhận nhiệm vụ xác định vùng tổn thương, trong khi các mô hình phân lớp thực hiện phân loại dựa trên những vùng tổn thương đã được xác định.

Tiếp theo, phương pháp cộng tác giữa phân đoạn và phân lớp được xây dựng và triển khai, thể hiện sự phối hợp chặt chẽ giữa hai giai đoạn để đạt hiệu suất cao. Cách tiếp cận này không chỉ giảm thiểu tác động của nhiễu trong dữ liệu mà còn tận dụng tối đa khả năng của từng mô hình, giúp hệ thống đạt được độ chính xác cao hơn trong chẩn đoán các bệnh lý da liễu.

Cuối cùng, chương đã trình bày quy trình huấn luyện chi tiết cho cả hai thành phần mô hình, với sự áp dụng các hàm mất mát như Dice Loss để tối ưu hóa quá trình phân đoạn, cũng

như các kỹ thuật huấn luyện hiệu quả cho phân lớp. Những phương pháp này không chỉ đảm bảo tính chính xác mà còn tăng cường khả năng tổng quát hóa của mô hình khi làm việc với dữ liệu phức tạp.

Chương 2 đã xây dựng nền tảng kỹ thuật quan trọng cho việc triển khai và đánh giá hệ thống. Trong chương tiếp theo, Kết quả thực nghiệm, luận văn sẽ trình bày chi tiết về quá trình thử nghiệm, kết quả đạt được, và so sánh hiệu suất của hệ thống trên bộ dữ liệu ISIC-2018. Điều này giúp đánh giá toàn diện hiệu quả của mô hình cộng tác trong bài toán chẩn đoán bệnh da liễu.

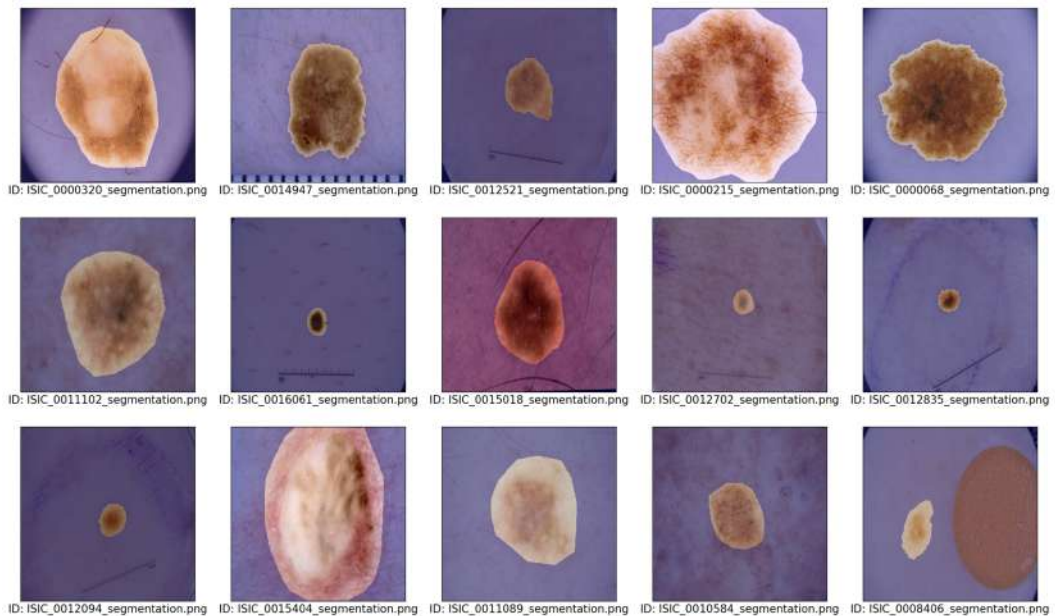
## Chương 3

# KẾT QUẢ THỰC NGHIỆM

Chương này trình bày các kết quả thực nghiệm trên bộ dữ liệu ISIC-2018 để đánh giá hiệu suất của mô hình cộng tác giữa phân đoạn và phân lớp. Nội dung bao gồm giới thiệu bộ dữ liệu, mô tả phương pháp thực nghiệm, và phân tích kết quả đạt được cho từng thành phần cũng như toàn bộ hệ thống. Thảo luận chi tiết về những điểm mạnh, hạn chế và tiềm năng cải tiến cũng được đưa ra nhằm cung cấp cái nhìn toàn diện về hiệu quả của mô hình.

### 3.1 Bộ dữ liệu sử dụng

#### 3.1.1 Tổng quan dữ liệu



Hình 3.1: Task 1 của cuộc thi ISIC-2018

Bộ dữ liệu **ISIC-2018** được phát triển cùng với cuộc thi *ISIC-2018 Challenge* [38], nhằm thúc đẩy nghiên cứu ứng dụng học sâu trong chẩn đoán bệnh da liễu. Cuộc thi này đã tạo tiền



đề cho việc phát triển và cải tiến các phương pháp phân đoạn và phân loại tổn thương da từ ảnh nội soi (dermoscopy). *ISIC-2018 Challenge* bao gồm 3 nhiệm vụ chính:

- **Task 1 - Lesion Boundary Segmentation:** Nhiệm vụ này yêu cầu các mô hình dự đoán mặt nạ phân đoạn của các tổn thương da từ các ảnh nội soi. Dữ liệu huấn luyện bao gồm 2,594 hình ảnh cùng với 12,970 mặt nạ phân đoạn tương ứng (mỗi ảnh đi kèm 5 mặt nạ ground truth). Ngoài ra, có 100 ảnh cho tập xác thực và 1,000 ảnh trong tập kiểm thử (Hình 3.1 minh họa).
- **Task 2 - Lesion Attribute Detection:** Mô hình cần phân loại và định vị các thuộc tính của tổn thương da thông qua các mặt nạ nhị phân. Các thuộc tính bao gồm: lưới (*network*), lưới âm tính (*negative network*), vân (*streaks*), nang giống milia (*milia-like cysts*), và các đốm (*dots and globules*). Dữ liệu huấn luyện bao gồm 2,594 hình ảnh, 100 ảnh để xác thực và 1,000 ảnh cho tập kiểm thử.
- **Task 3 - Disease Classification:** Đây là nhiệm vụ phân loại bệnh lý da liễu từ ảnh nội soi. Tập dữ liệu cho Task 3 chủ yếu dựa trên bộ dữ liệu nổi tiếng *HAM10000* [39] (Xem thêm tại B.1.1). Bộ dữ liệu *HAM10000* là một trong những nguồn dữ liệu lớn nhất được sử dụng cho các nghiên cứu về phân loại bệnh da liễu. Nó bao gồm 10,015 hình ảnh nội soi, mỗi ảnh được gán một nhãn thuộc một trong 7 loại bệnh da liễu phổ biến. Bộ dữ liệu này được quét bằng máy quét *Nikon Coolscan 5000 ED*, lưu dưới định dạng JPEG với độ sâu màu 8-bit, độ phân giải 300 DPI, và kích thước ban đầu là 15 × 10 cm.

**Quá trình tiền xử lý** của các hình ảnh bao gồm cắt vùng tổn thương (lesion) để tập trung vào vùng trung tâm với kích thước 800×600 pixel ở độ phân giải 72 DPI, đồng thời áp dụng các chỉnh sửa thủ công cho biểu đồ histogram nhằm cải thiện chất lượng hình ảnh.

Bộ dữ liệu *HAM10000* có dung lượng khoảng 2.9 GB và được công bố công khai để phục vụ cho các bài toán phân loại tổn thương da. Nó là một cơ sở dữ liệu quan trọng cho Task 3 của cuộc thi ISIC-2018, và là nền tảng cho việc xây dựng các mô hình học sâu trong chẩn đoán bệnh da liễu.

**Dữ liệu bổ sung** của Task 3 bao gồm 10,015 mục được nhóm lại theo tổn thương, cung cấp thông tin chi tiết về loại tổn thương và phương pháp xác nhận chẩn đoán (*diagnosis confirm type*) cho từng hình ảnh. Các thông tin này được cung cấp thông qua tệp CSV bổ sung, giúp hỗ trợ quá trình phân tích và chẩn đoán bệnh lý da.

Dữ liệu cuộc thi ISIC-2018 được cung cấp công khai cho các nhiệm vụ huấn luyện và xác thực, trong khi dữ liệu kiểm thử chỉ được công bố sau khi cuộc thi hoàn tất nhằm đảm bảo tính khách quan trong quá trình đánh giá các mô hình.

**Giấy phép sử dụng dữ liệu:**

- Dữ liệu của Task 1 và Task 2 được cung cấp theo giấy phép **CC-0**, cho phép sử dụng tự do mà không cần ghi công.
- Dữ liệu Task 3, bao gồm tập *HAM10000*, được cung cấp theo giấy phép **CC-BY-NC**, yêu cầu ghi công tác giả và chỉ sử dụng cho các mục đích phi thương mại.

**Yêu cầu trích dẫn bộ dữ liệu:** Khi sử dụng bộ dữ liệu ISIC-2018, các nghiên cứu cần trích dẫn đầy đủ các tài liệu sau:

- **HAM10000 Dataset:** (c) by ViDIR Group, Department of Dermatology, Medical University of Vienna; <https://doi.org/10.1038/sdata.2018.161>.
- **MSK Dataset:** (c) Anonymous; <https://arxiv.org/abs/1710.05006>; <https://arxiv.org/abs/1902.03368>.

Cuộc thi ISIC-2018 đã đặt nền tảng quan trọng cho sự phát triển của các hệ thống chẩn đoán bệnh da liễu tự động. Thành công của cuộc thi này đã mở đường cho việc phát triển tiếp bộ dữ liệu **ISIC-2019**, với mục tiêu tiếp tục cải thiện độ chính xác và hiệu quả trong chẩn đoán tổn thương da thông qua các kỹ thuật học sâu tiên tiến [39], [40]. Bộ dữ liệu ISIC-2019 (Xem thêm tại B.1.2) được xây dựng dựa trên nền tảng của ISIC-2018, với sự mở rộng về quy mô và độ phức tạp, góp phần tạo ra các phương pháp tiên tiến hơn trong lĩnh vực chẩn đoán bệnh lý da liễu. Một số hình ảnh về các loại bệnh trong bộ dữ liệu được mô tả như hình 3.2.

### 3.1.2 Thống kê mô tả và nhận xét dữ liệu

Bộ dữ liệu **ISIC-2018** chứa thông tin chi tiết về các nhiệm vụ phân đoạn và phân loại tổn thương da, cụ thể như sau:

**Task 1: Phân đoạn tổn thương da** Nhiệm vụ này bao gồm việc dự đoán mặt nạ phân đoạn của các vùng tổn thương trên ảnh da. Các thông tin cụ thể về dữ liệu như sau:

- **Dữ liệu huấn luyện:** Gồm 2,594 ảnh, dung lượng 10.4 GB. Có tổng cộng 12,970 vùng bệnh, trung bình mỗi ảnh chứa khoảng 5 vùng tổn thương. Dữ liệu này được cung cấp công khai.
- **Dữ liệu xác thực:** Gồm 100 ảnh với dung lượng 228 MB, cũng được công khai.
- **Dữ liệu kiểm thử:** Gồm 1,000 ảnh, dung lượng 2.2 GB, nhưng dữ liệu này không được công khai trước cuộc thi.



(a) Melanoma (MEL) - Khối u ác tính



(b) Melanocytic nevus (NV) - Tế bào hắc tố



(c) Basal cell carcinoma (BCC) - Ung thư biểu mô tế bào đáy



(d) Actinic Keratosis (AK) - Dày sừng ánh sáng



(e) Benign keratosis (BKL) - Dày sừng lành tính



(f) Dermatofibroma (DF) - U xơ da



(g) Vascular Lesion (VASC) - Tổn thương mạch máu

Hình 3.2: Một số ví dụ về các bệnh da liễu trong ISIC-2018

**Task 2: Phát hiện thuộc tính tổn thương** Nhiệm vụ này yêu cầu các mô hình vừa phân đoạn vừa phân loại các thuộc tính tổn thương. Dữ liệu sử dụng trong Task 2 là dữ liệu chung với Task 1, bao gồm các ảnh và mặt nạ phân đoạn nhưng có thêm yêu cầu phân loại thuộc tính cụ thể của vùng tổn thương.

**Task 3: Phân loại bệnh lý da** Task này sử dụng bộ dữ liệu *HAM10000*, một tập dữ liệu chứa các ảnh nội soi da đã được gán nhãn chi tiết về các loại bệnh lý da khác nhau. Bộ dữ liệu *HAM10000* chứa 7 nhãn phân loại chính:

- **ak**: Actinic Keratoses (Dày sừng quang hóa) - 327 ảnh
- **bcc**: Basal cell carcinoma (Ung thư biểu mô tế bào đáy) - 514 ảnh
- **bkl**: Benign keratosis (Dày sừng lành tính) - 1,099 ảnh
- **df**: Dermatofibroma (U xơ da) - 115 ảnh
- **nv**: Melanocytic nevus (Bớt hắc tố) - 6,705 ảnh
- **mel**: Melanoma (Ung thư da ác tính) - 1,113 ảnh
- **vasc**: Vascular skin lesions (Tổn thương mạch máu) - 142 ảnh

Thông tin dữ liệu cho nhiệm vụ phân loại bao gồm:

- **Dữ liệu huấn luyện**: Gồm 10,000 ảnh từ bộ dữ liệu *HAM10000*, được sử dụng để huấn luyện các mô hình phân loại bệnh lý.
- **Dữ liệu xác thực**: Gồm 200 ảnh với dung lượng 51 MB, dữ liệu này được công khai.
- **Dữ liệu kiểm thử**: Gồm 1,512 ảnh với dung lượng 401 MB, nhưng dữ liệu này không được công khai trước cuộc thi.

## 3.2 Thiết lập thực nghiệm

### 3.2.1 Tách bộ dữ liệu

Trong nghiên cứu này, luận văn sử dụng các tách dữ liệu được cung cấp sẵn bởi bộ dữ liệu *ISIC-2018*. Bộ dữ liệu *ISIC-2018* đã được tổ chức thành các tập con bao gồm tập huấn luyện (*training set*), tập kiểm tra (*validation set*), và tập kiểm định (*test set*) với mục đích phục vụ các nhiệm vụ như phân đoạn và phân lớp hình ảnh tổn thương da.

## Cấu trúc tách dữ liệu

### 1. Phân đoạn hình ảnh:

- Dữ liệu phân đoạn bao gồm các ảnh da liễu (*input images*) và nhãn phân đoạn tương ứng (*segmentation masks*).
- Bộ dữ liệu được chia thành:
  - **Tập huấn luyện:** 2594 ảnh.
  - **Tập kiểm tra:** 100 ảnh.
  - **Tập kiểm định:** 1000 ảnh.

### 2. Phân lớp hình ảnh:

- Dữ liệu phân lớp gồm các ảnh da liễu gắn nhãn thuộc một trong bảy loại bệnh lý da.
- Bộ dữ liệu được tổ chức như sau:
  - **Tập huấn luyện:** 10000 ảnh.
  - **Tập kiểm tra:** 200 ảnh.
  - **Tập kiểm định:** 1512 ảnh.

**Tiền xử lý liên quan** Mặc dù sử dụng các tách dữ liệu sẵn có, các bước tiền xử lý như chuẩn hóa (*normalization*), thay đổi kích thước ảnh (*resize*), và áp dụng tăng cường dữ liệu (*data augmentation*) vẫn được thực hiện trước khi đưa vào mô hình. Điều này đảm bảo rằng dữ liệu đầu vào phù hợp với yêu cầu của từng mô hình học sâu và tăng cường khả năng tổng quát hóa của mô hình.

### 3.2.2 Cấu hình phần cứng và môi trường huấn luyện

Để đảm bảo quá trình huấn luyện mô hình diễn ra hiệu quả và đạt hiệu suất tốt nhất, toàn bộ thí nghiệm trong nghiên cứu này được thực hiện trên hệ thống máy tính có cấu hình phần cứng mạnh mẽ, bao gồm các thành phần chính như sau:

- **VGA (Card đồ họa):** ASUS TUF Gaming GeForce RTX 4070 Ti SUPER 16GB GDDR6X - Đây là card đồ họa với hiệu suất cao, hỗ trợ quá trình huấn luyện các mô hình học sâu yêu cầu tính toán GPU với dung lượng bộ nhớ lớn.
- **CPU:** AMD Ryzen 7 7800X3D, xung nhịp cơ bản 4.2GHz và tối đa 5.0GHz, với 8 nhân và 16 luồng xử lý, phù hợp cho các tác vụ tính toán đa luồng và xử lý dữ liệu lớn.

- **RAM:** Adata XPG Lancer 32GB (2x16GB) DDR5, tốc độ 6000Mhz - Dung lượng RAM đủ lớn để xử lý các mô hình phức tạp và đảm bảo tốc độ truy xuất dữ liệu nhanh chóng.
- **Mainboard:** MSI PRO B650M-A WIFI DDR5 - Hỗ trợ tốt cho CPU và RAM với chuẩn DDR5, đảm bảo hệ thống hoạt động ổn định.
- **Ổ cứng HDD:** Seagate Barracuda 2TB 7200RPM, SATA3 6Gb/s - Dung lượng lớn giúp lưu trữ dữ liệu và các kết quả huấn luyện.
- **Ổ cứng SSD:** Kioxia Exceria Plus G3 Gen 4x4 1TB, tốc độ đọc/ghi: 5000/3900MB/s - Ổ cứng SSD tốc độ cao giúp cải thiện thời gian truy xuất dữ liệu trong quá trình huấn luyện và lưu trữ mô hình.

### 3.2.3 Cấu hình huấn luyện mô hình

Các mô hình học sâu trong nghiên cứu này được huấn luyện với các thông số và cấu hình như sau:

- **Optimizer:** Bộ tối ưu hoá Adam được sử dụng để cập nhật các tham số của mô hình trong quá trình huấn luyện, với **learning rate** ban đầu được đặt là 0.001. Adam là một thuật toán tối ưu hoá hiệu quả, thường được sử dụng trong các mô hình học sâu nhờ khả năng xử lý tốt cả các bài toán đòi hỏi tính toán phức tạp và dữ liệu lớn.
- **Hàm mất mát (Loss function):**
  - **Cross-Entropy Loss** được sử dụng cho nhiệm vụ phân lớp, giúp đo lường sự khác biệt giữa nhãn thực tế và nhãn dự đoán của mô hình.
  - **Dice Loss** được áp dụng cho nhiệm vụ phân đoạn, đặc biệt phù hợp với bài toán phân đoạn ảnh y học vì nó tập trung tối ưu hoá việc dự đoán các vùng tổn thương, cân bằng tốt giữa Precision và Recall.
- **Cấu hình của các phương pháp làm giàu dữ liệu (Data Augmentation):** Để tăng cường tính đa dạng và khả năng tổng quát của dữ liệu đầu vào, các phương pháp làm giàu dữ liệu được triển khai với các cấu hình như sau:
  - **Lật ảnh ngang (Horizontal Flip):** Ảnh và mặt nạ được lật ngẫu nhiên theo chiều ngang với xác suất  $p = 0.5$ .
  - **Lật ảnh dọc (Vertical Flip):** Ảnh và mặt nạ được lật ngẫu nhiên theo chiều dọc với xác suất  $p = 0.5$ .
  - **Xoay ảnh 90 độ (Rotation by 90 Degrees):** Cả ảnh và mặt nạ được xoay ngẫu nhiên một góc 90 độ ngược chiều kim đồng hồ với xác suất  $p = 0.5$ .

- **Thu nhỏ vùng quan trọng (Downsizing ROI):** Đối với mỗi ảnh, vùng chứa tổn thương được xác định dựa trên mặt nạ, sau đó được cắt và phóng to trở lại kích thước ban đầu. Tỷ lệ cắt được điều chỉnh ngẫu nhiên trong khoảng 1% đến 5% của kích thước ảnh. Phép biến đổi này được áp dụng với xác suất  $p = 1.0$ .
- **Áp dụng ngẫu nhiên các phép biến đổi:** Các phép biến đổi được áp dụng ngẫu nhiên theo một thứ tự xáo trộn. Việc áp dụng ngẫu nhiên đảm bảo rằng mỗi ảnh trong tập dữ liệu đầu vào sẽ được biến đổi theo cách khác nhau, tăng cường tính đa dạng.

Các cấu hình trên được lựa chọn để cân bằng giữa việc làm giàu dữ liệu và giữ nguyên các đặc điểm quan trọng của tổn thương, đảm bảo rằng dữ liệu tăng cường không làm mất đi tính chính xác của thông tin y học.

### 3.3 Kết quả thực nghiệm

#### 3.3.1 Kết quả phân đoạn hình ảnh

##### 3.3.1.1 Hiệu suất tổng quan của mô hình

Mô hình U-Net được huấn luyện với tập dữ liệu bao gồm 2,594 ảnh cho tập huấn luyện và 100 ảnh cho tập kiểm thử, mỗi ảnh có kích thước được chuẩn hóa là 256x256 pixel. Việc huấn luyện mô hình được thực hiện với 500 *epochs*, sử dụng *batch size* là 60. Quá trình huấn luyện không chỉ dựa vào việc tối ưu hóa độ chính xác phân loại mà còn tập trung vào việc cải thiện khả năng phân đoạn chính xác vùng tổn thương da.

Các siêu tham số chính của mô hình bao gồm:

- **Kích thước ảnh:** 256x256
- **Số epochs:** 500
- **Batch size:** 60
- **Số lượng lớp:** 2 (Normal - Da bình thường, Abnormal - Da có tổn thương)
- **Optimizer:** Adam với learning rate khởi tạo ban đầu là 0.0001

Các chỉ số chính được ghi nhận sau quá trình huấn luyện trên tập kiểm thử bao gồm:

- **Loss trên tập kiểm thử:** 0.0969
- **Độ chính xác (accuracy):** 95.73%
- **Dice Coefficient:** 0.9359

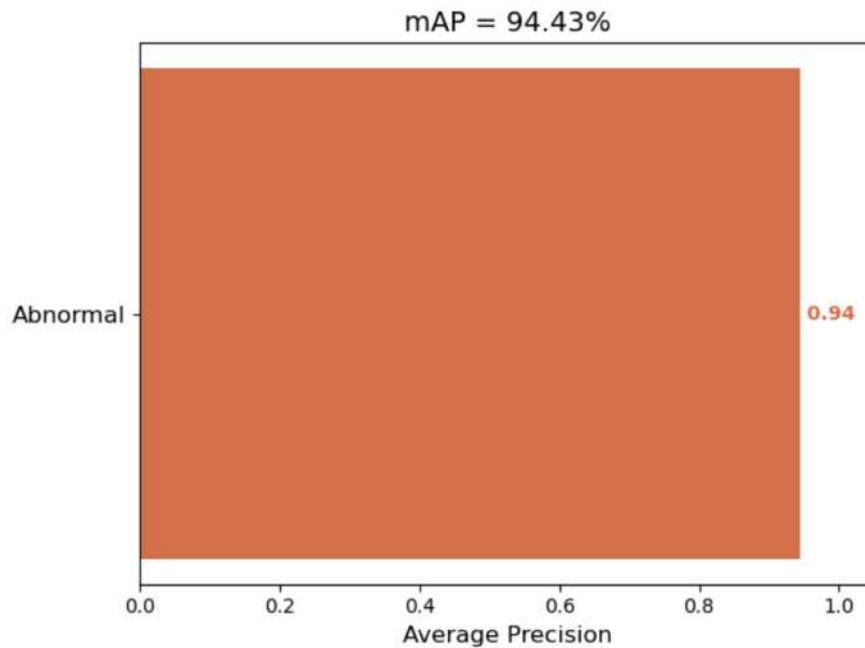
- **Soft Dice Coefficient:** 0.9171
- **Log-Average Miss Rate (LAMR - Phụ lục A.4.7.2):** 0.07

Các chỉ số này cho thấy rằng mô hình U-Net có khả năng phân đoạn hiệu quả các vùng da bất thường với độ chính xác cao, đặc biệt với Dice Coefficient đạt trên 93.59%. Chỉ số **Log-Average Miss Rate (LAMR)** đạt 0.07 cũng phản ánh khả năng giảm thiểu sai sót trong phân đoạn, cho thấy mô hình có khả năng nhận diện chính xác phần lớn các vùng tổn thương và tránh bỏ sót nhiều vùng tổn thương quan trọng.

### 3.3.1.2 Đánh giá chi tiết thông qua các chỉ số mAP, Precision và Recall

Để đánh giá tổng quan hơn về hiệu suất của mô hình trong việc phân đoạn các vùng bất thường, chỉ số **Mean Average Precision (mAP)** (xem thêm tại A.4.7.1) đã được sử dụng. Chỉ số mAP không chỉ đo lường độ chính xác mà còn giúp đánh giá hiệu suất tổng quát của mô hình trong việc phát hiện chính xác các vùng tổn thương da từ ảnh.

Kết quả tính toán mAP cho lớp **Abnormal** (vùng da có tổn thương) cho thấy rằng: **Mean Average Precision (mAP)** là 94.43% (hình 3.3)



Hình 3.3: Biểu đồ chỉ số mAP của mô hình

Kết quả mAP đạt 94.43% cho thấy khả năng phát hiện chính xác và chi tiết của mô hình đối với các vùng da có dấu hiệu tổn thương. Chỉ số này đánh giá mức độ mô hình có thể tìm kiếm và phân đoạn đúng các vùng tổn thương trong ảnh đầu vào, đồng thời hạn chế các dự đoán sai (false positives).



### ***Precision và Recall theo từng mức độ phân đoạn***

Chỉ số **Precision** và **Recall** theo từng giai đoạn của phân đoạn ảnh cho lớp **Abnormal** cũng được ghi nhận để cung cấp cái nhìn chi tiết hơn về hiệu suất của mô hình:

- **Precision:** Ban đầu đạt giá trị 1.00 ở nhiều trường hợp nhưng giảm dần khi số lượng đối tượng cần nhận dạng tăng lên, dao động từ 0.86 đến 1.00. Điều này cho thấy mô hình có xu hướng nhận dạng chính xác vùng tổn thương ở các ảnh đơn giản hoặc ít nhiễu, trong khi với các ảnh phức tạp hơn, độ chính xác có thể giảm xuống.
- **Recall:** Bắt đầu từ giá trị thấp 0.01 ở những bước đầu của quá trình phân đoạn nhưng tăng dần và đạt mức 0.95 ở các bước cuối cùng. Điều này cho thấy mô hình có khả năng nhận diện hầu hết các đối tượng bất thường khi số lượng ảnh trong tập kiểm thử tăng lên và mô hình càng được huấn luyện lâu hơn.

### ***Phân tích chi tiết về số lượng đối tượng nhận dạng***

Bên cạnh đó, mô hình cũng được đánh giá dựa trên số lượng đối tượng bất thường được nhận dạng so với số lượng thực tế trong tập kiểm thử. Kết quả phân tích cho thấy:

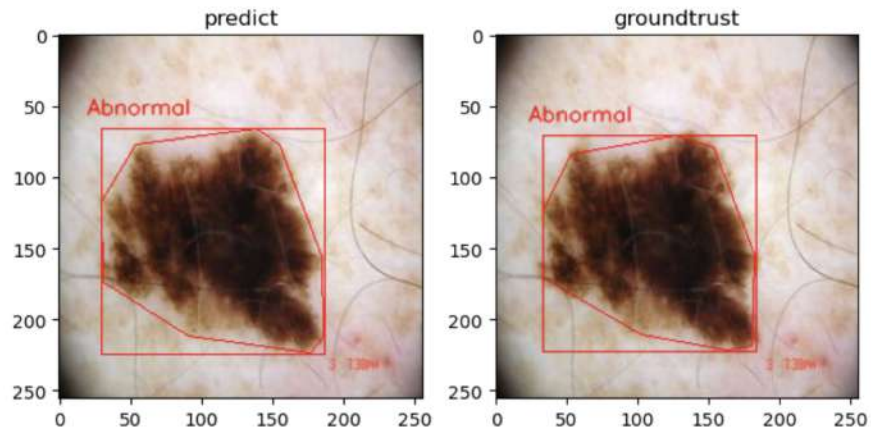
- **Tổng số đối tượng thực tế trong tập kiểm thử:** 100 (Abnormal)
- **Số đối tượng được mô hình phát hiện:** 111
  - **Số đối tượng phát hiện đúng (True Positives):** 95
  - **Số đối tượng phát hiện sai (False Positives):** 16

Những kết quả này phản ánh rằng mô hình có khả năng phát hiện hầu hết các đối tượng tổn thương da, tuy nhiên vẫn có một số trường hợp dự đoán nhầm. Số lượng *false positives* chiếm 14.4%, cho thấy mô hình cần được cải thiện thêm để giảm thiểu các lỗi phân đoạn không cần thiết.

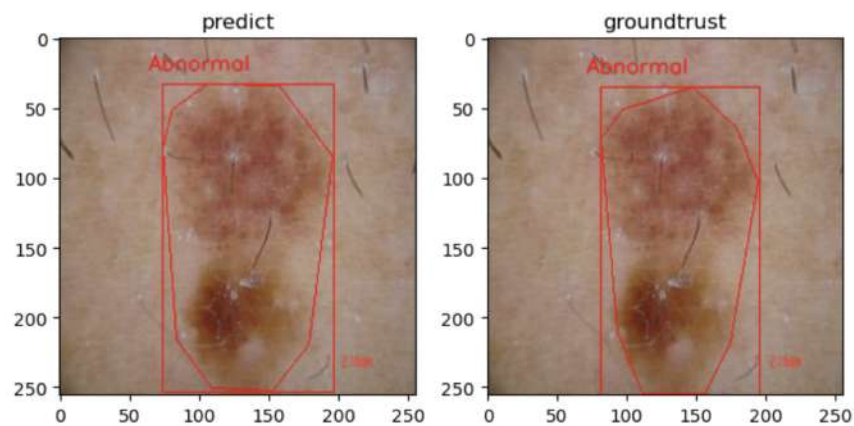
#### ***3.3.1.3 Trực quan hóa kết quả phân đoạn***

Kết quả phân đoạn của mô hình U-Net được trực quan hóa để đánh giá khả năng dự đoán các vùng bất thường. Hình ảnh đầu vào, kết quả phân đoạn và mặt nạ (*mask*) *ground truth* đã được so sánh trực tiếp để cung cấp cái nhìn rõ ràng hơn về hiệu suất mô hình.

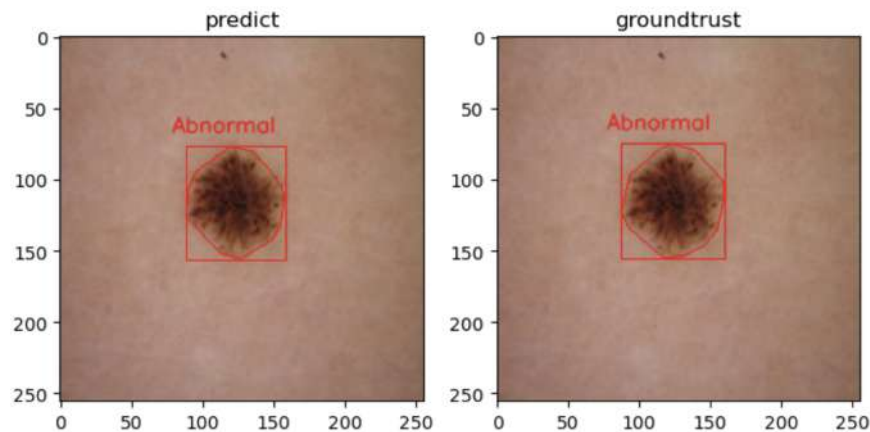
Các hình ảnh minh họa (Hình 3.4) cho thấy mô hình có thể xác định chính xác các vùng tổn thương với độ chi tiết cao. Kết quả trực quan hóa cũng cho thấy rằng, trong hầu hết các trường hợp, các biên giới của vùng tổn thương được mô hình phân đoạn rất sát với thực tế, điều này phản ánh khả năng phát hiện tốt và chính xác của mô hình U-Net.



(a) Ảnh thử nghiệm 1



(b) Ảnh thử nghiệm 2



(c) Ảnh thử nghiệm 3

Hình 3.4: Trực quan hóa kết quả phân đoạn cho một số ảnh trong tập kiểm thử.

#### 3.3.1.4 Thảo luận về hiệu suất mô hình

Mô hình U-Net được đề xuất trong nghiên cứu đã thể hiện hiệu suất cao trong việc phân đoạn các vùng da bất thường từ ảnh da liễu. Với chỉ số mAP đạt **94.43%**, Dice coefficient trên **93.59%**, chỉ số **Log-Average Miss Rate (LAMR)** đạt **0.07**, và các chỉ số *Precision*, *Recall* tốt, có thể thấy rằng mô hình này có tiềm năng lớn trong ứng dụng phân tích ảnh y học tự

động.

Tuy nhiên, vẫn còn tồn tại một số hạn chế, bao gồm tỷ lệ *false positives* vẫn tương đối cao (16/111). Điều này có thể làm giảm độ tin cậy trong các ứng dụng thực tế. Việc cải tiến mô hình hoặc sử dụng các biện pháp hậu xử lý (*post-processing*) để giảm thiểu các dự đoán sai có thể giúp cải thiện hiệu suất tổng thể.

Kết luận, kết quả phân đoạn của mô hình cho thấy tiềm năng ứng dụng lớn trong việc tự động phát hiện và chẩn đoán các bệnh lý về da thông qua ảnh y học, với hiệu suất cao và khả năng phân đoạn chi tiết, chính xác các vùng tổn thương.

### 3.3.2 Kết quả phân lớp hình ảnh

#### 3.3.2.1 Kết quả chạy thực nghiệm khi sử dụng phương pháp tăng cường dữ liệu

Luận văn sử dụng mô hình cơ sở (basemodel) là ResNet. Hàm `preprocessing_function` được chọn là `tf.keras.applications.resnet.preprocess_input` bởi vì trong quá trình thực nghiệm, hàm này đã chứng tỏ hiệu quả trong việc chuẩn hóa đầu vào, giúp mô hình học tốt hơn từ dữ liệu hình ảnh. Cụ thể, hàm `preprocess_input` thực hiện các thao tác chuẩn hóa bằng cách đưa dữ liệu về cùng một dải giá trị, phù hợp với mô hình ResNet V2 đã được huấn luyện trên tập dữ liệu ImageNet. Điều này giúp tăng cường khả năng khái quát hóa của mô hình, đảm bảo hiệu suất tốt hơn khi áp dụng trên tập dữ liệu thử nghiệm.

Hàm `preprocess_input` này không chỉ được áp dụng cho mô hình ResNet V2 mà còn được sử dụng cho mọi mô hình thử nghiệm khác trong luận văn, bao gồm các phiên bản ResNet khác và VGG16. Điều này giúp đảm bảo tính nhất quán trong quy trình xử lý dữ liệu và tăng tính so sánh giữa các mô hình khi áp dụng các kỹ thuật tăng cường dữ liệu (Data Augmentation). Các kết quả thực nghiệm về việc sử dụng các kỹ thuật này trên các mô hình ResNet và VGG16 được thể hiện như sau:

Các kỹ thuật Data Augmentation				Accuracy
<code>rotation_range</code>	<code>horizontal_flip</code>	<code>vertical_flip</code>	<code>preprocessing_function</code>	
90	True	True	True	<b>0.725</b>
90	True		True	0.724
90		True	True	0.712
90			True	0.724
			True	<b>0.749</b>

Bảng 3.1: Kết quả tăng cường dữ liệu với mô hình ResNet50

**Nhận xét:** Bảng 3.1 trình bày kết quả khi áp dụng các kỹ thuật tăng cường dữ liệu (Data Augmentation) cho mô hình **ResNet50**. Kết quả tốt nhất đạt **0.749** khi cả hai kỹ thuật lật ngang (`horizontal_flip`) và lật dọc (`vertical_flip`) được áp dụng, cùng với `rotation_range`

= 90. Khi chỉ sử dụng một kỹ thuật lật hoặc không sử dụng lật ảnh, độ chính xác giảm nhẹ, đạt mức thấp nhất là **0.712** khi chỉ dùng `vertical_flip`.

Các kỹ thuật Data Augmentation				Accuracy
<code>rotation_range</code>	<code>horizontal_flip</code>	<code>vertical_flip</code>	<code>preprocessing_function</code>	
90	True	True	True	0.730
90	True		True	<b>0.737</b>
90		True	True	<b>0.753</b>
90			True	0.735
			True	<b>0.754</b>

Bảng 3.2: Kết quả tăng cường dữ liệu với mô hình ResNet101

**Nhận xét:** Trong bảng 3.2, mô hình **ResNet101** đạt độ chính xác cao nhất **0.754** khi không sử dụng `rotation_range`, chỉ áp dụng các phép lật ảnh. Kết quả cũng cho thấy rằng khi sử dụng thêm phép xoay, độ chính xác giảm nhẹ, mặc dù vẫn giữ ở mức khá cao với **0.753**.

Các kỹ thuật Data Augmentation				Accuracy
<code>rotation_range</code>	<code>horizontal_flip</code>	<code>vertical_flip</code>	<code>preprocessing_function</code>	
90	True	True	True	0.718
90	True		True	0.721
90		True	True	<b>0.728</b>
90			True	0.728
			True	<b>0.737</b>

Bảng 3.3: Kết quả tăng cường dữ liệu với mô hình ResNet152

**Nhận xét:** Từ bảng 3.3, mô hình **ResNet152** đạt độ chính xác tốt nhất là **0.737** khi không sử dụng phép xoay ảnh và chỉ áp dụng các phép lật ảnh. Khi kết hợp cả lật ảnh và xoay ảnh, kết quả độ chính xác giảm xuống còn **0.728**, chứng tỏ sự nhạy cảm của mô hình đối với việc tăng cường dữ liệu.

Các kỹ thuật Data Augmentation				Accuracy
<code>rotation_range</code>	<code>horizontal_flip</code>	<code>vertical_flip</code>	<code>preprocessing_function</code>	
90	True	True	True	0.722
90	True		True	0.714
90		True	True	0.714
90			True	0.728
			True	0.707

Bảng 3.4: Kết quả tăng cường dữ liệu với mô hình VGG16

**Nhận xét:** Bảng 3.4 cho thấy mô hình **VGG16** đạt kết quả cao nhất với độ chính xác **0.728** khi chỉ sử dụng các phép lật ảnh mà không áp dụng phép xoay. Khi sử dụng thêm phép xoay (`rotation_range`), kết quả giảm nhẹ, đạt mức **0.707**.

### ***Kết luận***

Qua các bảng từ 3.1 đến 3.4, có thể thấy rằng các kỹ thuật tăng cường dữ liệu (Data Augmentation) có tác động đáng kể đến hiệu suất của các mô hình học sâu, đặc biệt là các mô hình như **ResNet50**, **ResNet101**, **ResNet152**, và **VGG16**. Cụ thể:

- **ResNet101** đạt hiệu suất tốt nhất với độ chính xác cao nhất **0.754** khi chỉ sử dụng các phép lật ảnh mà không áp dụng xoay.
- Mô hình **VGG16** và **ResNet152** cũng đạt kết quả tốt khi không áp dụng quá nhiều kỹ thuật tăng cường dữ liệu cùng lúc.
- Việc kết hợp các kỹ thuật lật ảnh và xoay ảnh không luôn mang lại kết quả tốt hơn, cho thấy mỗi mô hình có độ nhạy cảm riêng với các phương pháp tăng cường dữ liệu.

Từ các kết quả này, luận văn có thể rút ra rằng việc lựa chọn các kỹ thuật tăng cường dữ liệu phù hợp với từng mô hình là rất quan trọng để cải thiện hiệu suất phân loại hình ảnh da liễu. Trong các thử nghiệm tiếp theo, việc tối ưu hóa các kỹ thuật tăng cường dữ liệu và thử nghiệm với nhiều cấu hình khác nhau sẽ là hướng đi tiềm năng nhằm nâng cao hiệu quả của các mô hình học sâu trong chẩn đoán bệnh da liễu.

#### ***3.3.2.2 Nhận xét và đánh giá bảng kết quả Dropout cho các mô hình ResNet và VGG16***

<b>Dropout (<math>p</math>)</b>	<b>ResNet50</b>	<b>ResNet101</b>	<b>ResNet152</b>	<b>VGG16</b>	<b>ResNet50V2</b>	<b>ResNet101V2</b>	<b>ResNet152V2</b>
0.05	0.722	0.749	0.742	0.713	0.708	0.727	0.711
0.075	0.743	0.740	0.737	0.709	-	-	-
0.1	<b>0.747</b>	0.737	0.736	0.710	0.695	<b>0.743</b>	0.709
0.15	0.741	0.733	0.717	0.718	0.707	0.720	0.710
0.2	0.739	<b>0.78</b>	<b>0.751</b>	<b>0.718</b>	0.714	0.722	0.715
0.25	-	<b>0.78</b>	0.726	<b>0.731</b>	0.708	0.722	<b>0.720</b>
0.3	-	0.755	0.740	0.715	0.704	0.723	<b>0.726</b>

Bảng 3.5: Kết quả Dropout cho các mô hình ResNet và VGG16 (Các giá trị nổi bật được in đậm)

Bảng 3.5 trình bày kết quả thử nghiệm hiệu suất (Accuracy) của các mô hình ResNet và VGG16 khi áp dụng các giá trị khác nhau của Dropout. Dưới đây là các nhận xét và đánh giá chi tiết cho từng mô hình:

**1. ResNet50:** Kết quả tốt nhất cho mô hình ResNet50 đạt được khi sử dụng **Dropout = 0.1** với độ chính xác cao nhất là **0.747**. Khi giá trị Dropout nhỏ hơn (0.05) hoặc lớn hơn (0.15, 0.2), hiệu suất giảm dần. Điều này cho thấy rằng việc sử dụng giá trị Dropout vừa phải (0.1) giúp mô hình tránh được overfitting mà vẫn giữ được hiệu suất cao.

**2. ResNet101:** Mô hình ResNet101 đạt hiệu suất tốt nhất khi sử dụng **Dropout = 0.2** và **0.25**, với độ chính xác cao nhất là **0.78**. Ở các giá trị Dropout thấp hơn (0.1, 0.15), hiệu suất mô hình giảm, nhưng vẫn duy trì ở mức tốt (**0.737 - 0.740**). Kết quả này cho thấy ResNet101 hoạt động tốt hơn khi có mức Dropout cao hơn, giúp giảm overfitting.

**3. ResNet152:** Hiệu suất cao nhất của ResNet152 đạt được với **Dropout = 0.2**, với độ chính xác **0.751**. Hiệu suất ở các giá trị Dropout khác (0.05, 0.075, 0.1) vẫn duy trì ở mức tốt, nhưng thấp hơn một chút so với khi Dropout = 0.2. Giá trị này cho thấy rằng mô hình ResNet152 cũng hưởng lợi từ mức Dropout trung bình đến cao.

**4. VGG16:** Với mô hình VGG16, hiệu suất tốt nhất đạt được khi sử dụng **Dropout = 0.15** và **0.2**, với độ chính xác **0.718**. Khi sử dụng Dropout thấp hơn hoặc cao hơn (0.05, 0.075, 0.3), hiệu suất giảm, chứng tỏ mô hình VGG16 yêu cầu một mức Dropout nhất định để tránh overfitting nhưng cũng không nên quá cao.

**5. ResNet50V2:** Mô hình ResNet50V2 đạt hiệu suất tốt nhất khi **Dropout = 0.2**, với độ chính xác **0.714**. Tại các mức Dropout khác (0.05, 0.1, 0.15), hiệu suất giảm dần, điều này cho thấy ResNet50V2 yêu cầu mức Dropout vừa phải để đạt hiệu suất tối ưu.

**6. ResNet101V2:** ResNet101V2 đạt hiệu suất cao nhất tại **Dropout = 0.1**, với độ chính xác **0.743**. Khi tăng Dropout lên 0.15 hoặc 0.2, hiệu suất giảm nhẹ. Điều này cho thấy mô hình ResNet101V2 cần một mức Dropout thấp để đạt kết quả tốt nhất.

**7. ResNet152V2:** Hiệu suất cao nhất của ResNet152V2 đạt được khi **Dropout = 0.3**, với độ chính xác **0.726**. Ở các giá trị Dropout thấp hơn, hiệu suất giảm, đặc biệt là ở mức Dropout = 0.05. Mô hình này có xu hướng cải thiện khi Dropout tăng cao hơn.

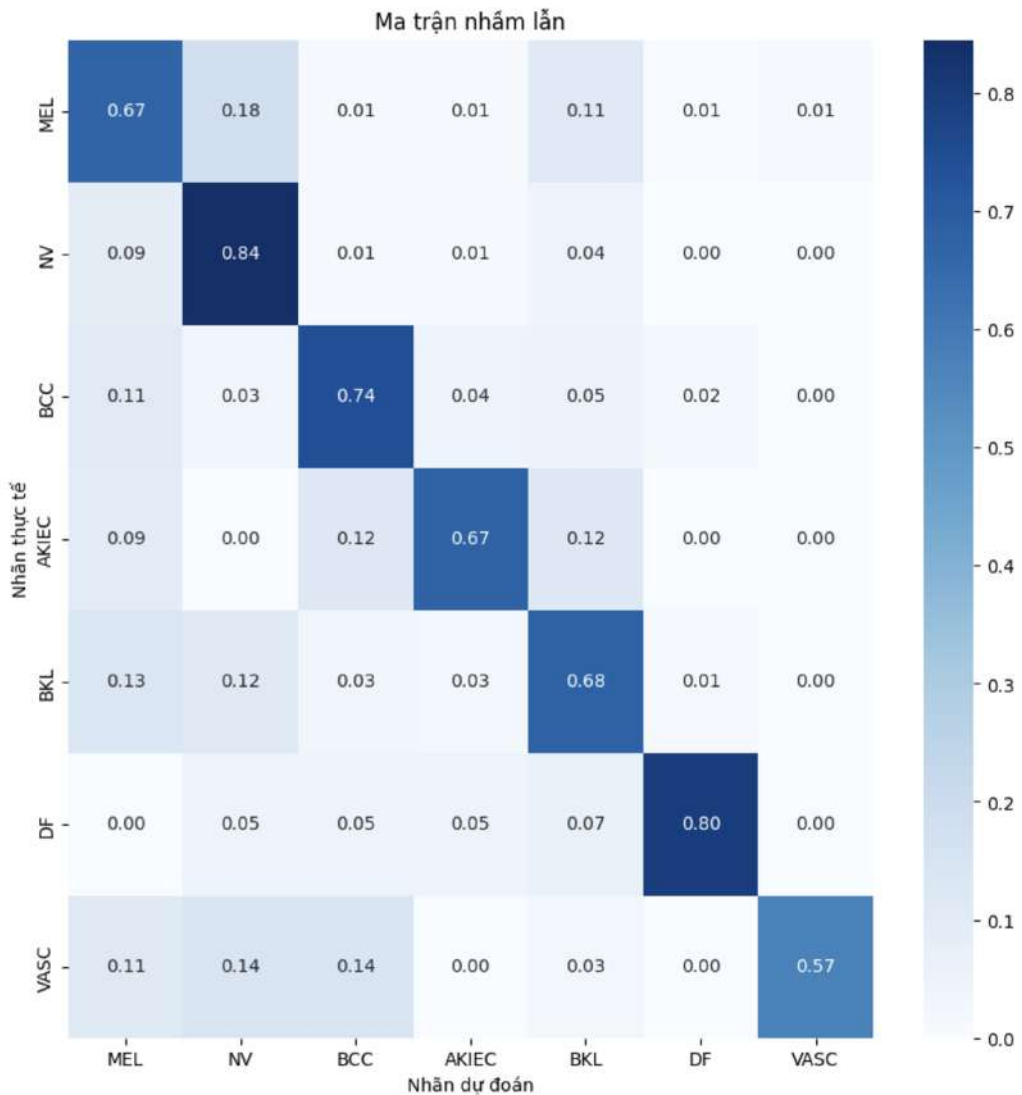
### ***Đánh giá tổng quát***

Nhìn chung, các mô hình ResNet101 và ResNet152 có xu hướng đạt hiệu suất tốt hơn khi Dropout được tăng lên (khoảng từ 0.2 đến 0.3), điều này cho thấy chúng cần mức độ regularization cao để tránh overfitting. Mặt khác, các mô hình ResNet50 và VGG16 hoạt động tốt nhất với các giá trị Dropout vừa phải (khoảng từ 0.1 đến 0.2), thể hiện rằng quá nhiều regularization có thể làm giảm hiệu suất của chúng. Mô hình ResNetV2 có xu hướng hoạt động tốt hơn ở mức Dropout thấp hơn, đặc biệt là với ResNet101V2 và ResNet50V2, điều này cho thấy kiến trúc của phiên bản V2 có khả năng chống overfitting tốt hơn mà không cần quá nhiều Dropout.

### Kết luận

Từ các kết quả trong bảng 3.5, có thể thấy rằng việc lựa chọn giá trị Dropout phù hợp phụ thuộc vào kiến trúc mô hình. Các mô hình sâu hơn như ResNet101 và ResNet152 yêu cầu mức Dropout cao hơn để đạt hiệu suất tối ưu, trong khi các mô hình như ResNet50 và VGG16 cần mức Dropout thấp hơn để hoạt động tốt nhất.

#### 3.3.2.3 Đánh giá mô hình tốt nhất



Hình 3.5: Ma trận nhầm lẫn của mô hình ResNet-101.

**Ma trận nhầm lẫn (Confusion Matrix)** Ma trận nhầm lẫn được thể hiện trong Hình 3.5 cung cấp cái nhìn chi tiết về hiệu suất của mô hình ResNet-101 trên từng lớp của bộ dữ liệu ISIC-2018. Các giá trị trên đường chéo chính biểu thị tỷ lệ dự đoán đúng (recall) cho từng lớp, trong khi các giá trị ngoài đường chéo thể hiện tỷ lệ nhầm lẫn giữa các lớp. Một số nhận xét nổi bật từ ma trận nhầm lẫn:



<b>Lớp</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>Support</b>
MEL	0.47	0.67	0.55	171
NV	0.92	0.84	0.88	909
BCC	0.68	0.74	0.71	93
AKIEC	0.58	0.67	0.62	43
BKL	0.67	0.68	0.67	217
DF	0.81	0.80	0.80	44
VASC	0.87	0.57	0.69	35
<b>Accuracy</b>		0.78 (1512 mẫu)		
<b>Macro Avg</b>	0.72	0.71	0.71	1512
<b>Weighted Avg</b>	0.81	0.78	0.79	1512

Bảng 3.6: Báo cáo phân loại của mô hình ResNet-101.

- Lớp **NV** (Nevus) đạt tỷ lệ dự đoán đúng cao nhất với recall là 84%, cho thấy mô hình phân loại lớp này rất hiệu quả.
- Lớp **DF** (Dermatofibroma) và **BCC** (Basal Cell Carcinoma) cũng có recall cao, lần lượt là 80% và 74%.
- Lớp **VASc** (Vascular Lesions) có recall thấp nhất với chỉ 57%, phản ánh khó khăn của mô hình trong việc phân biệt lớp này với các lớp khác như **NV** và **MEL**.
- Đáng chú ý, lớp **MEL** (Melanoma), một loại ung thư da ác tính quan trọng, có recall là 67%. Tuy nhiên, tỷ lệ nhầm lẫn với lớp **NV** lên đến 18%, cho thấy cần cải thiện hiệu suất phân loại giữa hai lớp có đặc điểm hình ảnh tương đồng này.

**Báo cáo phân loại (Classification Report)** Bảng 3.6 trình bày các chỉ số đánh giá hiệu suất của mô hình, bao gồm precision, recall và F1-score. Các kết quả đáng chú ý bao gồm:

- Lớp **NV** đạt hiệu suất tổng thể cao nhất với precision là 92%, recall là 84% và F1-score là 88%, khẳng định mô hình nhận diện rất tốt tổn thương thuộc lớp này.
- Các lớp **DF** và **BCC** cũng đạt F1-score khá cao, lần lượt là 80% và 71%, cho thấy hiệu suất phân loại tốt.
- Một số lớp như **MEL**, **AKIEC** và **VASc** có F1-score tương đối thấp (dưới 70%), đặc biệt là lớp **VASc** chỉ đạt 57%. Điều này cho thấy mô hình cần cải thiện trong việc nhận diện các tổn thương thuộc các lớp này.
- Precision của lớp **MEL** chỉ đạt 47%, thấp hơn đáng kể so với các lớp khác, cho thấy mô hình thường nhầm lẫn **MEL** với các lớp khác, đặc biệt là **NV**.



## Đánh giá tổng thể

- **Độ chính xác toàn cục (Accuracy):** Mô hình đạt độ chính xác 78%, phản ánh hiệu suất tổng thể ở mức khá.
- **Macro Average:** Precision, recall và F1-score trung bình lần lượt đạt 0.72, 0.71 và 0.71. Điều này cho thấy hiệu suất giữa các lớp là tương đối cân bằng, tuy nhiên vẫn tồn tại sự chênh lệch rõ rệt giữa các lớp có số lượng mẫu nhỏ (như **AKIEC**, **VASc**) và các lớp phổ biến hơn (như **NV**).
- **Weighted Average:** Các chỉ số weighted average lần lượt đạt 0.81, 0.78 và 0.79. Điều này phản ánh rằng các lớp có số lượng mẫu lớn, như **NV**, đóng góp đáng kể vào hiệu suất tổng thể của mô hình.

### 3.3.3 Kết quả mô hình cộng tác phân đoạn - phân lớp

Mô hình kết hợp giữa phân đoạn và phân lớp đã được thử nghiệm trên tập dữ liệu da liễu nhằm đánh giá hiệu quả của việc phân đoạn các vùng da tổn thương và phân loại chúng theo từng loại bệnh lý cụ thể. Các thử nghiệm được thực hiện với mô hình phân đoạn U-Net kết hợp với các mô hình phân lớp khác nhau, trong đó **ResNet101** đạt kết quả tốt nhất.

**1. Kết quả phân đoạn** Mô hình U-Net được sử dụng để phân đoạn các vùng tổn thương trên ảnh da liễu trong giai đoạn đầu. Kết quả phân đoạn đạt được các chỉ số nổi bật:

- **mAP (mean Average Precision): 94.43%**, thể hiện độ chính xác cao trong việc phát hiện các vùng da tổn thương.
- **Dice Coefficient: 93.59%**, phản ánh sự tương đồng cao giữa vùng tổn thương dự đoán và thực tế.
- **Log-Average Miss Rate (LAMR): 0.07**, cho thấy mô hình bỏ sót rất ít các vùng tổn thương.

Những kết quả này chứng minh hiệu suất vượt trội của mô hình U-Net trong việc phân đoạn và cung cấp dữ liệu chất lượng cao cho giai đoạn phân lớp tiếp theo.

**2. Kết quả phân lớp sau phân đoạn** Sau khi phân đoạn, các vùng tổn thương được đưa vào mô hình ResNet101 để phân loại theo từng loại bệnh lý. Kết quả phân lớp với mô hình ResNet101 sau phân đoạn như sau:

- **Độ chính xác (Accuracy): 78%**, cao nhất trong số các mô hình thử nghiệm.

- **Macro Average:** Precision, recall và F1-score trung bình lần lượt đạt **0.72**, **0.71**, và **0.71**.
- **Weighted Average:** Precision, recall và F1-score có trọng số lần lượt đạt **0.81**, **0.78**, và **0.79**.

Mô hình ResNet101 khi kết hợp với phân đoạn đã đạt được độ chính xác cao hơn so với các thử nghiệm không sử dụng phân đoạn. Đặc biệt, việc sử dụng phân đoạn đã giúp mô hình tập trung hơn vào các vùng tổn thương, giảm nhiễu từ các vùng da bình thường.

**3. Kết quả khi không sử dụng phân đoạn** Khi không sử dụng phân đoạn, toàn bộ ảnh được đưa trực tiếp vào mô hình ResNet101 để phân loại. Kết quả thu được như bảng 3.7.

Lớp	Precision	Recall	F1-score	Support
MEL	0.48	0.74	0.58	171
NV	0.94	0.79	0.86	909
BCC	0.69	0.73	0.71	93
AKIEC	0.65	0.60	0.63	43
BKL	0.59	0.76	0.67	217
DF	0.79	0.68	0.73	44
VASC	0.88	0.66	0.75	35
<b>Độ chính xác</b>		0.76 (1512 mẫu)		
<b>Macro Average</b>	0.72	0.71	0.70	1512
<b>Weighted Average</b>	0.81	0.76	0.78	1512

Bảng 3.7: Báo cáo phân loại của ResNet101 khi không sử dụng phân đoạn.

Khi không sử dụng phân đoạn, độ chính xác đạt **76%**, thấp hơn so với **78%** khi kết hợp phân đoạn từ U-Net. Một số điểm đáng chú ý từ kết quả:

- Lớp **NV** (Nevus) có hiệu suất tốt nhất với F1-score đạt **0.86**, nhờ số lượng mẫu lớn và sự khác biệt rõ rệt với các lớp khác.
- Lớp **VASc** (Vascular Lesions) đạt F1-score **0.75**, cao hơn một số lớp khác nhờ độ đặc trưng cao trong ảnh.
- Lớp **MEL** (Melanoma) có F1-score thấp nhất (**0.58**), cho thấy khó khăn của mô hình trong việc phân biệt Melanoma với lớp **NV**.
- Các chỉ số trung bình: Precision đạt **0.72**, Recall đạt **0.71**, và F1-score trung bình đạt **0.70**, cho thấy hiệu suất của mô hình tương đối ổn định nhưng vẫn còn hạn chế ở các lớp có ít mẫu như **AKIEC** và **DF**.

Nhìn chung, việc không sử dụng phân đoạn dẫn đến hiệu suất phân loại thấp hơn, đặc biệt là ở các lớp có tổn thương nhỏ hoặc các đặc điểm tương đồng như **MEL** và **NV**. Điều này nhấn mạnh tầm quan trọng của việc sử dụng phân đoạn để giúp mô hình tập trung hơn vào các vùng tổn thương có ý nghĩa.

**4. Đánh giá các chỉ số và hạn chế** Mặc dù kết quả khi kết hợp phân đoạn và phân lớp được cải thiện, vẫn còn một số hạn chế cần lưu ý:

- **Nhầm lẫn giữa các lớp:** Mô hình gặp khó khăn trong việc phân biệt giữa các lớp có đặc điểm tương đồng, đặc biệt là **MEL** và **NV**.
- **Tỷ lệ False Positives:** Một số vùng da bình thường vẫn bị dự đoán nhầm là tổn thương. Điều này có thể được giảm thiểu thông qua các phương pháp hậu xử lý (*post-processing*) hoặc cải thiện mô hình phân đoạn.
- **Tổn thương nhỏ:** Các tổn thương có kích thước nhỏ hoặc đặc điểm không rõ ràng vẫn là thách thức lớn đối với mô hình, làm giảm hiệu suất trong một số trường hợp.

**5. Kết luận** Mô hình ResNet101 kết hợp với phân đoạn từ U-Net đã chứng minh hiệu quả vượt trội, đạt độ chính xác **78%**, cải thiện rõ rệt so với mô hình phân lớp trực tiếp (**76%**). Việc sử dụng phân đoạn không chỉ giúp giảm nhiễu mà còn tăng cường khả năng tập trung vào các vùng tổn thương quan trọng, từ đó nâng cao hiệu suất phân loại.

Tuy nhiên, để đạt được độ tin cậy cao hơn trong các ứng dụng thực tế, cần tiếp tục cải thiện khả năng phân biệt giữa các lớp tương đồng, giảm tỷ lệ False Positives và tối ưu hóa xử lý các tổn thương nhỏ. Những cải tiến này sẽ giúp hệ thống trở nên hiệu quả và đáng tin cậy hơn trong việc hỗ trợ chẩn đoán bệnh lý da liễu.

### 3.4 Kết luận

Nghiên cứu này đã chứng minh được hiệu quả vượt trội của mô hình kết hợp giữa phân đoạn và phân lớp trong chẩn đoán các bệnh lý da liễu. Kết quả thực nghiệm cho thấy việc sử dụng mô hình U-Net để phân đoạn vùng tổn thương đã cải thiện đáng kể hiệu suất của các mô hình phân lớp như ResNet-50, ResNet-101, ResNet-152 và VGG-16. Bằng cách cô lập chính xác các vùng tổn thương, bước phân đoạn giúp giảm nhiễu từ các vùng da không liên quan, từ đó nâng cao độ chính xác trong phân loại.

Sự kết hợp này đặc biệt quan trọng trong các bài toán chẩn đoán y học, khi việc phát hiện và phân loại chính xác vùng tổn thương là yếu tố then chốt để đưa ra các chẩn đoán chính xác. So với các thử nghiệm chỉ sử dụng mô hình phân lớp trực tiếp trên toàn bộ ảnh, việc phân đoạn trước đã giúp tăng độ chính xác phân loại một cách rõ rệt. Ví dụ, đối với các mô hình ResNet, độ chính xác trung bình đã tăng từ 65-70% lên trên 78%.

Mặc dù hệ thống đã đạt được những kết quả tích cực, nhưng vẫn tồn tại một số thách thức cần được khắc phục, bao gồm tỷ lệ *false positives* còn tương đối cao và khó khăn trong việc phân loại các vùng tổn thương nhỏ hoặc có đặc điểm tương tự với vùng da lành. Những thách thức này đòi hỏi cần có sự cải tiến trong mô hình phân đoạn và áp dụng các phương pháp hậu xử lý để giảm thiểu sai sót.

Tiềm năng ứng dụng của mô hình kết hợp phân đoạn và phân lớp là rất lớn, đặc biệt trong các hệ thống chẩn đoán tự động. Mô hình này có thể hỗ trợ đắc lực cho các bác sĩ da liễu trong việc phát hiện sớm các bệnh lý da liễu, đồng thời giúp giảm tải khối lượng công việc và nâng cao chất lượng chẩn đoán. Tuy nhiên, để áp dụng vào thực tiễn, cần mở rộng thử nghiệm trên các bộ dữ liệu đa dạng hơn, bao gồm các loại da và bệnh lý phức tạp hơn, nhằm đảm bảo tính tổng quát và độ tin cậy của mô hình.

Tóm lại, nghiên cứu này đã cho thấy tiềm năng lớn của việc kết hợp giữa phân đoạn và phân lớp trong chẩn đoán bệnh lý da liễu. Các hướng phát triển trong tương lai sẽ tập trung vào việc cải thiện mô hình phân đoạn, tối ưu hóa kỹ thuật huấn luyện và thử nghiệm trên các tập dữ liệu lớn hơn để đảm bảo mô hình hoạt động hiệu quả trong các ứng dụng lâm sàng thực tế.

## Chương 4

# KẾT LUẬN

### 4.1 Kết luận

Luận văn đã trình bày một hệ thống chẩn đoán tự động bệnh lý da liễu dựa trên việc kết hợp mô hình phân đoạn và phân lớp hình ảnh da liễu, sử dụng các bộ dữ liệu tiêu chuẩn quốc tế như ISIC 2018 và HAM10000. Hệ thống áp dụng mô hình U-Net cho phân đoạn vùng tổn thương da và các mô hình phân lớp mạnh mẽ như ResNet-50, ResNet-101, ResNet-152 và VGG-16 cho việc phân loại các vùng tổn thương. Các kết quả thực nghiệm đã chứng minh rằng việc phân đoạn trước khi phân loại mang lại cải thiện đáng kể về hiệu suất, giúp độ chính xác phân loại tăng lên đáng kể so với các phương pháp chỉ phân lớp trực tiếp trên toàn bộ ảnh.

Cụ thể, việc sử dụng mô hình U-Net để phân đoạn các vùng tổn thương đã giúp cô lập và làm nổi bật các khu vực quan trọng, từ đó cải thiện đáng kể độ chính xác của các mô hình phân lớp. Đặc biệt, các mô hình ResNet, với kiến trúc sâu và các kết nối tắt (skip connections), đã đạt được hiệu suất cao nhất trong số các mô hình thử nghiệm, với độ chính xác trung bình vượt qua mức 78%.

Bên cạnh đó, luận văn cũng đã xem xét một số hạn chế của hệ thống, bao gồm tỷ lệ *false positives* cao trong một số trường hợp và khó khăn trong việc phát hiện các tổn thương nhỏ hoặc có đặc điểm tương đồng với vùng da lành. Kết quả này chỉ ra rằng vẫn còn nhiều thách thức trong việc áp dụng mô hình này vào thực tế, đặc biệt là trong môi trường lâm sàng.

Ngoài ra, hệ thống AI hỗ trợ chẩn đoán bệnh da liễu được xây dựng trong luận văn đã được mô tả chi tiết trong phụ lục C, với dữ liệu thực tế (ảnh chụp da liễu) từ dự án thuộc khuôn khổ đề tài **”Nghiên cứu phát triển hệ thống trí tuệ nhân tạo hỗ trợ chẩn đoán bệnh Vẩy nến, Viêm da cơ địa và Ung thư da tại Việt Nam”** (Mã số: KC-4.0-44/19-25). Dự án này có sự tham gia của **Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam** và **Bệnh viện Da liễu Trung ương**, tạo cơ sở dữ liệu đa dạng và phong phú, giúp tối ưu hóa hệ thống AI cho các bệnh lý da đặc thù tại Việt Nam.

## 4.2 Hướng phát triển

Dựa trên các kết quả đã đạt được và những hạn chế còn tồn tại, có nhiều hướng phát triển tiềm năng cho nghiên cứu này trong tương lai, đặc biệt khi áp dụng với dữ liệu da liễu thực tế tại Việt Nam và phát triển một hệ thống AI chuyên dụng.

### 4.2.1 Thu thập và xây dựng bộ dữ liệu da liễu thực tế tại Việt Nam

Một trong những hướng quan trọng để phát triển hệ thống chẩn đoán tự động này là xây dựng và mở rộng bộ dữ liệu da liễu thực tế tại Việt Nam. Hiện nay, hầu hết các bộ dữ liệu được sử dụng trong nghiên cứu là từ các nguồn quốc tế, nơi môi trường và đặc điểm da của bệnh nhân có thể khác biệt so với Việt Nam. Do đó, việc thu thập một bộ dữ liệu da liễu từ các bệnh viện, phòng khám da liễu và các cơ sở y tế trong nước là điều cần thiết để đảm bảo hệ thống có thể hoạt động chính xác và hiệu quả hơn với các đặc điểm da và bệnh lý tại Việt Nam.

Bộ dữ liệu này cần được chuẩn hóa, bao gồm các hình ảnh nội soi da với đầy đủ các thông tin về chẩn đoán và gán nhãn bệnh lý, tương tự như các bộ dữ liệu ISIC. Dữ liệu này không chỉ giúp cải thiện hiệu quả của hệ thống mà còn đóng góp vào việc phát triển các nghiên cứu y học trong nước.

### 4.2.2 Phát triển hệ thống AI chẩn đoán tự động bệnh lý da liễu

Hướng tiếp theo là phát triển một hệ thống AI chẩn đoán tự động dựa trên mô hình kết hợp giữa phân đoạn và phân lớp, được tối ưu hóa cho dữ liệu da liễu Việt Nam. Hệ thống này sẽ giúp hỗ trợ các bác sĩ trong việc chẩn đoán và điều trị bệnh lý da liễu, đặc biệt ở các khu vực thiếu thốn nguồn lực y tế. Hệ thống có thể được triển khai dưới dạng ứng dụng di động hoặc tích hợp vào các thiết bị y tế, cho phép người dùng chụp ảnh da của họ và nhận được chẩn đoán sơ bộ.

**Những bước phát triển chính của hệ thống AI bao gồm:**

- **Cải tiến mô hình phân đoạn:** Cần nghiên cứu các mô hình phân đoạn tiên tiến hơn như Mask R-CNN hoặc các biến thể U-Net mới, nhằm tăng độ chính xác trong việc phát hiện các vùng tổn thương nhỏ hoặc khó phân biệt.
- **Tối ưu hóa mô hình phân lớp:** Việc sử dụng các mô hình phân lớp như ResNet-152 hay các mô hình học sâu tiên tiến hơn như EfficientNet có thể mang lại độ chính xác cao hơn khi xử lý các tập dữ liệu lớn và đa dạng.
- **Học tăng cường từ phản hồi thực tế:** Hệ thống có thể được cải tiến thông qua việc học tăng cường từ phản hồi của các bác sĩ trong quá trình sử dụng, giúp mô hình ngày càng hoàn thiện và phù hợp hơn với thực tế lâm sàng.

- **Ứng dụng học chuyển giao (transfer learning):** Hệ thống có thể áp dụng học chuyển giao để tận dụng các mô hình đã được huấn luyện trên các bộ dữ liệu quốc tế, sau đó tinh chỉnh với dữ liệu trong nước, giúp rút ngắn thời gian huấn luyện và tăng cường độ chính xác.

### 4.2.3 Đánh giá và thử nghiệm lâm sàng

Cuối cùng, để đảm bảo tính khả thi và độ tin cậy của hệ thống AI chẩn đoán tự động, việc đánh giá và thử nghiệm lâm sàng là bước không thể thiếu. Các thử nghiệm lâm sàng cần được thực hiện trên nhiều loại bệnh nhân với các đặc điểm da khác nhau, qua đó đánh giá khả năng tổng quát hóa của hệ thống trong môi trường thực tế.

Việc hợp tác với các bệnh viện và phòng khám da liễu trong nước là cần thiết để thử nghiệm hệ thống, thu thập thêm dữ liệu phản hồi, và tiếp tục tối ưu hóa các mô hình chẩn đoán nhằm đáp ứng các tiêu chuẩn y tế cao nhất.

## 4.3 Kết luận chung

Nghiên cứu này đã chứng minh được tiềm năng lớn của việc kết hợp phân đoạn và phân lớp trong chẩn đoán bệnh lý da liễu. Mô hình đề xuất đã đạt hiệu suất cao trong việc phân loại các loại bệnh lý da phổ biến, và mở ra những tiềm năng ứng dụng thực tế trong hệ thống y tế Việt Nam. Tuy nhiên, để hệ thống có thể áp dụng rộng rãi, việc thu thập dữ liệu thực tế từ Việt Nam và tối ưu hóa hệ thống cho các điều kiện lâm sàng trong nước là cần thiết.

Trong tương lai, việc phát triển một hệ thống AI chẩn đoán tự động dựa trên dữ liệu da liễu Việt Nam, cùng với việc thử nghiệm lâm sàng và cải tiến các mô hình hiện có, sẽ là những bước quan trọng để đưa hệ thống này vào ứng dụng thực tế, góp phần nâng cao chất lượng chẩn đoán và điều trị bệnh lý da liễu tại Việt Nam.

# TÀI LIỆU THAM KHẢO

- [1] P. E. LeBoit **and others**, *World Health Organization Classification of Tumours: Pathology and Genetics of Skin Tumours*. Lyon: IARC Press, 2006, **pages** 10–19.
- [2] L. T. Vinh, “Khảo sát mô hình bệnh tật tại Bệnh viện Da liễu Trung ương giai đoạn 2017 – 2019,” Bệnh viện Da liễu Trung ương, Đề tài nghiên cứu cấp cơ sở, 2020.
- [3] C. Cortes **and** V. Vapnik, “Support-vector networks,” *Machine learning*, **journal** 20, **number** 3, **pages** 273–297, 1995.
- [4] Y. LeCun, Y. Bengio **and** G. Hinton, “Deep learning,” *Nature*, **journal** 521, **number** 7553, **pages** 436–444, 2015.
- [5] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, **journal** 9, **number** 1, **pages** 62–66, 1979.
- [6] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on pattern analysis and machine intelligence*, **number** 6, **pages** 679–698, 1986.
- [7] S. Beucher, “The morphological approach to segmentation: the watershed transformation,” **in** *Mathematical morphology in image processing* CRC Press, 1992, **pages** 433–481.
- [8] T. Cover **and** P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, **journal** 13, **number** 1, **pages** 21–27, 1967.
- [9] L. Breiman, J. Friedman, C. J. Stone **and** R. A. Olshen, *Classification and regression trees*. Wadsworth International Group, 1984.
- [10] N. Dalal **and** B. Triggs, “Histograms of oriented gradients for human detection,” **in** *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)* IEEE, **volume** 1, 2005, **pages** 886–893.
- [11] T. Ojala, M. Pietikainen **and** D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern recognition*, **journal** 29, **number** 1, **pages** 51–59, 1996.
- [12] L. Breiman, “Random forests,” *Machine learning*, **journal** 45, **number** 1, **pages** 5–32, 2001.
- [13] S. Z. Li, *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.



- [14] A. Krizhevsky, I. Sutskever **and** G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, **jourvol** 25, **pages** 1097–1105, 2012.
- [15] O. Ronneberger, P. Fischer **and** T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” **in** *International Conference on Medical image computing and computer-assisted intervention* Springer, 2015, **pages** 234–241.
- [16] O. Oktay, J. Schlemper, L. Le Folgoc **and** others, “Attention U-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.
- [17] Z. Zhang, Q. Liu **and** Y. Wang, “Road extraction by deep residual U-Net,” *IEEE Geoscience and Remote Sensing Letters*, **jourvol** 15, **number** 5, **pages** 749–753, 2018.
- [18] J. Long, E. Shelhamer **and** T. Darrell, “Fully convolutional networks for semantic segmentation,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, **pages** 3431–3440, 2015.
- [19] V. Badrinarayanan, A. Kendall **and** R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, **jourvol** 39, **number** 12, **pages** 2481–2495, 2017.
- [20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy **and** A. L. Yuille, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.
- [21] D. P. Kingma **and** J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv preprint arXiv:1412.6980*, 2014. **url:** <https://arxiv.org/abs/1412.6980>.
- [22] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin **and** M. J. Cardoso, “Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations,” **in** *Deep learning in medical image analysis and multimodal learning for clinical decision support* Springer, 2017, **pages** 240–248. DOI: [10.1007/978-3-319-67558-9\\_28](https://doi.org/10.1007/978-3-319-67558-9_28).
- [23] A. Esteva, B. Kuprel, R. A. Novoa **and** others, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, **jourvol** 542, **number** 7639, **pages** 115–118, 2017. DOI: [10.1038/nature21056](https://doi.org/10.1038/nature21056). **url:** <https://doi.org/10.1038/nature21056>.
- [24] K. He, X. Zhang, S. Ren **and** J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, **pages** 770–778, 2016.
- [25] G. Huang, Z. Liu, L. Van Der Maaten **and** K. Q. Weinberger, “Densely connected convolutional networks,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, **pages** 4700–4708, 2017.

- [26] J. Yosinski, J. Clune, Y. Bengio **and** H. Lipson, “How transferable are features in deep neural networks?” *in Advances in neural information processing systems* 2014, **pages** 3320–3328.
- [27] D. P. Kingma **and** J. Ba, *Adam: A Method for Stochastic Optimization*, 2017. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG]. **url:** <https://arxiv.org/abs/1412.6980>.
- [28] Y. Lecun, L. Bottou, Y. Bengio **and** P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, **journal** 86, **number** 11, **pages** 2278–2324, 1998. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [29] J. Kawahara **and** G. Hamarneh, “Multi-resolution-tract CNN with Hybrid Pretrained and Skin-lesion Trained Layers,” *in International Workshop on Machine Learning in Medical Imaging* Springer, 2016.
- [30] A. Esteva, B. Kuprel, R. A. Novoa **and others**, “Dermatologist-level Classification of Skin Cancer with Deep Neural Networks,” *Nature*, **journal** 542, **number** 7639, **pages** 115–118, 2017.
- [31] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park **and** S. E. Chang, “Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm,” *Journal of Investigative Dermatology*, **journal** 138, **number** 7, **pages** 1529–1538, 2018. DOI: [10.1016/j.jid.2018.01.028](https://doi.org/10.1016/j.jid.2018.01.028).
- [32] T. J. Brinker, A. Hekler, A. H. Enk **and** C. von Kalle, “Enhanced Classifier Training to Improve Precision of a Convolutional Neural Network to Identify Images of Skin Lesions,” *PLoS ONE*, **journal** 14, **number** 6, e0218713, 2019.
- [33] Z. Yu, D. Ni, S. Chen **and others**, “Hybrid Dermoscopy Image Classification Framework Based on Deep Convolutional Neural Network and Fisher Vector,” *in Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI)* 2017, **pages** 301–304.
- [34] A. Menegola, M. Fornaciali, R. Pires, F. V. Bittencourt, S. Avila **and** E. Valle, “Knowledge Transfer for Melanoma Screening with Deep Learning,” *arXiv preprint arXiv:1703.07479*, 2017.
- [35] K. Dopytalska, P. Sobolewski, A. Błaszczak, E. Szymańska **and** I. Walecka, “Psoriasis in Special Localizations,” *Reumatologia/Rheumatology*, **journal** 56, **number** 6, **pages** 392–398, 2018. DOI: [10.5114/reum.2018.80718](https://doi.org/10.5114/reum.2018.80718).
- [36] X. Glorot, A. Bordes **and** Y. Bengio, “Deep Sparse Rectifier Neural Networks,” *in Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* G. Gordon, D. Dunson **and** M. Dudík, **editors**, **journal** Proceedings of Machine Learning Research, **volume** 15, Fort Lauderdale, FL, USA: PMLR, **november** 2011, **pages** 315–323. **url:** <https://proceedings.mlr.press/v15/glorot11a.html>.

- [37] J. S. Bridle, “Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition,” in *NATO Neurocomputing* 1989. **url:** <https://api.semanticscholar.org/CorpusID:59636530>.
- [38] *ISIC 2018 Challenge*, 2018. **url:** <https://challenge.isic-archive.com/landing/2018/>.
- [39] P. Tschandl, C. Rosendahl and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific Data*, **jourvol** 5, **number** 1, 2018. DOI: [10.1038/sdata.2018.161](https://doi.org/10.1038/sdata.2018.161).
- [40] N. C. Codella, V. Rotemberg, P. Tschandl and others, “Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), held in Melbourne, Australia, April 2017,” *IEEE Transactions on Medical Imaging*, **jourvol** 38, **number** 10, **pages** 2366–2378, 2019. DOI: [10.1109/TMI.2019.2905860](https://doi.org/10.1109/TMI.2019.2905860).
- [41] A. Krizhevsky, I. Sutskever and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems* F. Pereira, C. Burges, L. Bottou and K. Weinberger, **editors**, **volume** 25, Curran Associates, Inc., 2012. **url:** [https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf).
- [42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, **jourvol** 15, **number** 56, **pages** 1929–1958, 2014. **url:** <http://jmlr.org/papers/v15/srivastava14a.html>.
- [43] M. Everingham and others, “The Pascal Visual Object Classes (VOC) Challenge,” *International Journal of Computer Vision*, **jourvol** 88, **number** 2, **pages** 303–338, 2010.
- [44] T.-Y. Lin, M. Maire, S. Belongie and others, “Microsoft COCO: Common Objects in Context,” in *European conference on computer vision* Springer, 2014, **pages** 740–755.
- [45] P. Dollar, C. Wojek, B. Schiele and P. Perona, “Pedestrian Detection: An Evaluation of the State of the Art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **jourvol** 36, **number** 4, **pages** 743–761, 2014.
- [46] P. Tschandl, C. Rosendahl and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific Data*, **jourvol** 5, **number** 1, **august** 2018. DOI: [10.1038/sdata.2018.161](https://doi.org/10.1038/sdata.2018.161). **url:** <https://doi.org/10.1038/sdata.2018.161>.

- [47] K. Ali, Z. A. Shaikh, A. A. Khan **and** A. A. Laghari, “Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer,” *Neuroscience Informatics*, **journal** 2, **number** 4, **page** 100 034, 2022, ISSN: 2772-5286. DOI: [10 . 1016 / j . neuri . 2021 . 100034](https://doi.org/10.1016/j.neuri.2021.100034). **url**: [https : // www . sciencedirect . com / science / article/pii/S2772528621000340](https://www.sciencedirect.com/science/article/pii/S2772528621000340).
- [48] R. Garg, S. Maheshwari **and** A. Shukla, “Decision Support System for Detection and Classification of Skin Cancer Using CNN,” **in** *Innovations in Computational Intelligence and Computer Vision* M. K. Sharma, V. S. Dhaka, T. Perumal, N. Dey **and** J. M. R. S. Tavares, **editors**, Singapore: Springer Singapore, 2021, **pages** 578–586, ISBN: 978-981-15-6067-5.
- [49] L. Faes, S. K. Wagner, D. J. Fu, X. Liu, E. Korot, J. R. Ledsam **and others**, “Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study,” *The Lancet Digital Health*, **journal** 1, **number** 5, e232–e242, 2019, ISSN: 2589-7500. DOI: [10 . 1016 / S2589 - 7500 \(19\) 30108 - 6](https://doi.org/10.1016/S2589-7500(19)30108-6). **url**: [https : // www . sciencedirect . com / science / article / pii / S2589750019301086](https://www.sciencedirect.com/science/article/pii/S2589750019301086).
- [50] M. Combalia, N. C. F. Codella, V. Rotemberg **and others**, *BCN20000: Dermoscopic Lesions in the Wild*, 2019. DOI: [10 . 48550 / arXiv . 1908 . 02288](https://doi.org/10.48550/arXiv.1908.02288). **url**: [https : // arxiv . org / abs / 1908 . 02288](https://arxiv.org/abs/1908.02288).
- [51] M. A. Kassem, K. M. Hosny **and** M. M. Fouad, “Skin Lesions Classification Into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning,” *IEEE Access*, **journal** 8, **pages** 114 822–114 832, 2020. DOI: [10 . 1109 / ACCESS . 2020 . 3003890](https://doi.org/10.1109/ACCESS.2020.3003890).
- [52] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska **and** M. H. Yap, “Analysis of the ISIC image datasets: Usage, benchmarks and recommendations,” *Medical Image Analysis*, **journal** 75, **page** 102 305, 2022, ISSN: 1361-8415. DOI: [10 . 1016 / j . media . 2021 . 102305](https://doi.org/10.1016/j.media.2021.102305). **url**: [https : // www . sciencedirect . com / science / article / pii / S1361841521003509](https://www.sciencedirect.com/science/article/pii/S1361841521003509).
- [53] A. Gong, X. Yao **and** W. Lin, “Dermoscopy Image Classification Based on StyleGANs and Decision Fusion,” *IEEE Access*, **journal** 8, **pages** 70 640–70 650, 2020. DOI: [10 . 1109 / ACCESS . 2020 . 2986916](https://doi.org/10.1109/ACCESS.2020.2986916).

# PHỤ LỤC A | Cơ sở lý thuyết

Trong phần này, tôi trình bày khái quát một số lý thuyết và công nghệ chính mà mình sử dụng. Mục đích của chương này là cung cấp cho người đọc không đúng chuyên ngành kiến thức để đọc hiểu các phần tiếp theo và cũng giúp cho người phản biện đánh giá được mức độ lĩnh hội kiến thức của tôi.

## A.1 Đại số tuyến tính

### A.1.1 Định nghĩa ma trận

Một ma trận  $\mathbf{A}$  cấp  $mn$  trên trường  $\mathbb{K}$  ( $\mathbb{K}$  - là trường thực  $\mathbb{R}$  hoặc phức  $\mathbb{C}$ ) là một bảng chữ nhật gồm  $mn$  phần tử trong  $\mathbb{K}$  được viết thành  $m$  dòng và  $n$  cột như sau:

$$A_{m,n} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

Trong đó:

-  $a_{i,j}$  là phần tử của ma trận  $\mathbf{A}$  nằm ở giao điểm của dòng  $i$  và cột  $j$ . -  $m$  là số dòng của ma trận  $\mathbf{A}$ . -  $n$  là số cột của ma trận  $\mathbf{A}$ . -  $[a_{i,1} \ a_{i,2} \ \cdots \ a_{i,n}]$  là dòng thứ  $i$  của ma trận  $\mathbf{A}$ . -

$\begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix}$  là cột thứ  $j$  của ma trận  $\mathbf{A}$ .

### A.1.2 Phép cộng ma trận

Cho hai ma trận  $\mathbf{A}$ ,  $\mathbf{B}$  cùng cỡ  $mn$ , ta có tổng  $\mathbf{A} + \mathbf{B}$  là ma trận  $\mathbf{C}$  có cùng kích thước ( $mn$ ) với phần tử trong vị trí tương ứng bằng tổng của hai phần tử tương ứng của mỗi ma trận:  $c_{i,j} = a_{i,j} + b_{i,j}$  với  $1 \leq i \leq m$  và  $1 \leq j \leq n$

### A.1.3 Phép nhân ma trận

Cho ma trận  $\mathbf{A}_{mp}$  và ma trận  $\mathbf{B}_{pn}$ , trong đó số cột của ma trận  $\mathbf{A}$  bằng số hàng của ma trận  $\mathbf{B}$ . Tích  $\mathbf{AA}$  là ma trận  $\mathbf{C}$  có  $m$  hàng và  $n$  cột, phần tử  $c_{i,j}$  được xác định theo tích vô hướng của hàng tương ứng trong  $\mathbf{A}$  với cột tương ứng trong  $\mathbf{B}$ :

$$c_{i,j} = \sum_{k=1}^p a_{i,k} b_{k,j}$$

Ngoài ra có một phép nhân khác được gọi là \*element-wise\* được sử dụng khá nhiều trong học máy. Tích \*element-wise\* của hai ma trận cùng kích thước  $\mathbf{A}$ ,  $\mathbf{B}$  được kí hiệu là  $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$ , trong đó:

$$c_{i,j} = a_{i,j} b_{i,j}$$

## A.2 Giải tích

### A.2.1 Đạo hàm

#### A.2.1.1 Định nghĩa

Cho hàm số  $y = f(x)$  xác định trên khoảng  $(a; b)$  (khoảng  $(a; b) = \{x \in \mathbb{R} | a < x < b\}$ ). Xét giá trị  $x_0$  và giá trị  $x \in (a; b)$ ,  $x \neq x_0$ .

Đặt  $\Delta x = x - x_0$  thì  $x = x_0 + \Delta x$  và  $\Delta x$  được gọi là số gia đối số.

Đặt  $\Delta y = f(x) - f(x_0)$  và  $\Delta y$  được gọi là số gia hàm số.

Xét tỷ số  $\frac{\Delta y}{\Delta x}$ . Nếu khi  $\Delta x \rightarrow 0$ , tỷ số đó dần tới một giới hạn thì giới hạn đó được gọi là đạo hàm của hàm số  $y = f(x)$  tại điểm  $x_0$  ký hiệu là  $f'(x)$

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

#### A.2.1.2 Đạo hàm riêng

Đạo hàm riêng của một hàm số đa biến là ta coi tất cả các biến khác như là hằng số và tiến hành đạo hàm theo biến đã chọn.

Đạo hàm riêng của  $f$  đối với biến  $x$  được ký hiệu khác nhau bởi:  $f'_x = \frac{df}{dx}$ .

**Vector gradient:** Cho một hàm số  $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ . Trong trường hợp này  $f$  có các đạo hàm riêng  $\frac{df}{dx_j}$  đối với mỗi biến  $x_j$  ( $1 \leq j \leq n$ ) thì vector chứa các đạo hàm riêng này là vector gradient.

$$\Delta_x f(x) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

### A.2.1.3 Chain rule

Đạo hàm hàm hợp là công thức để tính đạo hàm của hàm số gồm nhiều hàm số kết hợp với nhau. Đó là, nếu  $f, g$  là hai hàm số và hàm  $h(x) = f(g(x))$  thì ta có

$$h'(x) = f(g(x))' = f'(g(x)) \cdot g'(x)$$

hay chúng ta có công thức quen thuộc hơn với cách đặt  $z = f(y), y = g(x)$ :

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} = f'(y)g'(x) = f'(g(x)) \cdot g'(x)$$

## A.3 Xác suất

### A.3.1 Các phép toán xác suất

#### A.3.1.1 Tổng xác suất

Tổng xác suất là xác suất của sự kiện hợp. Cho tập sự kiện  $A_i, i = \overline{1, n}$  khi đó ta có:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i+1} \sum_{i \leq k_1 \leq \dots \leq k_i \leq n} P\left(\bigcap_{j=1}^i A_{k_j}\right)$$

Trong đó, tổng  $\sum_{i \leq k_1 \leq \dots \leq k_i \leq n} P(\bigcap_{j=1}^i A_{k_j})$  là tổng của tất cả các xác suất giao của tập con gồm  $i$  phần tử từ tập  $1, 2, \dots, n$

#### A.3.1.2 Xác suất có điều kiện

Là xác suất của một sự kiện xảy ra khi biết xác suất của sự kiện đã xảy ra. Xác suất của sự kiện  $A$  khi biết  $B$  đã xảy ra được ký hiệu là  $P(A|B)$ . Công thức tính xác suất của  $A$  được xác định như sau:

$$P(A|B) = \frac{P(AB)}{P(B)} \text{ trong đó } P(B) \geq 0$$

Nếu  $A$  và  $B$  là độc lập, tức  $A$  không phụ thuộc vào  $B$  thì  $P(A|B) = P(A)$  và  $P(B|A) = P(B)$ . Xác suất có điều kiện cũng có các tính chất như xác suất thông thường:

$$- P\left(\bigcup_{i=1}^n A_i | B\right) = \sum_{i=1}^n (-1)^{i+1} \sum_{i \leq k_1 \leq \dots \leq k_i \leq n} P\left(\bigcap_{j=1}^i A_{k_j} | B\right) - P(\overline{A} | B) = 1 - P(A|B)$$

#### A.3.1.3 Tích xác suất

Tích xác suất là xác suất của sự kiện giao. Từ công thức xác suất có điều kiện ta có thể tính được xác suất giao như sau:

$$P(AB) = P(B)P(A|B) = P(A)P(B|A)$$

Trường hợp tổng quát, cho  $A_i, i = \overline{1, n}$  thì tích xác suất của chúng được tính như sau:

$$P\left(\bigcup_{i=1}^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_2A_1) \dots P(A_n|A_1A_2 \dots A_{n-1})$$

Hay viết gọn thành:

$$P\left(\bigcup_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i | \bigcap_{j=1}^{i-1} A_j)$$

Tích xác suất còn được gọi là quy tắc chuỗi xác suất bởi cách biểu diễn liên hoàn thành chuỗi như trên.

Nếu  $A_j$  là độc lập từng đôi một thì ta có:

$$P\left(\bigcup_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

### A.3.2 Xác suất hậu nghiệm - Bayes

Từ công thức tính tích xác suất ta có suy luận sau:

$$P(A)P(B|A) = P(B)P(A|B)$$

Từ đó, ta có thể tính xác suất của A khi biết B như sau:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Trong đó: -  $P(A|B)$ : Xác suất có điều kiện -  $P(A)$ : Xác suất tiền nghiệm -  $P(B)$ : Xác suất hậu nghiệm -  $P(B|A)$ : Khả dĩ (likelihood)

## A.4 Học máy

### A.4.1 Khái niệm

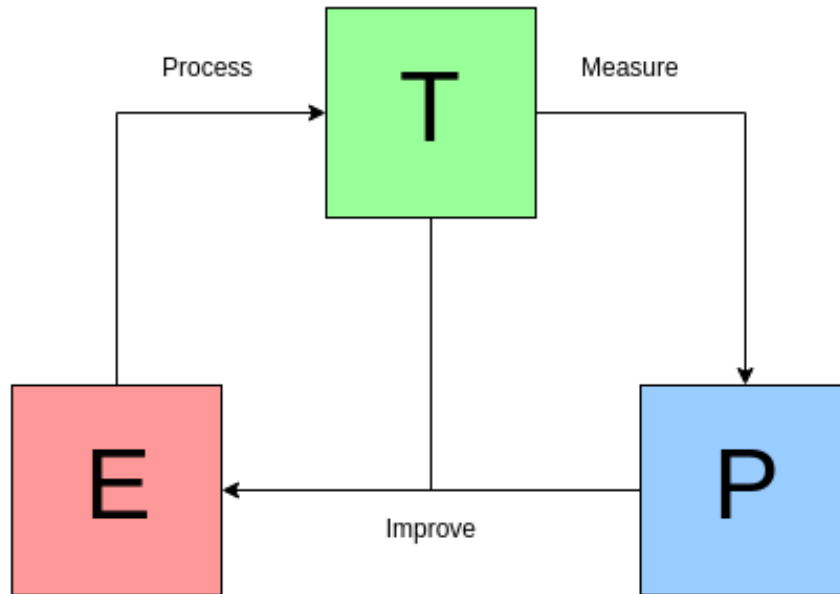
Học máy được đề ra vào năm 1959 bởi Arthur Samuel. Sau đó, Tom M. Mitchell đã đưa ra một định nghĩa chính thức hơn về các thuật toán được nghiên cứu trong lĩnh vực học máy:

"Một chương trình máy tính được cho là học hỏi từ kinh nghiệm E đối với một số loại nhiệm vụ T và hiệu suất đo P. Nếu hiệu suất của nó trong các nhiệm vụ trong T, được đo



bằng P, sẽ cải thiện kinh nghiệm E <sup>1</sup>."

Như vậy, quy trình học máy như hình A.1



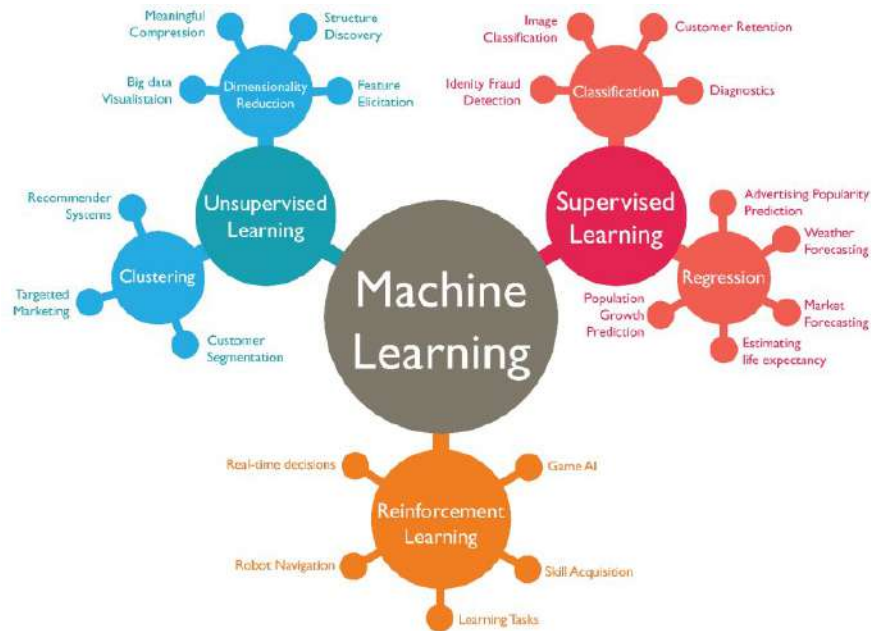
Hình A.1: Quy trình máy học

#### A.4.2 Các phương pháp học

Có ba phương pháp học phổ biến là học có giám sát (supervised learning), học không giám sát (unsupervised learning), học tăng cường (reinforcement learning) như hình A.2:

- **Học có giám sát:** Là phương pháp sử dụng những dữ liệu đã được gán nhãn từ trước để suy luận ra quan hệ giữa đầu vào và đầu ra. Các dữ liệu này được gọi là dữ liệu huấn luyện và chúng là cặp các dữ liệu đầu vào/đầu ra. Học có giám sát sẽ xem xét các tập huấn luyện này để từ đó có thể đưa ra dự đoán đầu ra cho một đầu vào mới chưa gặp bao giờ. Ví dụ dự đoán giá nhà, phân loại email,...
- **Học không giám sát:** Khác với học có giám sát, học không giám sát sử dụng những dữ liệu chưa được gán nhãn từ trước để suy luận. Phương pháp này thường được sử dụng để tìm cấu trúc của tập dữ liệu. Tuy nhiên lại không có phương pháp đánh giá được cấu trúc tìm ra được là đúng hay sai. Ví dụ như phân cụm dữ liệu, triết xuất thành phần chính của một chất nào đó.
- **Học tăng cường:** Phương pháp học tăng cường tập trung vào việc làm sao để cho một tác tử trong môi trường có thể hành động sao cho lấy được phần thưởng nhiều nhất có thể. Khác với học có giám sát nó không có cặp dữ liệu gán nhãn trước làm đầu vào và cũng không có đánh giá các hành động là đúng hay sai.

<sup>1</sup><http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf>



Hình A.2: Phương pháp máy học

### A.4.3 Xây dựng sản phẩm ứng dụng học máy

#### A.4.3.1 Các giai đoạn trong quy trình xây dựng sản phẩm

Một sản phẩm ứng dụng học máy cần trải qua 5 bước chính:

- Thu thập dữ liệu thô
- Tiền xử lý dữ liệu
- Lựa chọn thuật toán học máy
- Huấn luyện và đánh giá mô hình
- Triển khai mô hình

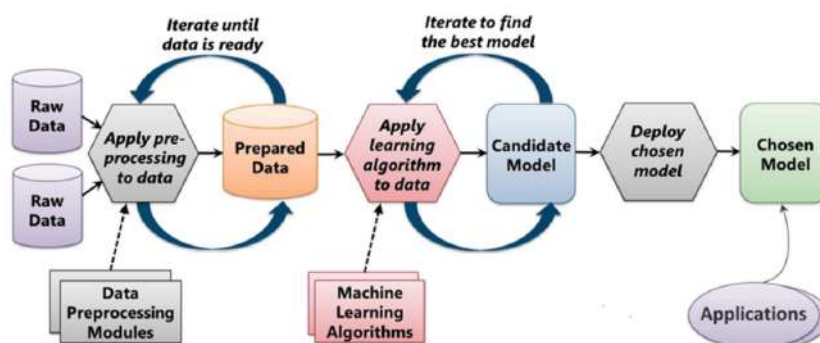
Các bước được minh họa như hình [A.3](#)

#### A.4.3.2 Dữ liệu

Bất cứ một bài toán học máy nào cũng đều cần có dữ liệu để huấn luyện, ta có thể coi nó là điều kiện tiên quyết. Dữ liệu sau khi có được cần phải:

- **Chuẩn hóa:** Tất cả các dữ liệu đầu vào đều cần được chuẩn hoá để máy tính có thể xử lý được. Quá trình chuẩn hoá bao gồm số hoá dữ liệu, co giãn thông số cho phù hợp với bài toán. Việc chuẩn hoá này ảnh hưởng trực tiếp tới tốc độ huấn luyện cũng như cả hiệu quả huấn luyện.

## The Machine Learning Process



From "Introduction to Microsoft Azure" by David Chappell

Hình A.3: Các bước để có một sản phẩm ứng dụng học máy

- **Phân chia:** Việc mô hình được chọn rất khớp với tập dữ liệu đang có không có nghĩa là giả thuyết của ta là đúng mà có thể xảy ra tình huống dữ liệu thật lại không khớp. Vấn đề này trong học máy được gọi là quá khớp (Overfitting). Vì vậy khi huấn luyện người ta phải phân chia dữ liệu ra thành 3 loại để có thể kiểm chứng được phần nào mức độ tổng quát của mô hình.
- **Tập huấn luyện** (Train set ): Dùng để học khi huấn luyện.
- **Tập kiểm chứng** (Validation set): Dùng để kiểm chứng mô hình khi huấn luyện.
- **Tập kiểm tra** (Test set): Dùng để kiểm tra xem mô hình đã phù hợp chưa sau khi huấn luyện.

Tập kiểm tra ta phải lọc riêng ra và không được sử dụng trong quá trình huấn luyện. Còn tập huấn luyện và tập kiểm chứng thì nên xáo trộn đối cho nhau để mô hình của ta được huấn luyện với các mẫu ngẫu nhiên nhất có thể.

### A.4.4 Các vấn đề trong học máy

Khi ta xây dựng một mô hình thì sẽ thường xảy ra các hiện tượng không mong muốn như mô hình dự đoán không chính xác, mô hình chỉ chính xác trên tập huấn luyện hay giá trị mất mát quá lớn,... Để tránh các trường hợp trên xảy ra, người xây dựng mô hình cần nắm được một số khái niệm cũng như kỹ thuật để giải quyết khi các vấn đề trên xảy ra.

- **Underfitting** là hiện tượng mô hình chưa được phù hợp với tập dữ liệu huấn luyện và cả các mẫu mới khi dự đoán. Nguyên nhân có thể là do mô hình chưa đủ độ phức tạp

cần thiết để bao quát được tập dữ liệu.

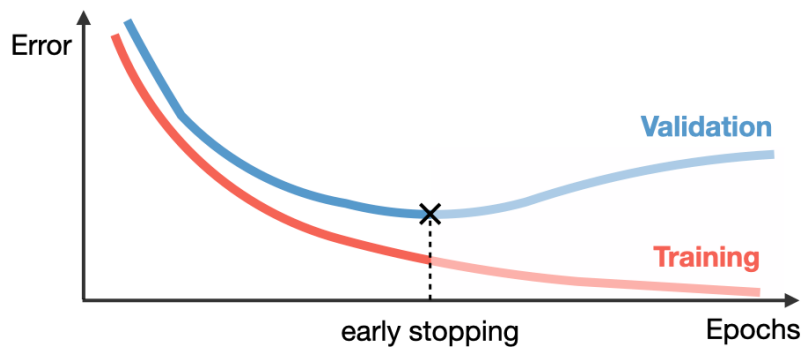
- **Overfitting** là hiện tượng mô hình quá khớp với tập dữ liệu huấn luyện (training set), việc này sẽ gây ra hậu quả vô cùng nghiêm trọng nếu tập dữ liệu huấn luyện xuất hiện 13nhiều. Mô hình sẽ chỉ chú trọng vào việc xấp xỉ với tập dữ liệu huấn luyện mà quên đi mục đích ban đầu là tổng quát hóa, làm cho mô hình sẽ không thật sự tốt đối với dữ liệu nằm ngoài dữ liệu huấn luyện (dữ liệu test và dữ liệu thực tế). Overfitting xảy ra khi độ phức tạp của mô hình quá lớn hoặc quá ít dữ liệu.
- **Good fitting** là mô hình nằm giữa 2 mô hình chưa khớp (underfitting) và quá khớp (overfitting) cho ra kết quả hợp lý với cả tập dữ liệu huấn luyện và các giá trị mới, tức là nó mang được tính tổng quát như hình 1 ở giữa phía trên. Lý tưởng nhất là khớp được với nhiều dữ liệu mẫu và cả các dữ liệu mới. Tuy nhiên trên thực tế được mô hình như vậy rất hiếm.

#### A.4.5 Phương pháp giải quyết

Các phương pháp để giải quyết các vấn đề trên nói chung là làm thay đổi nhỏ các tham số hay kiến trúc mô hình, chấp nhận hy sinh độ chính xác trong quá trình huấn luyện để giảm độ phức tạp của mô hình giúp tránh được hiện tượng không mong muốn mà vẫn giữ được tính tổng quát của mô hình. Một số kỹ thuật như: early stopping, thêm số hạng vào hàm mất mát, drop-out,...

##### A.4.5.1 Early stopping

Khi ta dùng một phương pháp tối ưu hàm số để giảm thiểu giá trị mất mát thì  $J_{train}$ ,  $J_{val}$  sẽ cùng giảm theo thời gian nhưng nếu sau một thời gian  $J_{val}$  tăng lên còn  $J_{train}$  tiếp tục giảm thì đó là lúc bắt đầu dẫn đến *overfitting*. Cách đơn giản nhất để giảm thiểu *overfitting* đó là dừng huấn luyện tại ngay thời điểm bắt đầu *overfitting* và phương pháp này được gọi là **early stopping**<sup>2</sup>



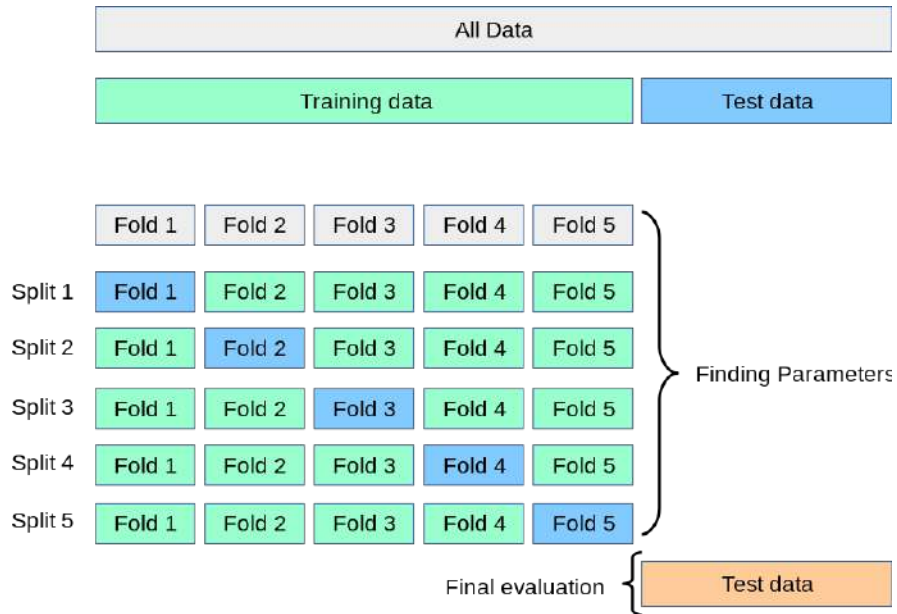
Hình A.4: Early stopping

<sup>2</sup>[https://en.wikipedia.org/wiki/Early\\_stopping](https://en.wikipedia.org/wiki/Early_stopping)

### A.4.5.2 Đánh giá chéo K-Fold

Đánh giá chéo  $k$ -fold, có tên tiếng anh là  $k$ -fold cross-validation, thường được áp dụng khi chúng ta có ít dữ liệu để huấn luyện và đánh giá làm cho mô hình khó có thể trích xuất được đặc trưng.

Để giải quyết vấn đề này, chúng ta sẽ chia dữ liệu thành  $k$  phần bằng nhau như hình A.5, được gọi là *fold*. Một phần sẽ được làm tập đánh giá,  $k - 1$  phần còn lại sẽ làm tập huấn luyện. Sau mỗi vòng lặp tập đánh giá sẽ được chuyển sang một tập khác.



Hình A.5: K-Fold

### A.4.5.3 Tăng cường dữ liệu

Phương pháp tăng cường dữ liệu (Data Augmentation) là một kỹ thuật quan trọng trong học sâu, đặc biệt khi áp dụng cho các bài toán thị giác máy tính như phân loại ảnh y tế. Mục tiêu của phương pháp này là tạo ra nhiều biến thể khác nhau từ các mẫu dữ liệu gốc bằng cách áp dụng các phép biến đổi ngẫu nhiên nhưng có chủ đích. Điều này giúp mở rộng tập dữ liệu mà không cần phải thu thập thêm dữ liệu thực tế, điều thường gặp nhiều khó khăn trong các bài toán y tế. Tăng cường dữ liệu không chỉ giúp cải thiện hiệu quả học tập của mô hình mà còn góp phần giảm thiểu hiện tượng quá khớp, đặc biệt trong các tình huống tập dữ liệu có kích thước nhỏ hoặc tính đa dạng thấp. Ví dụ, bài báo [41] là một trong nhiều bài báo giới thiệu việc sử dụng phương pháp tăng cường dữ liệu để cải thiện hiệu suất của mô hình học sâu. Lợi ích của tăng cường dữ liệu như sau:

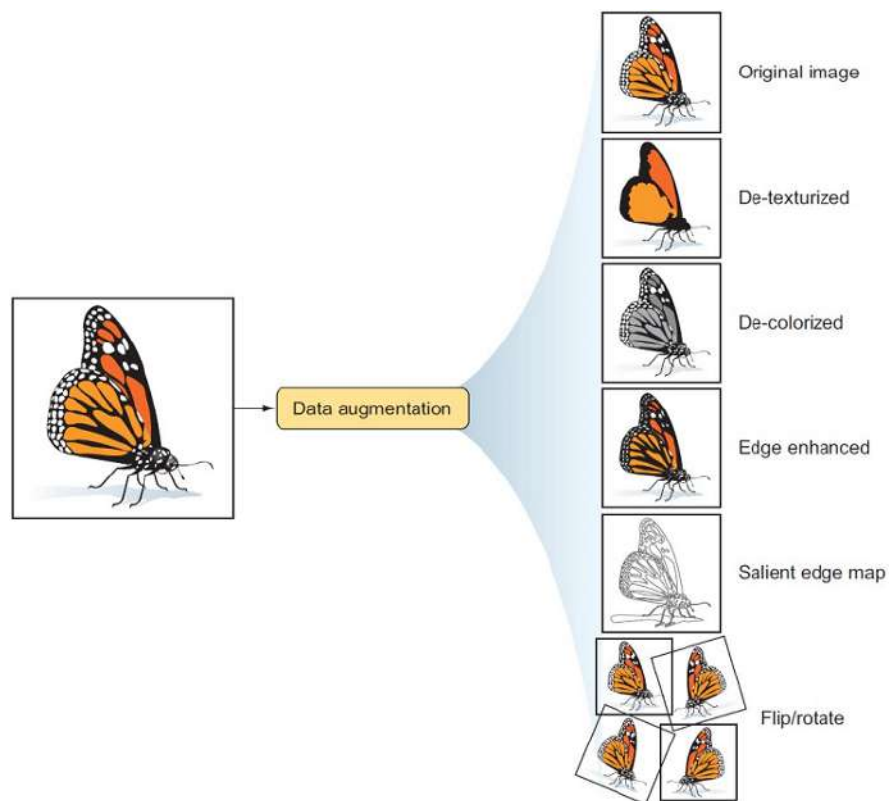
- Cải thiện khả năng tổng quát hóa: Bằng cách tạo ra nhiều biến thể của cùng một hình ảnh, mô hình học được các đặc trưng đa dạng và phong phú hơn, giúp cải thiện khả năng tổng quát hóa trên dữ liệu chưa từng thấy.

- Giảm yêu cầu thu thập dữ liệu mới: Dữ liệu y tế chất lượng cao thường khó thu thập và tốn kém. Do đó, tăng cường dữ liệu cung cấp một giải pháp hiệu quả để tăng cường số lượng mẫu mà không phải mở rộng tập dữ liệu thực tế.
- Tăng khả năng chống nhiễu: Việc huấn luyện mô hình trên các hình ảnh đã được biến đổi giúp mô hình ít nhạy cảm hơn với các nhiễu hoặc biến dạng ngẫu nhiên có thể xuất hiện trong quá trình chụp hoặc xử lý ảnh y tế.

Tăng cường dữ liệu (Data augmentation) là tạo thêm dữ liệu từ dữ liệu có sẵn.

- Giả sử ta có tập dữ liệu nhưng số lượng quá ít sẽ khiến cho mô hình của ta không học được các đặc trưng do đầu vào không đủ. Như vậy ta cần phải tìm cách làm tăng dữ liệu lên.
- Nếu ta đã có lượng lớn dữ liệu, nhưng chúng lại mất cân bằng. Có nghĩa là sự chênh lệch số lượng đầu vào lớn giữa các lớp với nhau. Khi đó mô hình sẽ ưu tiên những lớp mà có số lượng phần tử nhiều hơn. Ta cần phải tìm cách để dữ liệu chúng ta được cân bằng.

Với hai trường hợp điển hình trên, thì ta sẽ làm tăng cường dữ liệu thông qua một số thủ thuật như: co giãn kích thước ảnh, dịch chuyển ảnh, tăng giảm độ sáng, thêm độ nhiễu cho ảnh, ... như hình minh họa [A.6](#)

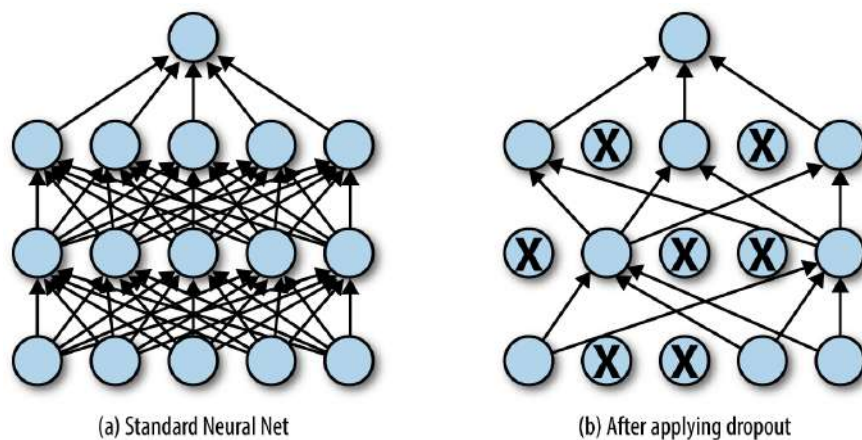


Hình A.6: Data augmentation

#### A.4.5.4 Drop-out

Dropout là một kỹ thuật phổ biến và hiệu quả trong việc ngăn chặn hiện tượng quá khớp (overfitting) trong mạng nơ-ron sâu. Phương pháp này được giới thiệu bởi Geoffrey Hinton và cộng sự vào năm 2012 [42], và nhanh chóng trở thành một công cụ mạnh mẽ trong quá trình huấn luyện các mô hình học sâu.

Drop-out<sup>3</sup> là một kỹ thuật Regularization để chống lại vấn đề **overfitting**. Cách dropout thực hiện là xóa bỏ một số unit trong các step training ứng với một giá trị xác suất  $p$  cho trước. Các mạng mới sau khi áp dụng dropout được gọi là subsample. Thông thường xác suất ở layer input bằng 0.8 hay ta loại bỏ khoảng 20% số unit, ở hidden layer thì xác suất là 0.5 có nghĩa là ta loại bỏ 50% số unit ở layer sử dụng dropout. Ta có ví dụ như hình A.7



Hình A.7: Drop-out

Ý tưởng chính của dropout là trong mỗi lần cập nhật trọng số trong quá trình huấn luyện, một số lượng nhất định các nơ-ron (neuron) trong mạng nơ-ron sẽ được "tạm thời" loại bỏ ngẫu nhiên, nghĩa là chúng sẽ không tham gia vào quá trình tính toán trong lượt forward và backward propagation. Việc này tạo ra một mạng con nhỏ hơn từ mạng chính, giúp mô hình tránh được việc phụ thuộc quá mức vào một số nơ-ron hoặc đường dẫn cụ thể, đồng thời giúp mạng học các đặc trưng tốt hơn và tổng quát hơn trên dữ liệu chưa thấy trước.

#### Tác động của Dropout đối với Mô hình

Sử dụng dropout giúp mô hình trở nên ít phụ thuộc vào các nơ-ron cụ thể, từ đó giảm thiểu việc ghi nhớ các mẫu huấn luyện và cải thiện khả năng tổng quát hóa đối với các mẫu dữ liệu mới. Trong bài toán phân loại ảnh y tế, điều này đặc biệt quan trọng, bởi vì số lượng dữ liệu thường hạn chế, dẫn đến nguy cơ cao về overfitting.

Qua quá trình thử nghiệm, việc thêm dropout vào mô hình ResNet50 đã giúp cải thiện hiệu suất của mô hình trên tập kiểm tra, đồng thời giảm sự chênh lệch giữa độ chính xác trên tập huấn luyện và tập xác thực. Dropout đã góp phần làm tăng khả năng dự đoán chính

<sup>3</sup>[https://en.wikipedia.org/wiki/Dilution\\_\(neural\\_networks\)](https://en.wikipedia.org/wiki/Dilution_(neural_networks))



xác của mô hình, đặc biệt khi kết hợp với các kỹ thuật khác như tăng cường dữ liệu (data augmentation) và chuẩn hóa đầu vào (preprocessing input).

Dropout là một phương pháp hiệu quả để giảm overfitting và cải thiện khả năng tổng quát hóa của các mô hình học sâu, đặc biệt trong các bài toán phân loại ảnh y tế với số lượng dữ liệu hạn chế. Bằng cách vô hiệu hóa ngẫu nhiên một số nơ-ron trong quá trình huấn luyện, dropout giúp mô hình trở nên mạnh mẽ hơn và ít phụ thuộc vào các đặc trưng cụ thể của tập huấn luyện, từ đó tăng độ chính xác và hiệu quả trên tập kiểm tra.

### ***Cách hoạt động của dropout***

Dropout hoạt động theo nguyên tắc đơn giản: với xác suất  $p$  (thường gọi là "dropout rate"), một số nơ-ron được chọn ngẫu nhiên và bị loại bỏ tạm thời trong quá trình huấn luyện. Các nơ-ron bị loại bỏ sẽ không tham gia vào quá trình tính toán giá trị đầu ra của lớp và cũng không tham gia vào việc điều chỉnh trọng số. Khi mô hình hoàn thành quá trình huấn luyện và chuyển sang pha kiểm tra hoặc dự đoán (inference), tất cả các nơ-ron sẽ được sử dụng, nhưng với trọng số được điều chỉnh theo tỷ lệ tương ứng với xác suất dropout để đảm bảo sự nhất quán trong kết quả đầu ra. Dropout hoạt động với nguyên tắc chi tiết như sau:

- Dropout được áp dụng trên một layer của mạng neural networks với một xác suất  $p$  cho trước (có thể sử dụng nhiều Drop-Out khác nhau cho những layer khác nhau, nhưng trên 1 layer sẽ chỉ có 1 dropout).
- Tại mỗi step trong quá trình training, khi thực hiện feedforward đến layer sử dụng dropout, thay vì tính toán tất cả unit có trên layer, tại mỗi unit ta "gieo xúc xắc" với xác suất  $p$  xem unit đó được tính (active) hay không được tính (deactive). Những unit active ta tính toán bình thường còn với những unit deactivate thì ta set giá trị tại unit đó bằng 0.
- Trong quá trình test thì tất cả các unit đều được active và chúng ta mong muốn đầu ra của các units giống với đầu ra mong đợi trong quá trình training. Ví dụ đầu ra của một unit (trước khi dropout) là  $a$ , khi áp dụng dropout thì đầu ra mong đợi của unit đó sẽ là  $pa + (1 - p)0$ , vì unit bị deactivate thì giá trị của unit đó là 0. Do đó trong quá trình test chúng ta điều chỉnh đầu ra  $a \rightarrow pa$  để giống với đầu ra mong đợi.

Thời gian test khá là quan trọng nên nếu chúng ta điều chỉnh đầu ra ở các layer áp dụng dropout thì hiệu suất test sẽ bị giảm đi. Vì thế thay vì chỉnh trong quá trình test thì chúng ta sẽ thực hiện việc này trong quá trình training. Ta sẽ lấy *dropout mask* (vector xác suất được khởi tạo ngẫu nhiên, tại vị trí có giá trị nhỏ hơn  $p$  sẽ được giữ nguyên còn lớn hơn  $p$  sẽ set lại giá trị vị trí đó là 0) chia cho  $p$  trong quá trình training. Trường hợp này được gọi là *inverted dropout*.



### A.4.6 Phương pháp đánh giá mô hình

Khi xây dựng một mô hình học máy, chúng ta cần một phép đánh giá để xem mô hình sử dụng có hiệu quả không và để so sánh khả năng của các mô hình. Hiệu năng của mô hình thường được đánh giá dựa trên tập dữ liệu kiểm thử (test data) do không chứa dữ liệu đã được huấn luyện. Cụ thể, giả sử đầu ra của mô hình khi đầu vào là tập kiểm thử là một vector dự đoán đầu ra với mỗi phần tử là nhãn được dự đoán của một điểm dữ liệu trong tập kiểm thử. Ta sẽ so sánh giữa vector dự đoán vừa tính được với vector chữ nhãn thật của dữ liệu. Một số metrics đánh giá mô hình được tổng hợp như tổng hình A.8

		True condition		Prevalence = $\frac{\sum \text{Condition positive}}{\sum \text{Total population}}$	Accuracy (ACC) = $\frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$
		Condition positive	Condition negative		
Predicted condition	Predicted condition positive	<b>True positive</b> , Power	<b>False positive</b> , Type I error	Positive predictive value (PPV), Precision = $\frac{\sum \text{True positive}}{\sum \text{Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\sum \text{False positive}}{\sum \text{Predicted condition positive}}$
	Predicted condition negative	<b>False negative</b> , Type II error	<b>True negative</b>	False omission rate (FOR) = $\frac{\sum \text{False negative}}{\sum \text{Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\sum \text{True negative}}{\sum \text{Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\sum \text{True positive}}{\sum \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\sum \text{False positive}}{\sum \text{Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR-}}$
		False negative rate (FNR), Miss rate $= \frac{\sum \text{False negative}}{\sum \text{Condition positive}}$	True negative rate (TNR), Specificity (SPC) $= \frac{\sum \text{True negative}}{\sum \text{Condition negative}}$	Negative likelihood ratio (LR-) $= \frac{\text{FNR}}{\text{TNR}}$	

Hình A.8: Các metrics đánh giá mô hình

#### A.4.6.1 Định nghĩa cơ bản

Giả sử trong một bài toán chẩn đoán bệnh.

- *Positive (P)*: Số lượng kết quả chẩn đoán là dương tính thực tế.
- *Negative (N)*: Số lượng kết quả chẩn đoán là âm tính thực tế.
- *True positive (TP)*: Số lượng kết quả chẩn đoán đúng là dương tính (thực tế cũng là dương tính) trên toàn bộ mẫu kiểm tra.
- *False positive (FP)*: Số lượng kết quả chẩn đoán sai là dương tính (thực tế là âm tính) trên toàn bộ mẫu kiểm tra.
- *True negative (TN)*: Số lượng kết quả chẩn đoán đúng là âm tính (thực tế cũng là âm tính) trên toàn bộ mẫu kiểm tra.

- *False negative* (FN): Số lượng kết quả chuẩn đoán sai là âm tính (thực tế là dương tính) trên toàn bộ mẫu kiểm tra.

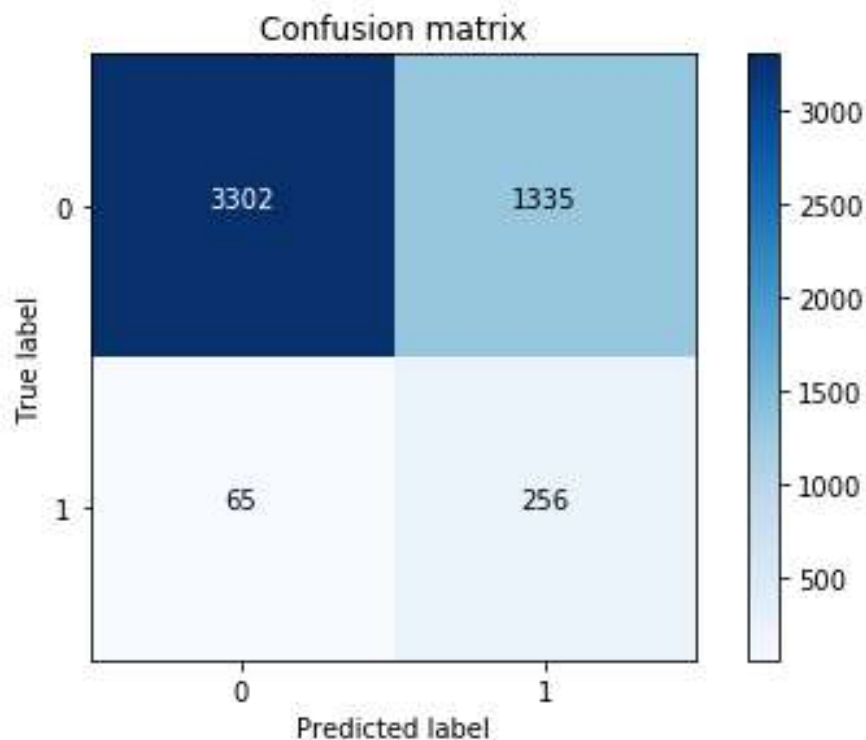
#### A.4.6.2 Accuracy

Cách này đơn giản, tự nhiên nhất và hay được sử dụng nhất cho dữ liệu cân bằng. Cách đánh giá này đơn giản tính tỉ lệ giữa số lượng các dữ liệu được dự đoán đúng và tổng số dữ liệu trong tập dữ liệu kiểm thử.

$$\text{Accuracy} = \frac{\text{Tổng số dữ liệu đầu vào dự đoán đúng}}{\text{Tổng số dữ liệu đầu vào}} = \frac{TP + TN}{TP + FP + TN + FN}$$

#### A.4.6.3 Confusion matrix - Ma trận nhầm lẫn

Cách tính sử dụng accuracy như ở trên chỉ cho chúng ta biết được bao nhiêu phần trăm lượng dữ liệu được phân loại đúng mà không chỉ ra được cụ thể mỗi loại được phân loại như thế nào, lớp nào được phân loại đúng nhiều nhất, và dữ liệu thuộc lớp nào thường bị phân loại nhầm vào lớp khác. Để có thể đánh giá được các giá trị này, chúng ta sử dụng một ma trận được gọi là ma trận nhầm lẫn hay confusion matrix. Về cơ bản, confusion matrix thể hiện có bao nhiêu điểm dữ liệu thực sự thuộc vào một class, và được dự đoán là rơi vào một class.



Hình A.9: Ví dụ về ma trận nhầm lẫn

Nhìn vào hình A.9, ta có thể biết rằng tổng số dữ liệu kiểm tra cho lớp '0' là 4637, số lượng đoán chính xác là 3302, 1335 là số lượng bị đoán sai và đoán vào lớp '1'. Đối với lớp

'1' thì có tổng số dữ liệu kiểm tra là 321, trong đó đoán chính xác là 256, đoán nhầm vào lớp '0' là 65.

#### A.4.6.4 Sensitivity, recall, hit rate, hoặc true positive rate (TPR)

**Sensitivity** (Độ nhạy) đo tỷ lệ chẩn đoán đúng dương tính với trên toàn bộ mẫu kiểm tra dương tính thật.

$$\text{sensitivity} = \text{recall} = \text{TPR} = \frac{TP}{P} = \frac{TP}{TP + FN}$$

#### A.4.6.5 Specificity, selectivity hoặc true negative rate (TNR)

**Specificity** (độ đặc hiệu) đo tỷ lệ chuẩn đoán đúng âm tính trên toàn bộ mẫu kiểm tra âm tính thật.

$$\text{specificity} = \text{TNR} = \frac{TN}{N} = \frac{TN}{TN + FP}$$

#### A.4.6.6 Precision hoặc positive predictive value (PPV)

**Precision** (độ chính xác) đo tỷ lệ chuẩn đoán đúng dương tính trên toàn bộ kết quả trả về là dương tính.

$$\text{precision} = \frac{TP}{TP + FP}$$

#### A.4.6.7 F-measure

**F-measure** là một thước đo kết hợp giữa **precision** và **recall**, trả về một giá trị thống nhất.

$$F_{\beta} = (1 + \beta^2) * \frac{\text{precision} * \text{recall}}{\beta^2 * \text{precision} + \text{recall}}$$

Với bài toán bình thường,  $\beta$  thường được đặt là 1. Ở đây nếu dùng  $\beta = 2$  là để muốn ưu tiên *recall* giảm độ ảnh hưởng của *precision* nhằm một kết quả **Thà nhận nhầm còn hơn bỏ sót** phù hợp với bài toán trong y tế, kí hiệu là F2.

#### A.4.6.8 ROC Curve và AUC

Đường cong ROC <sup>4</sup> (Receiver Operating Characteristic Curve) là một biểu đồ cho thấy hiệu suất của mô hình phân loại ở tất cả các ngưỡng phân loại. Đường cong này vẽ hai tham số:

- Tỷ lệ True Positive (TPR, sensitivity)

<sup>4</sup><https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>

- Tỷ lệ False Positive (FPR, 1- specificity)

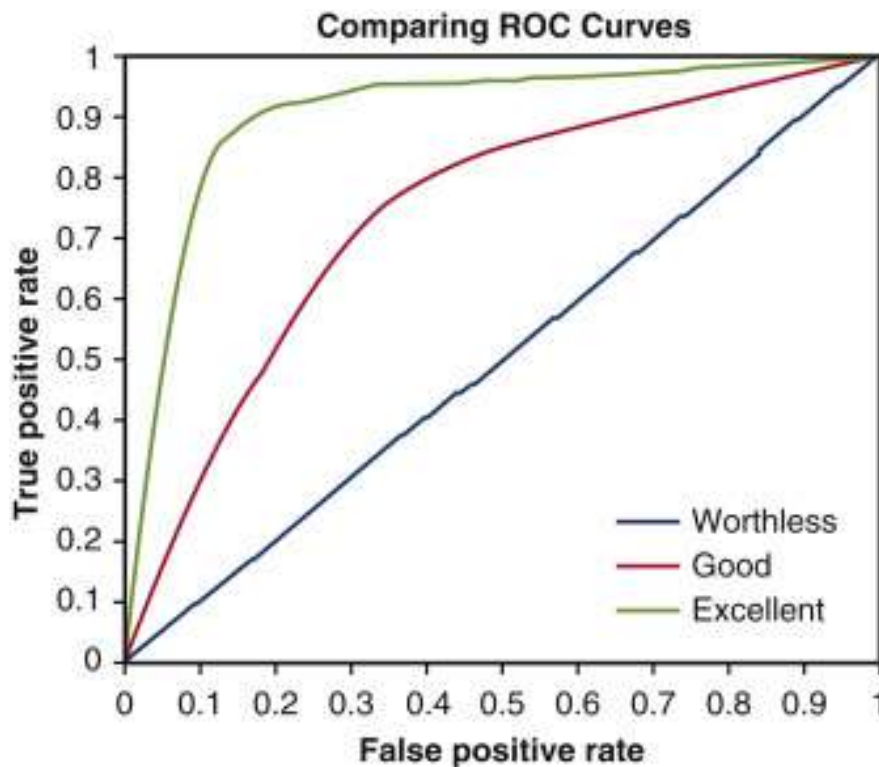
Với TPR được tính như sau:

$$\text{sensitivity} = TPR = \frac{TP}{TP + FN}$$

FPR được tính như sau:

$$1 - \text{specificity} = FPR = \frac{FP}{FP + TN}$$

Đường cong ROC vẽ đồ thị TPR so với FPR ở các ngưỡng phân loại khác nhau (như hình A.10). Việc hạ thấp ngưỡng phân loại sẽ phân loại nhiều hơn theo hướng tích cực, do đó làm tăng cả FN và TP.

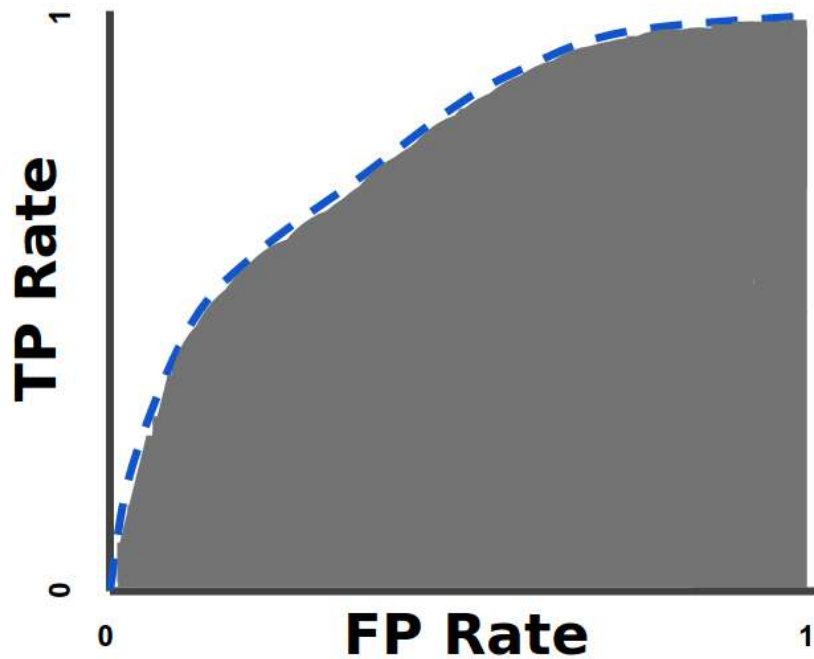


Hình A.10: Đường cong ROC

Đường cong ROC được sử dụng để so sánh các mô hình, đường càng cong càng tốt.

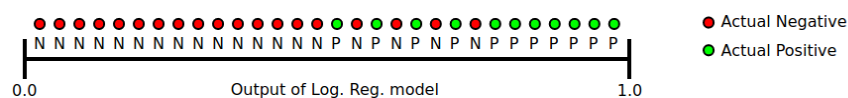
Để tính các điểm trong một đường cong ROC, chúng ta có thể đánh giá mô hình hồi quy logistic nhiều lần với các ngưỡng phân loại khác nhau, nhưng điều này sẽ không hiệu quả. Tuy nhiên, có một thuật toán dựa trên thuật toán sắp xếp có thể cung cấp thông tin này, được gọi là AUC (Area Under the ROC Curve)<sup>5</sup>. AUC đo toàn bộ khu vực bên dưới toàn bộ đường cong ROC (tính toán tích phân) từ (0,0) đến (1,1) (hình minh họa A.11).

<sup>5</sup><https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>



Hình A.11: AUC

AUC cung cấp một thước đo tổng hợp về hiệu suất trên tất cả các ngưỡng phân loại có thể. Một cách giải thích AUC là xác suất mà mô hình xếp hạng một mẫu positive ngẫu nhiên cao hơn một mẫu negative ngẫu nhiên. Ví dụ, đưa ra các mẫu sau (hình A.12), được sắp xếp từ trái sang phải theo thứ tự tăng dần theo dự đoán hồi quy logistic:



Hình A.12: Các dự đoán được xếp hạng theo thứ tự tăng dần của điểm hồi quy logistic

AUC đại diện cho xác suất rằng một mẫu positive (màu xanh lá cây) ngẫu nhiên được đặt ở bên phải của mẫu negative (màu đỏ) ngẫu nhiên. AUC có giá trị từ 0 đến 1. Một mô hình có dự đoán sai 100% thì AUC là 0.0; ngược lại, có dự đoán đúng 100% thì AUC là 1.0.

#### A.4.6.9 Dice Index

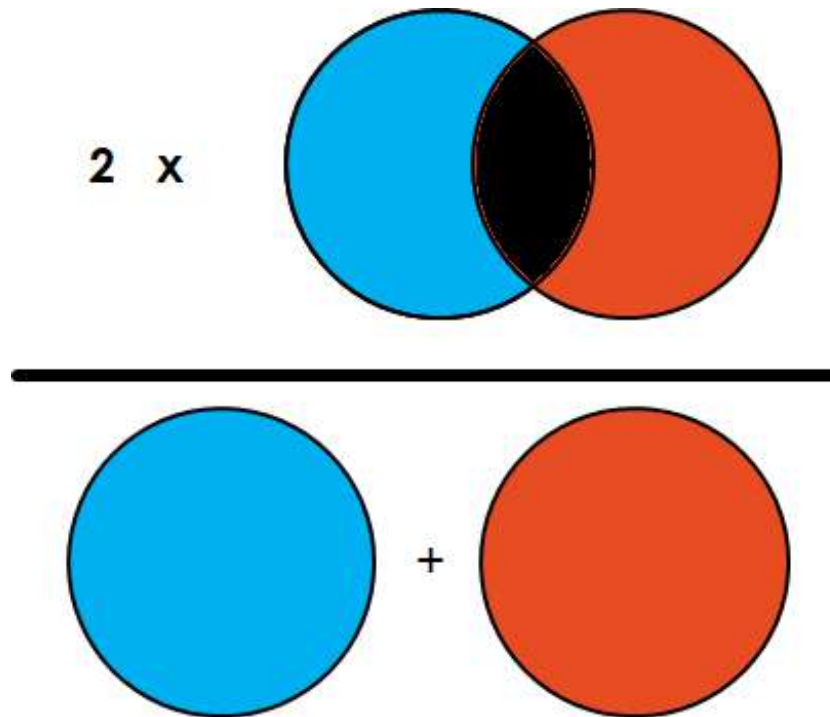
**Dice Index - Dice Coefficient** là hệ số được dùng phổ biến trong các bài toán *semantic segmentation*. Metric này thể hiện độ khớp của output của mô hình và nhãn trên một ảnh.

Công thức toán học của Dice Index là

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

ở đây  $X$  và  $Y$  là 2 tập cần so sánh. Trên dữ liệu ảnh, *Dice Coefficient* bằng 2 lần số pixel

giống nhau chia cho tổng số pixel có trong cả 2 ảnh. Các tính Dice Index được minh họa như hình ??



Hình A.13: Cách tính Dice Index

**Dice Index** mang đến một giá trị cân bằng cho tính toán độ hiệu quả của mô hình segmentation.

#### A.4.7 Giới thiệu các chỉ số mAP và LAMR

Trong các bài toán về phân đoạn ảnh và phát hiện đối tượng, hai chỉ số quan trọng để đánh giá hiệu suất của mô hình là *Mean Average Precision (mAP)* và *Log-Average Miss Rate (LAMR)*. Đây là những chỉ số đánh giá tổng quan giúp đo lường khả năng phát hiện, phân đoạn chính xác và khả năng bỏ sót đối tượng trong tập dữ liệu. Phần này sẽ trình bày chi tiết cách tính và ý nghĩa của từng chỉ số, cùng với các tài liệu trích dẫn hỗ trợ.

##### A.4.7.1 Mean Average Precision (mAP)

**Mean Average Precision (mAP)** là một chỉ số được sử dụng phổ biến trong các bài toán phát hiện và phân đoạn đối tượng. mAP được tính dựa trên kết hợp giữa Precision và Recall. Để tính toán chỉ số mAP, trước tiên ta cần xác định **Average Precision (AP)** cho từng lớp đối tượng. Quá trình này được thực hiện qua các bước sau:

- Tính **Precision** tại mỗi ngưỡng *Recall*:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (\text{A.1})$$

với  $TP$  là số lượng đối tượng phát hiện đúng (True Positives) và  $FP$  là số lượng phát hiện sai (False Positives).

- Tính **Recall** tại từng ngưỡng:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (\text{A.2})$$

với  $FN$  là số đối tượng bị bỏ sót (False Negatives).

- Vẽ đường cong Precision-Recall và tính diện tích dưới đường cong để xác định giá trị AP cho từng lớp đối tượng.

Cuối cùng, chỉ số mAP được tính bằng cách lấy trung bình giá trị AP của tất cả các lớp đối tượng:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (\text{A.3})$$

với  $N$  là số lượng lớp đối tượng.

Chỉ số mAP cung cấp cái nhìn tổng quát về khả năng của mô hình trong việc phát hiện và phân đoạn các vùng bất thường. Một giá trị mAP cao cho thấy mô hình có khả năng phát hiện chính xác và đầy đủ đối tượng. mAP thường được sử dụng trong các cuộc thi lớn về phát hiện đối tượng như *PASCAL VOC* và *COCO* [43], [44].

#### A.4.7.2 Log-Average Miss Rate (LAMR)

**Log-Average Miss Rate (LAMR)** là một chỉ số khác được sử dụng trong các bài toán phát hiện đối tượng, tập trung vào việc đánh giá khả năng phát hiện sai sót của mô hình. LAMR đo lường mức độ bỏ sót các đối tượng thực sự có trong ảnh và được tính toán dựa trên biểu đồ giữa *False Positive Per Image (FPPI)* và *Miss Rate (MR)*.

Các bước tính LAMR bao gồm:

- Tính **Miss Rate** tại mỗi ngưỡng:

$$\text{Miss Rate} = \frac{FN}{FN + TP} \quad (\text{A.4})$$

- Vẽ biểu đồ Miss Rate theo False Positive Per Image (FPPI), với  $\text{FPPI} = \frac{FP}{\text{Số lượng ảnh}}$ .

- Tính giá trị log trung bình của Miss Rate (log-average) trên các mức FPPI:

$$\text{LAMR} = \exp\left(\frac{1}{k} \sum_{i=1}^k \log(\text{MR}(\text{FPPI}_i))\right) \quad (\text{A.5})$$

trong đó  $k$  là số lượng các ngưỡng FPPI.

Giá trị LAMR càng thấp cho thấy mô hình càng ít bỏ sót các đối tượng quan trọng. Chỉ số này thường được sử dụng để đánh giá các mô hình phát hiện đối tượng trong các tình huống mà việc bỏ sót đối tượng gây hậu quả nghiêm trọng, chẳng hạn như trong lĩnh vực an ninh hoặc y học [45].

#### A.4.7.3 Ý nghĩa của mAP và LAMR trong bài toán phân đoạn ảnh y học

Trong lĩnh vực phân đoạn ảnh y học, chỉ số mAP và LAMR đóng vai trò bổ sung cho nhau. Chỉ số mAP giúp đánh giá khả năng của mô hình trong việc phát hiện và phân đoạn đúng các vùng bất thường, trong khi LAMR đo lường khả năng giảm thiểu các lỗi bỏ sót đối tượng. Việc kết hợp cả hai chỉ số này sẽ mang lại cái nhìn tổng quan về hiệu suất của mô hình, đảm bảo rằng mô hình không chỉ có độ chính xác cao mà còn phát hiện đầy đủ các vùng tổn thương, tránh bỏ sót các đối tượng quan trọng.

### A.4.8 Một số hàm tính lỗi phổ biến

#### A.4.8.1 Hàm sai số bình phương trung bình

Hàm sai số toàn phương trung bình hay mean squared error (MSE), là hàm tiếp cận dễ nhất để hiểu hàm mất mát. Nó được định nghĩa là trung bình của bình phương các sai số giữa giá trị ước lượng được và thực tế.

Giả sử  $\hat{\mathbf{y}}$  là một vector đầu ra dự đoán, và  $\mathbf{y}$  là vector đầu ra thực tế quan sát được, tương ứng với các dữ liệu đầu vào, thì MSE của phép dự báo có thể ước lượng theo công thức:

$$\text{MSE} = \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{y}})^2$$

#### A.4.8.2 Hàm cross-entropy

Cross entropy <sup>6</sup> được dùng để đo sự giống nhau của hai phân phối xác suất, giá trị của hàm số càng nhỏ thì hai xác suất càng gần nhau và thường dùng cho các bài toán phân loại nhiều nhãn. Giả sử ta có một phân bố xác suất  $\mathbf{p} = [p_1, p_2, \dots, p_n]$  với  $p_i \in [0, 1]$  và  $\sum_{i=1}^n p_i = 1$ . Nếu ta có một phân bố xác suất bất kỳ  $\mathbf{q} = [q_1, q_2, \dots, q_n]$  và  $q_i \neq 0, \forall i$  thì hàm số cross entropy giữa hai phân bố  $\mathbf{p}$  và  $\mathbf{q}$  được định nghĩa là:

<sup>6</sup>[https://en.wikipedia.org/wiki/Cross\\_entropy](https://en.wikipedia.org/wiki/Cross_entropy)



$$H(\mathbf{p}, \mathbf{q}) = - \sum_{i=1}^n p_i \log q_i$$

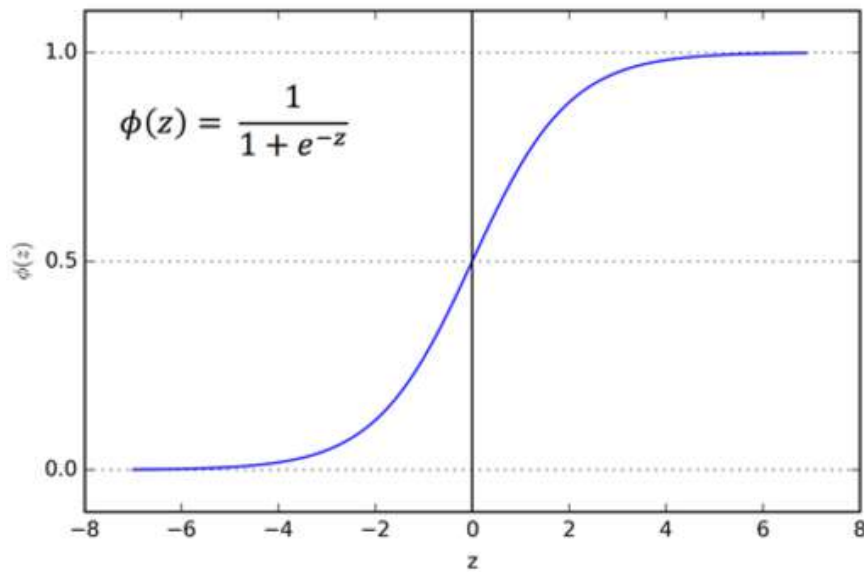
**Chú ý:** Hàm cross entropy không có tính đối xứng  $H(\mathbf{p}, \mathbf{q}) \neq H(\mathbf{q}, \mathbf{p})$ . Theo công thức trên chúng ta có thể thấy giá trị của  $\mathbf{q}$  không thể nhận giá trị là 0. Vì thế khi sử dụng cross entropy trong các bài toán học có giám sát, chúng ta phải để  $\mathbf{p}$  là đầu ra thực tế vì chỉ có vị trí nhãn là được đánh dấu 1, các vị trí còn lại được đánh dấu 0 (sử dụng one-hot encoding),  $\mathbf{q}$  là đầu ra dự đoán vì không có xác suất nào bằng 0 tuyệt đối.

## A.4.9 Một số hàm kích hoạt phổ biến

### A.4.9.1 Hàm sigmoid

Hàm sigmoid là hàm phi tuyến tính có đồ thị một dạng đường cong hình dạng chữ "S" như hình dưới và có công thức toán học là

$$f(x) = \frac{1}{1 + e^x}$$



Hình A.14: Hàm Sigmoid

Sự tăng trưởng của đồ thị gồm giai đoạn tăng trưởng ban đầu được xấp xỉ hàm mũ và khi quá trình bão hòa bắt đầu, sự phát triển sẽ chậm lại, và tới giai đoạn trưởng thành thì dừng hẳn.

Giá trị của hàm số luôn nằm trong khoảng  $(0, 1]$ , cụ thể hơn là với đầu vào lớn hàm số sẽ cho đầu ra gần với 1 còn với đầu vào nhỏ hàm số sẽ cho đầu ra gần với 0. Và đạo hàm của hàm sigmoid rất đẹp.

$$f'(x) = f(x)(1 - f(x))$$

#### A.4.9.2 Hàm softmax

Hàm softmax <sup>7</sup> hay hàm trung bình mũ là sự khái quát hóa của hàm sigmoid biến không gian K-chiều véc tơ với giá trị thực bất kỳ đến không gian K-chiều véc tơ mang giá trị trong phạm vi (0, 1). Hàm softmax có phương trình toán học như sau:

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}; \forall i = 1, \dots, n \text{ với } n \text{ là số lượng phần tử trong vector } \mathbf{x}$$

Hàm softmax có công dụng trong việc phân loại tập dữ liệu. Vì nếu tại giá trị phần tử  $x_i$  lớn vượt trội so với toàn bộ dữ liệu ở vector  $\mathbf{x}$  thì giá trị đầu ra  $y_i$  cũng sẽ lớn vượt trội so với đầu ra ở các phần tử khác. Do tổng giá trị đầu ra của hàm softmax luôn bằng một, luôn dương và mỗi đầu ra đều phụ thuộc vào tất cả các đầu vào nên ta có thể coi giá trị đầu ra thể hiện xác suất của dữ liệu rơi vào từng tập dữ liệu tương ứng.

#### A.4.9.3 Hàm ReLU

Hàm ReLU (rectified linear unit function) là hàm phi tuyến tính đơn giản nhất, có công thức toán học là

$$f(s) = \max(0, s)$$

Ta thấy hàm ReLU có đầu ra là 0 nếu đầu vào nhỏ hơn hoặc bằng 0, và đầu ra bằng đầu vào trong trường hợp ngược lại. Hàm ReLU có đạo hàm như sau:

$$f'(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{otherwise} \end{cases}$$

Hàm ReLU thường được sử dụng làm hàm truyền trong các hidden layer, còn ở layer cuối cùng thì ta sẽ sử dụng hàm khác để có thể tính toán được sắc xuất dự đoán vào vùng phân loại.

#### A.4.10 Một số tham số huấn luyện - Model hyperparameter

Trước hết, ta hiểu về model parameter là các giá trị của mô hình được sinh ra từ dữ liệu huấn luyện giúp thể hiện mối liên hệ giữa các đại lượng trong dữ liệu. Model hyperparameter khác với model parameter. Nó là các tham số được sử dụng trong quá trình huấn luyện, giúp mô hình tìm ra được các model parameter hợp lý nhất. Nó thường được lựa chọn thủ công bởi những người tham gia trong việc huấn luyện mô hình.

<sup>7</sup>[https://en.wikipedia.org/wiki/Softmax\\_function](https://en.wikipedia.org/wiki/Softmax_function)

Một vài model hyperparameter thường dùng trong học máy như: learning rate, batch size, epoch, iteration,...

- *learning rate*: Tốc độ học hay độ dốc của quá trình đạo hàm hàm tối ưu Gradient Descent.
- *batch size*: Số lượng mẫu mà thuật toán tối ưu Mini-batch Gradient Descent sử dụng cho một lần cập nhật trọng số.
- *epoch*: Mỗi epoch là mỗi lần duyệt qua hết số lượng mẫu trong tập huấn luyện.
- *iteration*: Là số lượng batch mà thuật toán phải duyệt trong một epoch.

### A.4.11 Một vài phương pháp tối ưu

Mục tiêu của bài toán tối ưu là tìm ra nghiệm sao cho tại đó hàm số đạt giá trị nhỏ nhất *global minimum* hoặc hàm số đạt giá trị lớn nhất *global maximum*. Tuy nhiên, việc tìm *global minimum* hay *global maximum* của các hàm số là rất phức tạp, thậm chí là bất khả thi. Thay vào đó, người ta thường cố gắng tìm các điểm lân cận của chúng ở một mức độ nào đó và coi đó là nghiệm cần tìm của bài toán.

Giả sử ta có  $N$  quan sát  $(\mathbf{x}, \mathbf{y})$  có ánh xạ  $\mathbf{x} \rightarrow \mathbf{y}$  với  $f(\theta, \mathbf{x}) = \mathbf{y}$ . Ta cần tìm *global minimum* cho hàm  $f(\theta; \mathbf{x}, \mathbf{y})$  trong đó  $\theta$  là một vector,  $\theta = [\theta_1, \theta_2, \dots, \theta_m]$ . Ta có một số phương pháp giúp ta giải quyết bài toán đó là: gradient descent, stochastic gradient descent, adam, rmsprop,...

Gradient của hàm số  $f$  tại một điểm  $\theta$  bất kỳ được ký hiệu là  $\Delta\theta f(\theta; \mathbf{x}, \mathbf{y})$

Thuật toán gradient descent giúp ta tìm  $\theta$  sao cho  $f(\theta, \mathbf{x})$  gần  $\mathbf{y}$  nhất.

**Algorithm 1** Gradient descent

- Require: Tập  $N$  quan sát  $(\mathbf{x}, \mathbf{y})$
- Require:  $\theta = [\theta_1, \theta_2, \dots, \theta_m]$ 
  - repeat
    - \* for all  $i = 1; i \leq N; i++$  do
      - $\Delta\theta_i := -\Delta_{\theta} f(\theta; \mathbf{x}^{(i)}, \mathbf{y}_i)$
    - \* end for
    - \*  $\Delta\theta = \frac{1}{N} \sum_{i=1}^N \Delta\theta_i$
    - \* Chọn bước nhảy  $\eta$
    - \* Cập nhật  $\theta := \theta - \eta\Delta\theta$
  - until thuật toán hội tụ

Trong đó  $\eta$  (đọc là eta) là một số dương được gọi là learning rate (tốc độ học) và giá trị của learning rate thường là 0.001. Việc lựa chọn learning rate rất quan trọng trong các bài toán thực tế. Việc lựa chọn giá trị này phụ thuộc nhiều vào từng bài toán và phải làm một vài thí nghiệm để chọn ra giá trị tốt nhất. Và dấu trừ tại  $\theta := -\Delta_{\theta}f(\mathbf{x})$  thể hiện việc chúng ta phải đi **ngược với đạo hàm** (Đây cũng chính là lý do phương pháp này được gọi là Gradient Descent - descent nghĩa là đi ngược).

Nếu dữ liệu có kích thước  $N$  lớn thì mỗi lần cập nhật  $\theta$  đòi hỏi chúng ta sử dụng tất cả các quan sát  $\mathbf{x}_i$  do đó khối lượng tính toán lớn do phải tính đạo hàm trên toàn bộ dữ liệu, thuật toán chạy chậm. Do vậy để tiết kiệm khối lượng tính toán, chúng ta sẽ cập nhật tính toán sau mỗi dữ liệu quan sát, phương pháp này gọi là stochastic gradient descent (SGD)

**Algorithm 2** Stochastic gradient descent

- Require: Tập  $N$  quan sát  $(\mathbf{x}, \mathbf{y})$
- Require:  $\theta = [\theta_1, \theta_2, \dots, \theta_m]$ 
  - repeat
    - \* Xáo trộn dữ liệu
    - \* for all  $i = 1; i \leq N; i++$  do
      - $\Delta\theta_i := -\Delta_{\theta}f(\theta; \mathbf{x}^{(i)}, \mathbf{y}_i)$
      - Chọn bước nhảy  $\eta$
      - Cập nhật  $\theta := \theta - \eta\Delta\theta_i$
      - if hội tụ then break
      - end if
    - \* end for
  - until thuật toán hội tụ

Khác với SGD, thay vì mỗi \*iteration\* ta tính toán trên một quan sát thì ta sẽ tính toán với  $k$  quan sát ( $1 < k \ll N$ ). Phương pháp này được gọi là *mini-batch gradient descent*.

**Algorithm 3** Mini-batch Gradient descent

- Require: Tập  $N$  quan sát  $(\mathbf{x}, \mathbf{y})$
- Require:  $\theta = [\theta_1, \theta_2, \dots, \theta_m]$ 
  - repeat
    - \* Xáo trộn dữ liệu
    - \* for all  $i = 1; i \leq N; i++$  do
      - $\Delta\theta_i := -\sum_{j=i}^{i+k} \Delta_{\theta}f(\theta; \mathbf{x}^{(j)}, \mathbf{y}_j)$
      - Chọn bước nhảy  $\eta$

- Cập nhật  $\theta := \theta - \eta \Delta \theta_i$
- if hội tụ then break
- end if
- \* end for
- until thuật toán hội tụ

**Adam** (Adaptive Moment Estimation) <sup>8</sup> là một bản cập nhật được đưa ra gần đây. Cách cập nhật của thuật toán Adam như sau:

$$m_t := \beta_1 m_{t-1} + (1 - \beta_1) \Delta \theta_t$$

$$v_t := \beta_2 v_{t-1} + (1 - \beta_2) \Delta \theta_t^2$$

$$\hat{m}_t := \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t := \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t := \theta_t + \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t$$

$m_t$ ,  $v_t$  là giá trị ước lượng giữa thời điểm trước và thời điểm sau của các gradient tương ứng. Các giá trị của  $m_t$ ,  $v_t$  đều được khởi tạo bằng 0. Các tác giả của Adam thấy rằng chúng bị lệch về 0, đặc biệt là các bước lặp đầu. Do vậy họ tạo ra  $\hat{m}_t$  và  $\hat{v}_t$  để chống lại việc giá trị của  $m_t$  và  $v_t$  lệch về 0. Các giá trị của  $\beta_1$ ,  $\beta_2$  được tác giả đề xuất là 0.9 cho  $\beta_1$ , 0.999 cho  $\beta_2$  và  $10^{-8}$  cho  $\epsilon$ .

## A.5 Mạng Nơ-ron nhân tạo

### A.5.1 Khái niệm

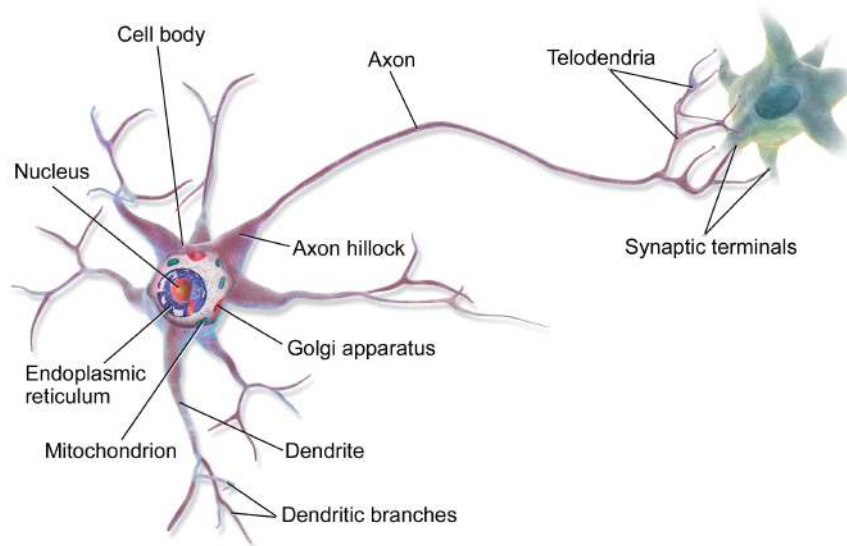
**Mạng nơ-ron nhân tạo** (ANN) hay hệ thống kết nối là các hệ thống máy tính được lấy cảm hứng mô phỏng từ các mạng nơ-ron sinh học cấu thành não động vật. Bản thân mạng nơ-ron không phải là một thuật toán, mà là một khuôn khổ cho nhiều thuật toán học máy khác nhau hoạt động cùng nhau và xử lý các đầu vào dữ liệu phức tạp. Các hệ thống như vậy "học" để thực hiện các nhiệm vụ bằng cách xem xét các ví dụ, nói chung mà không được lập trình với bất kỳ quy tắc cụ thể cho nhiệm vụ.

ANN dựa trên một tập hợp các đơn vị hoặc nút được kết nối được gọi là tế bào thần kinh nhân tạo, mô hình hóa lỏng lẻo các tế bào thần kinh trong não sinh học. Mỗi kết nối, giống như khớp thần kinh trong bộ não sinh học (hình A.16), có thể truyền tín hiệu từ nơ-ron nhân tạo này sang nơ-ron nhân tạo khác. Một tế bào thần kinh nhân tạo nhận được tín hiệu có thể xử lý nó và sau đó báo hiệu các tế bào thần kinh nhân tạo bổ sung được kết nối với nó.

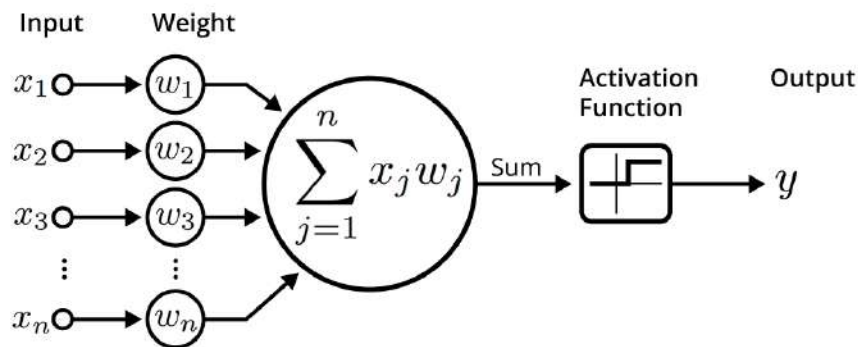
Ở hình A.16:

- Đầu vào (input): là các tín hiệu vào của nơ-ron, các tín hiệu này thường đưa vào dưới dạng một vector, kí hiệu là  $\mathbf{x}$ .

<sup>8</sup><https://arxiv.org/abs/1412.6980>



Hình A.15: Nơ-ron thần kinh



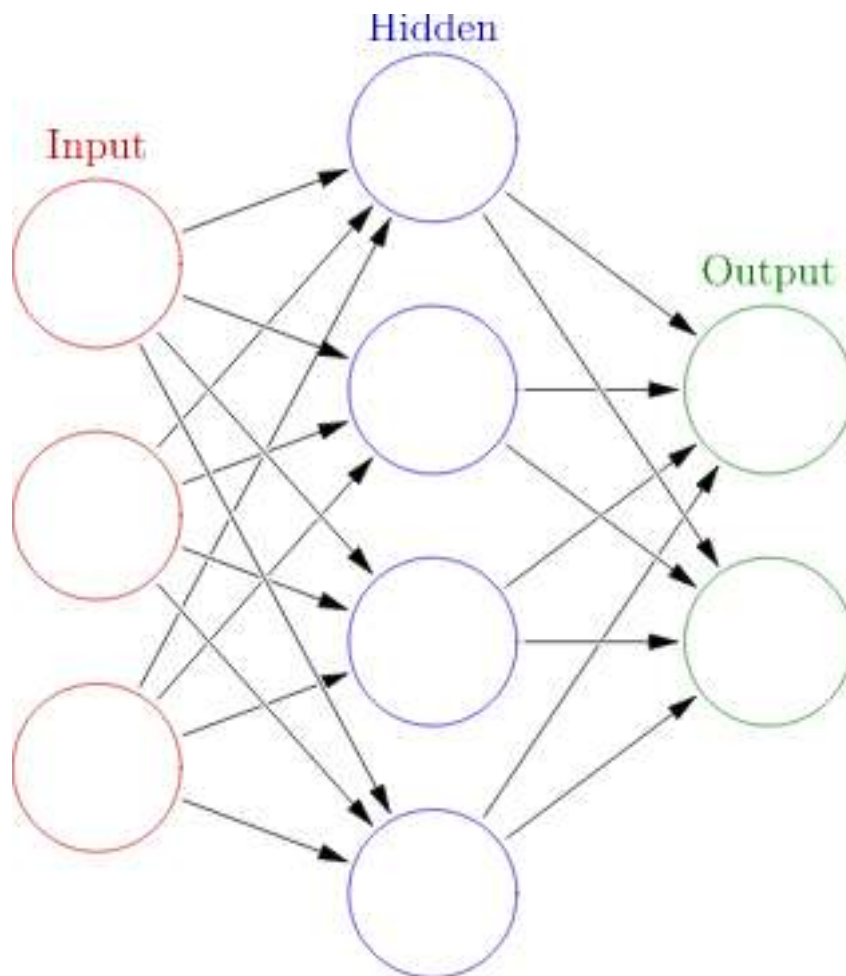
An illustration of an artificial neuron. Source: Becoming Human.

Hình A.16: Nơ-ron thần kinh nhân tạo

- Trọng số (weight): mỗi liên kết được thể hiện bởi một trọng số liên kết và thường được kí hiệu  $w$ . Thông thường, các trọng số này được khởi tạo một cách ngẫu nhiên theo phân phối chuẩn ở thời điểm khởi tạo mạng và được cập nhật liên tục trong quá trình học mạng.
- Hàm kích hoạt (activation function): Hàm này được dùng để giới hạn phạm vi đầu ra của mỗi nơ-ron. Nó nhận đầu vào là kết quả của hàm kết hợp và ngưỡng đã cho. Thông thường hàm kích hoạt sẽ là các hàm phi tuyến tính. Có rất nhiều hàm kích hoạt được sử dụng cho tới nay như: sigmoid, tanh, ReLU, softmax,...
- Đầu ra (output): là tín hiệu đầu ra của một nơ-ron, với mỗi nơ-ron sẽ có tối đa là một đầu ra. Nếu như nơ-ron đó ở các hidden layer thì đầu ra của nó được gọi là activation và được kí hiệu là  $a$ .

Trong các triển khai ANN thông thường, tín hiệu tại kết nối giữa các nơ-ron nhân tạo là một số thực và đầu ra của mỗi nơ-ron nhân tạo được tính bằng một số hàm phi tuyến tính của

tổng các đầu vào của nó. Các kết nối giữa các nơ-ron nhân tạo được gọi là các cạnh. Các tế bào thần kinh nhân tạo và các cạnh thường có **weight** điều chỉnh khi quá trình học tập diễn ra. **Weight** làm tăng hoặc giảm cường độ của tín hiệu tại một kết nối. Tế bào thần kinh nhân tạo có thể có một ngưỡng sao cho tín hiệu chỉ được gửi đi nếu tín hiệu tổng hợp vượt qua ngưỡng đó. Thông thường, các tế bào thần kinh nhân tạo được tập hợp thành từng lớp. Các lớp khác nhau có thể thực hiện các loại biến đổi khác nhau trên đầu vào của chúng. Tín hiệu đi từ lớp đầu tiên ( lớp đầu vào), đến lớp cuối cùng (lớp đầu ra ), có thể sau khi đi qua các lớp bên trong nhiều lần (hình minh họa A.17).



Hình A.17: Mạng lưới thần kinh nhân tạo

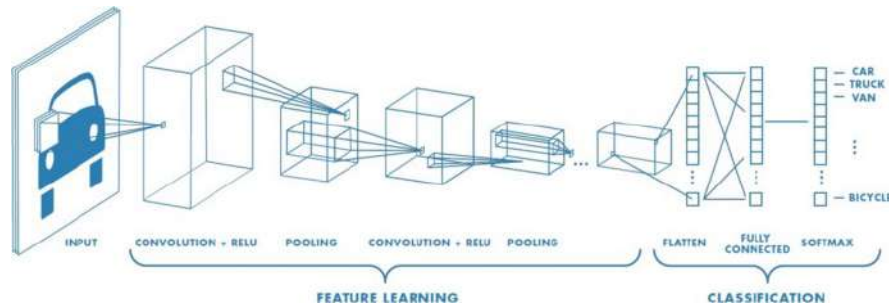
Một **Perceptron nhiều lớp** (MLP) là một lớp học của mạng nơ-ron nhân tạo feedforward. Một MLP bao gồm ít nhất ba lớp nút: lớp đầu vào, lớp ẩn và lớp đầu ra. Ngoại trừ các nút đầu vào, mỗi nút là một nơ-ron sử dụng chức năng kích hoạt phi tuyến. MLP sử dụng một kỹ thuật học tập có giám sát được gọi là backpropagation để đào tạo. Nhiều lớp và kích hoạt không tuyến tính của nó phân biệt MLP với một perceptron tuyến tính. Nó có thể phân biệt dữ liệu không thể phân tách tuyến tính.

## A.6 Mạng nơ-ron tích chập

Mạng nơ-ron tích chập (**Convolutional Neural Network-CNN**) là một trong những mô hình deep learning được thiết kế để trích xuất đặc trưng dữ liệu ảnh  $n$  chiều (hình minh họa A.18).

Tương tự như việc con người học cách nhận diện đối tượng, chúng ta cần cho thuật toán học rất nhiều hình ảnh trước khi nó có thể đưa ra dự đoán cho hình ảnh đầu vào mà nó chưa từng thấy. Máy tính nhìn theo cách khác con người. Trong thế giới máy tính chỉ có những con số. Mỗi hình ảnh có thể được biểu diễn dưới dạng mảng  $n$  chiều, những phần tử trong mảng được gọi là các điểm ảnh. Một bức ảnh dùng làm đầu vào được biểu diễn bởi ma trận các điểm ảnh (thường từ 0 đến 255 cho ảnh 8-bit) với kích thước  $[w * h * c]$ ; trong đó  $w$  là số lượng điểm ảnh trên chiều ngang,  $h$  là số lượng điểm ảnh trên chiều dọc,  $c$  gọi là kênh màu.

Về mặt kỹ thuật, mô hình CNN học sâu để đào tạo và kiểm tra, mỗi hình ảnh đầu vào sẽ chuyển nó qua một loạt các lớp tích chập (convolution layers) với các bộ lọc (Kernels), Pooling, các lớp được kết nối đầy đủ (fully connected layers-FC) và áp dụng hàm Softmax để phân loại một đối tượng có giá trị xác suất giữa 0 và 1. Hình dưới đây là toàn bộ luồng CNN để xử lý hình ảnh đầu vào và phân loại các đối tượng dựa trên các giá trị.



Hình A.18: Convolutional Neural Network-CNN

### A.6.1 Tầng tích chập - Convolutional layer

Tầng tích chập, tên tiếng anh là Convolution layer<sup>9</sup>, là khối cốt lõi, cơ bản của ConvNets, làm mạng nơ-ron tích chập trở nên khác biệt so với mạng nơ-ron truyền thống và hoạt động cực kỳ hiệu quả trong bài toán phân tích ảnh.

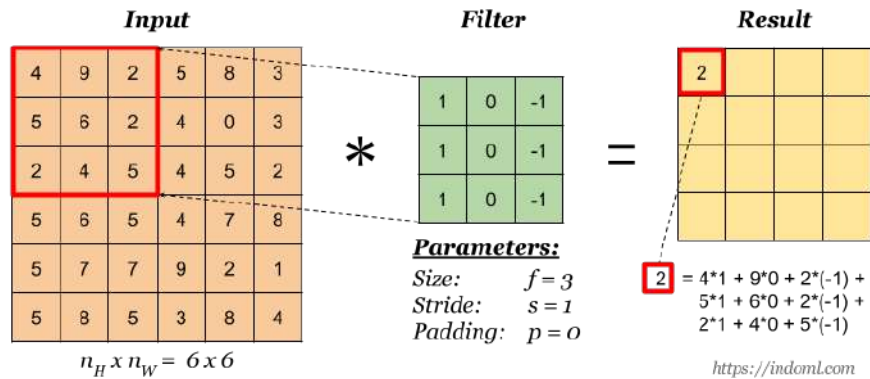
Tầng tích chập lấy dữ liệu đầu vào, thực hiện các phép chuyển đổi để tạo ra dữ liệu đầu vào cho tầng kế tiếp (đầu ra của tầng này là đầu vào của tầng sau). Phép biến đổi được sử dụng là phép tính tích chập. Mỗi tầng tích chập chứa một hoặc nhiều bộ lọc - bộ phát hiện đặc trưng (*filter - feature detector*) cho phép phát hiện và trích xuất những đặc trưng khác nhau của ảnh.

Thay vì kết nối toàn bộ điểm ảnh, tầng này sẽ sử dụng bộ lọc (*filter*) áp vào một vùng trong ảnh có kích thước bằng với bộ lọc và tiến hành tính *element-wise* giữa bộ lọc và vùng

<sup>9</sup>[stanford.edu/shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks](https://stanford.edu/shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks)



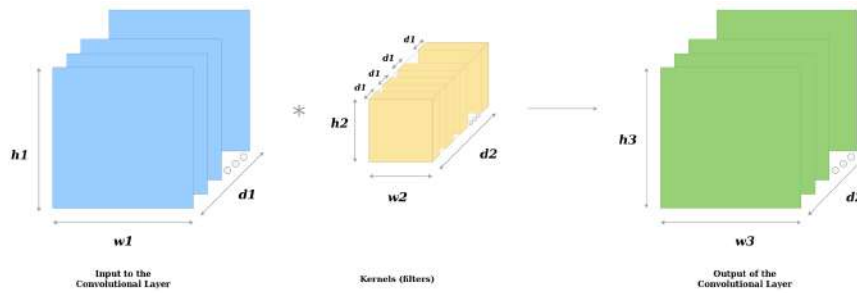
ảnh đó. Vùng ảnh này được gọi vùng tiếp nhận cục bộ (*local receptive filter*). Ta tiếp tục trượt bộ lọc để có thể quét hết các vùng trong ảnh. Mỗi bước trượt sẽ cho ra một giá trị duy nhất và tập đầu ra này được gọi là *feature map*. Khi đó ta thu được kết quả sẽ trích xuất ra các đặc trưng của ảnh, có thể là bộ lọc góc, cạnh, đường chéo, hình tròn, hình vuông,.... Tiếp tục làm như vậy cho các tầng tiếp theo thì các tầng tiếp theo sẽ lại trích xuất tiếp các đặc trưng của đặc trưng của các đối tượng đó, việc có nhiều layer như vậy cho phép chúng ta chia nhỏ đặc trưng của ảnh tới mức nhỏ nhất có thể. Hình A.19 là ảnh minh họa về Convolutional layer.



Hình A.19: Convolutional layer

- **Filter:** Bộ lọc hay còn gọi là filter hoặc kernel, giúp phát hiện và trích xuất các đặc trưng của ảnh, có thể là bộ lọc góc, cạnh, đường chéo, hình tròn, hình vuông. Kích thước của bộ lọc thường là các ma trận có kích thước nhỏ ( $3 \times 3, 5 \times 5, 7, \dots$ ) giúp ta thay đổi các giá trị đầu vào dựa vào các giá trị lân cận theo một nguyên tắc, công thức nào đó. Độ phức tạp của đặc trưng được phát hiện bởi bộ lọc tỉ lệ thuận với độ sâu của tầng tích chập mà nó thuộc về hay bộ lọc ở tầng tích chập càng sâu thì phát hiện các đặc trưng càng phức tạp.
- **Stride:** Bước trượt hay còn có tên tiếng anh là slide là khoảng cách dịch chuyển của bộ lọc sau mỗi lần tính và được trượt theo chiều dọc và chiều ngang. Các bước trượt là khoảng cách mà ta dịch chuyển filter trên ảnh đầu vào.
- **Padding:** Nó có tác dụng khắc phục được nhược điểm trên của bước trượt, giúp giữ lại được thông tin ở lề của ảnh. Thêm lề được thực hiện bằng cách ta thêm một đường viền phụ vào xung quanh ma trận đầu, việc này làm tăng kích thước của ma trận đầu vào, dẫn tới tăng kích thước ma trận đầu ra. Từ đó độ chênh lệch giữa ma trận đầu ra với ma trận đầu vào gốc giảm. Những ô nằm trên cạnh hoặc góc của ma trận đầu vào gốc cũng lùi sâu vào bên trong hơn, dẫn tới được sử dụng nhiều hơn trong việc tính toán ma trận đầu ra, tránh được việc mất mát thông tin.

Tích chập khối (hình A.20) cũng tương tự như tích chập hai chiều như phần trên đã trình bày nhưng có một điều quan trọng cần phải nhớ đó là độ sâu của bộ lọc phải bằng với độ sâu của đầu vào.



Hình A.20: Tích chập khối

Và chúng ta có công thức tính tổng quát cho tích chập khối:

- Kích thước đầu vào  $w_1 \times h_1 \times d_1$
- Kích thước bộ lọc  $w_2 \times h_2 \times d_1$
- Số bộ lọc  $d_2$
- Tốc độ trượt  $s$
- Thêm lề  $p$
- Kích thước đầu ra là

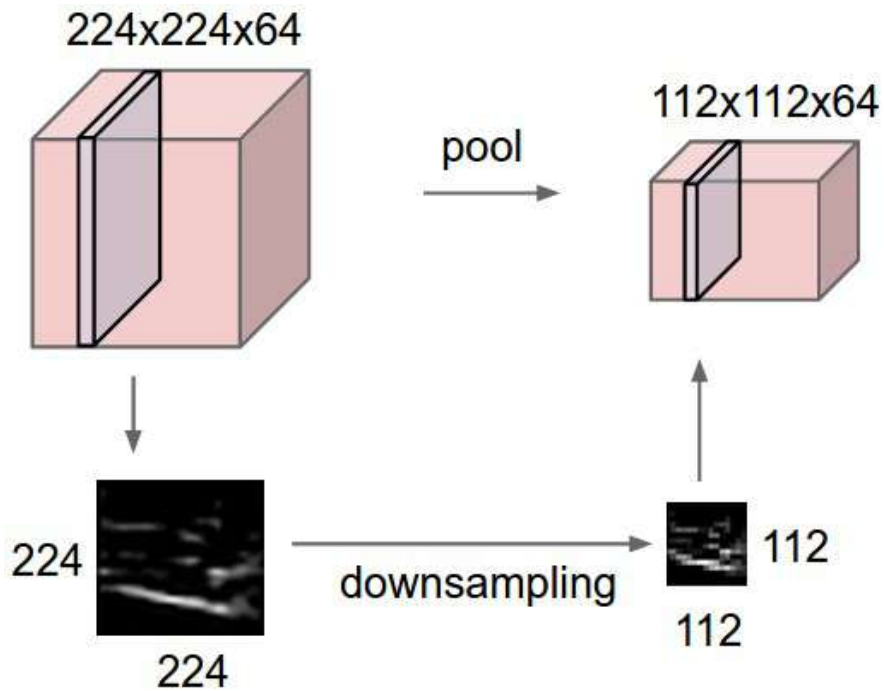
$$w_3 \times h_3 \times d_2 = \left( \frac{w_1 + 2p - w_2}{s} + 1 \right) \times \left( \frac{h_1 + 2p - h_2}{s} + 1 \right) \times d_2$$

## A.6.2 Tầng tổng hợp - Pooling layer

Tầng giảm số chiều *pooling layer* <sup>10</sup> trong mạng CNNs thực hiện công việc loại bỏ bớt những thông tin không cần thiết sau khi thực hiện tích chập và được chèn giữa các tầng convolutional với nhau hoặc sau một tập tầng convolutional. Nó có vai trò giảm kích thước dữ liệu. Với một bức ảnh kích thước lớn qua nhiều tầng pooling sẽ được thu nhỏ lại tuy nhiên vẫn giữ được những đặc trưng cần cho việc nhận dạng (thông qua cách lấy mẫu). Việc giảm kích thước dữ liệu sẽ làm giảm lượng tham số, tăng hiệu quả tính toán và góp phần kiểm soát hiện tượng quá khớp (*overfitting*). Tuy nhiên nếu lạm dụng loại layer này cũng có thể khiến data đi qua bị mất dữ liệu.

Cách thức hoạt động: Pooling layer sử dụng bộ lọc để trượt trên đầu vào nhưng không thực hiện element-wise, kết quả của mỗi bước trượt là một giá trị được xem là giá trị đại diện cho thông tin ảnh tại vùng đó (giá trị mẫu) được giữ lại và đó gọi là tiến hành lấy mẫu (*subsampling*). Các phương thức lấy phổ biến trong tầng giảm số chiều là *max pooling* (lấy giá trị lớn nhất), *min pooling* (lấy giá trị nhỏ nhất) và *average pooling* (lấy giá trị trung bình). Hình A.21 minh họa cho các hoạt động này.

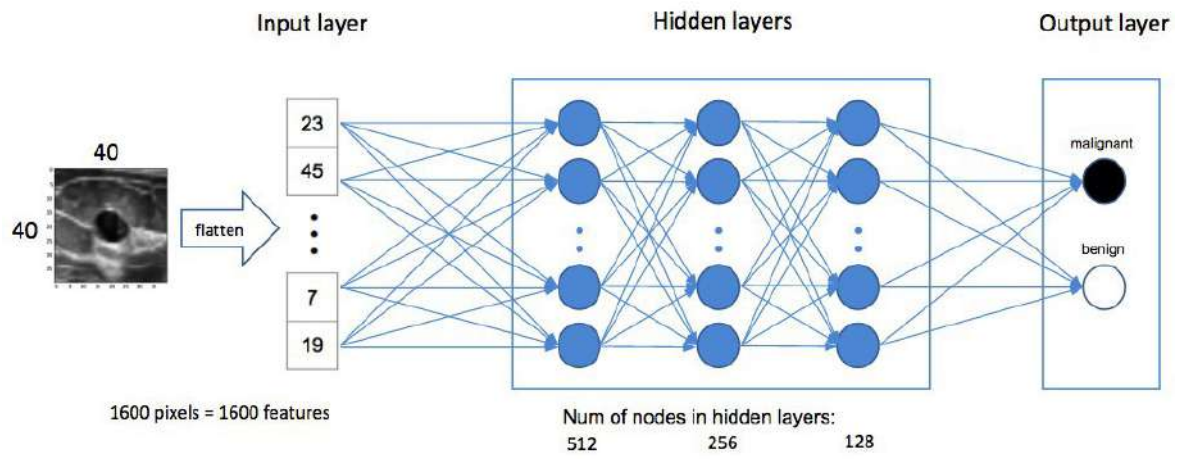
<sup>10</sup><https://cs231n.github.io/convolutional-networks>



Hình A.21: Pooling layer

### A.6.3 Tầng Fully-Connected (Fully-Connected layer)

Đầu ra từ các tầng chập và tầng tổng hợp thể hiện các đặc trưng cấp cao của ảnh đầu vào. Mục đích của tầng Fully-Connected là sử dụng các đặc trưng này để phân loại hình ảnh đầu vào thành các lớp khác nhau dựa trên tập dữ liệu huấn luyện. Bằng cách làm phẳng (flatten) đầu ra cuối cùng của giai đoạn trước đó và áp dụng mạng nơ-ron truyền thẳng với các thuật toán phân loại để phân loại đối tượng dựa vào vector ta vừa làm phẳng. Hay nói cách khác Fully-Connected chính là một mạng nơ-ron được gắn vào phần cuối của giai đoạn trích xuất đặc trưng với đầu vào là đầu ra của các tầng trước đó. Nó đóng vai trò như một mô hình phân tầng và tiến hành dựa trên dữ liệu đã được xử lý ở các tầng trước đó. Hình A.22 minh họa cho tầng này.



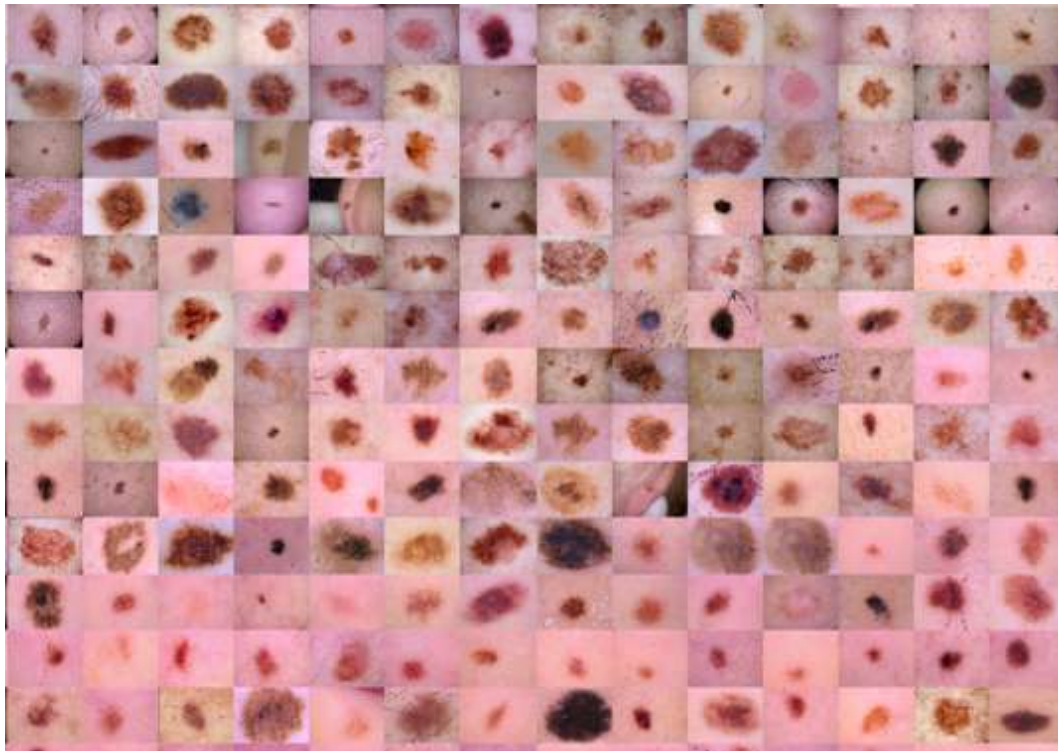
Hình A.22: Tầng Fully-Connected

# PHỤ LỤC B | Bộ dữ liệu HAM10000 và ISIC-2019

## B.1 Tổng quan

### B.1.1 HAM10000

HAM10000 (Human Against Machine with 10000 training images) là một bộ dữ liệu có khoảng 10,000 hình ảnh soi da cận cảnh, chất lượng cao về các tổn thương da nhằm phục vụ các nghiên cứu về Trí tuệ nhân tạo cho bài toán xác định bệnh da liễu. Bộ dữ liệu được cung cấp thu thập từ Khoa Da liễu tại Đại học Y khoa Vienna, Áo và phòng khám ung thư da của Cliff Rosendahl ở Queensland, Australia. Các nhà nghiên cứu đã sử dụng những phương pháp tự động hóa, học máy và đánh giá thử công để cho ra những hình ảnh chất lượng cao, màu sắc tương đồng, nhất quán, vị trí tổn thương ở giữa và miễn phí sử dụng. Chi tiết cụ thể về quy trình thu thập dữ liệu được tìm thấy trong bài báo của họ [46].



Hình B.1: Bộ dữ liệu HAM10000

Bộ dữ liệu gồm có 7 nhãn bệnh được ký hiệu như sau:

- ak: Actinic Keratoses (Solar Keratoses) - Dày sừng ánh sáng
- bcc: Basal cell carcinoma - Ung thư biểu mô tế bào đáy
- bkl: Benign keratosis - Dày sừng lành tính

- df: Dermatofibroma - U xơ da
- nv: Melanocytic nevus - Tế bào hắc tố
- mel: Melanoma - Khối u ác tính
- vasc: Vascular skin lesions - Tổn thương mạch máu

Các hình ảnh trong bộ dữ liệu HAM10000 được thu thập từ các bệnh nhân đến khám với các triệu chứng liên quan đến da và được chụp ở độ phân giải cao. Mỗi hình ảnh được gắn nhãn với loại bệnh lý da tương ứng.

Theo tác giả của HAM10000 [46], hơn 50% các tổn thương được xác nhận thông qua mô bệnh học (histo), nhãn thực tế của các trường hợp còn lại là kiểm tra, theo dõi (follow\_up), đồng thuận của chuyên gia (consensus), hoặc xác định bằng kính hiển vi in-vivo (confocal). Bộ dữ liệu cũng được bảo gồm các tổn thương có nhiều hình ảnh (một vùng tổn thương được chụp nhiều ảnh khác nhau), có thể được theo dõi bằng tệp metadata kèm theo. Không chỉ quan sát được ảnh thuộc tổn thương nào và nhãn bệnh tương ứng, tệp metadata cũng cung cấp thông tin về phương pháp xác định bệnh (histo, follow\_up, ...), giới tính, tuổi của bệnh nhân và vị trí da bị bệnh.

Bộ dữ liệu HAM10000 đã được sử dụng trong nhiều nghiên cứu về trí tuệ nhân tạo và học sâu, nhằm phát triển các mô hình học máy để phát hiện và phân loại các bệnh lý da. Các mô hình này có thể được sử dụng để hỗ trợ cho việc chẩn đoán và điều trị bệnh lý da. Google Scholar <sup>1</sup> chứa khoảng 2.5 nghìn kết quả khi tìm kiếm với từ khoá "HAM10000 dataset". Những nghiên cứu có thể kể đến như:

- Karar Ali cùng các cộng sự [47] đã đề xuất một mô hình mạng nơ-ron tích chập để dự đoán bệnh da liễu sử dụng bộ dữ liệu HAM10000 để đánh giá mô hình.
- Garg cùng các cộng sự [48] đã đề xuất một mô hình mạng tích chập chẩn đoán bệnh ngoài da với bộ dữ liệu HAM10000 để huấn luyện và đánh giá hiệu quả của mô hình.
- Faes cùng các cộng sự [49] cũng đã đề xuất một mô hình mạng nơ-ron tích chập để chẩn đoán bệnh ngoài da sử dụng bộ dữ liệu này đánh giá độ hiệu quả của mô hình.

HAM10000 được cung cấp cho những người tham gia thử thách phân loại ISIC-2018 [40] được tổ chức bởi hội nghị MICCAI hàng năm ở Granada, Tây Ban Nha. Vào năm 2019, thử thách ISIC-2019 được khởi động với bộ dữ liệu sử dụng đã bao gồm ISIC-2018 và thêm nhiều dữ liệu mới. Thông tin về nó sẽ được trình bày tại phần sau. Tóm lại, HAM10000 là một bộ dữ liệu phổ biến, nền tảng nhất đối với bài toán học máy về da liễu.

<sup>1</sup><https://scholar.google.com>

## B.1.2 ISIC-2019

Bộ dữ liệu nổi tiếng ISIC-2019 <sup>2</sup> cho da liễu có nguồn gốc từ một cuộc thi cùng tên được tổ chức bởi Hợp tác quốc tế về hình ảnh da (International Skin Imaging Collaboration - ISIC). Bộ dữ liệu bao gồm 2 bộ dữ liệu nổi tiếng khác:

- BCN\_20000 [50]: Bộ dữ liệu về hình ảnh soi da của các loại tổn thương da liễu phổ biến nhất như dày sừng quang sáng (actinic keratosis), ung thư biểu mô tế bào vảy (squamous cell carcinoma), dày sừng tiết bã (seborrheic keratosis), ...
- HAM10000 [46]: Bộ dữ liệu về hình ảnh soi gia đã được giới thiệu ở ??.

Bộ dữ liệu ISIC-2019 là một trong những bộ dữ liệu lớn và đa dạng nhất về các bệnh lý da, bao gồm nhiều loại bệnh lý da phổ biến và hiếm. Bộ dữ liệu này cung cấp cho các nhà nghiên cứu và chuyên gia y tế một tài nguyên quý giá để phát triển và kiểm tra các mô hình học máy và các công cụ trợ giúp chẩn đoán bệnh lý da. ISIC-2019 chứa tổng cộng 25,331 ảnh cho việc huấn luyện mô hình với 8 loại bệnh khác nhau. Tập kiểm tra chứa 8,238 ảnh nhưng nhãn của chúng không được công bố. Với 8 loại bệnh, ISIC-2019 có 9 lớp cần phân loại như sau.

- Melanoma (MEL) - Khối u ác tính
- Melanocytic nevus (NV) - Tế bào hắc tố
- Basal cell carcinoma (BCC) - Ung thư biểu mô tế bào đáy
- Actinic Keratosis (AK) - Dày sừng ánh sáng
- Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis) (BKL) - Dày sừng lành tính
- Dermatofibroma (DF) - U xơ da
- Vascular Lesion (VASC) - Tổn thương mạch máu
- Squamous cell carcinoma (SCC) - Ung thư biểu mô tế bào vảy
- None of the others (UNK) - Không biết

Một điểm đặc biệt của bộ dữ liệu ISIC-2019 là nó cung cấp các hình ảnh chụp bằng nhiều phương pháp khác nhau, bao gồm cả hình ảnh chụp bằng máy quét da, MRI và siêu âm. Điều này giúp các nhà nghiên cứu và chuyên gia y tế có thể phát triển các phương pháp phân tích hình ảnh đa dạng và đáng tin cậy hơn so với việc chỉ sử dụng một phương pháp chụp duy nhất.

<sup>2</sup><https://challenge.isic-archive.com/data/#2019>



Bên cạnh đó, bộ dữ liệu ISIC-2019 còn cung cấp thông tin chi tiết về từng bệnh nhân, bao gồm tuổi, giới tính và vị trí của bệnh lý da. Thông tin này giúp các nhà nghiên cứu và chuyên gia y tế có thể phân tích các yếu tố tác động đến bệnh lý da và phát triển các mô hình dự đoán khác nhau.

Bộ dữ liệu ISIC-2019 cũng được sử dụng rất nhiều trong nhiều nghiên cứu về trí tuệ nhân tạo và học sâu, nhằm phát triển các mô hình học máy để phát hiện và phân loại các bệnh lý da. Google Scholar <sup>3</sup> chứa khoảng 12 nghìn kết quả khi tìm kiếm với từ khoá "ISIC-2019 dataset". Những nghiên cứu có thể kể đến như:

- Kassem cùng các cộng sự [51] đã đề xuất một mô hình mạng nơ-ron tích chập để dự đoán bệnh da liễu sử dụng bộ dữ liệu ISIC-2019.
- Cassidy cùng các cộng sự [52] đã đề xuất một mô hình mạng tích chập chẩn đoán bệnh da với bộ dữ liệu ISIC-2019 để huấn luyện và đánh giá hiệu quả của mô hình.
- Gong cùng các cộng sự [53] cũng đã đề xuất một mô hình mạng nơ-ron tích chập để chẩn đoán bệnh ngoài da sử dụng bộ dữ liệu ISIC-2019 đánh giá độ hiệu quả của mô hình.

Một số hình ảnh về các loại bệnh trong bộ dữ liệu được mô tả như hình B.2.

## B.2 Thống kê mô tả và nhận xét dữ liệu

Trong phần này, tác giả tập trung vào phân tích, mô tả, thống kê dữ liệu cho bộ ISIC-2019 vì HAM10000 đã có trong ISIC-2019.

### B.2.1 ISIC-2019

Như đã trình bày ở phần trước, ISIC-2019 chứa tổng cộng 25,331 ảnh cho việc huấn luyện mô hình với 8 loại bệnh khác nhau. Tập kiểm tra chứa 8,238 ảnh nhưng nhãn của chúng không được công bố. Với 8 loại bệnh, ISIC-2019 có 9 lớp cần phân loại như sau.

- Melanoma (MEL) - Khối u ác tính
- Melanocytic nevus (NV) - Tế bào hắc tố
- Basal cell carcinoma (BCC) - Ung thư biểu mô tế bào đáy
- Actinic Keratosis (AK) - Dày sừng ánh sáng
- Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis) (BKL) - Dày sừng lành tính

---

<sup>3</sup><https://scholar.google.com>





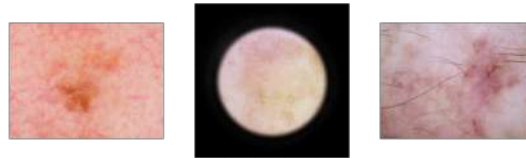
(a) Melanoma (MEL) - Khối u ác tính



(b) Melanocytic nevus (NV) - Tế bào hắc tố



(c) Basal cell carcinoma (BCC) - Ung thư biểu mô tế bào đáy



(d) Actinic Keratosis (AK) - Dày sừng ánh sáng



(e) Benign keratosis (BKL) - Dày sừng lành tính



(f) Dermatofibroma (DF) - U xơ da



(g) Vascular Lesion (VASC) - Tổn thương mạch máu



(h) Squamous cell carcinoma (SCC) - Ung thư biểu mô tế bào vảy

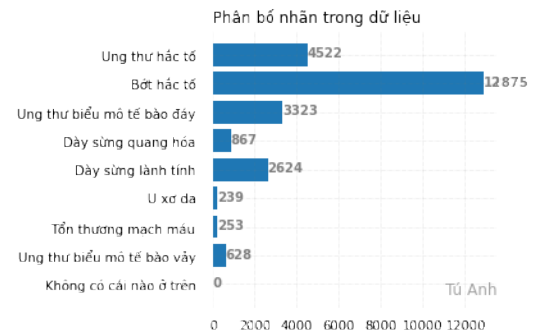
Hình B.2: Một số ví dụ về các bệnh da liễu trong ISIC-2019

- Dermatofibroma (DF) - U xơ da
- Vascular Lesion (VASC) - Tổn thương mạch máu
- Squamous cell carcinoma (SCC) - Ung thư biểu mô tế bào vảy
- None of the others (UNK) - Không biết

### B.2.1.1 Phân bố lớp

STT	Ký hiệu	Số lượng ảnh
0	MEL	4522
1	NV	12875
2	BCC	3323
3	AK	867
4	BKL	2624
5	DF	239
6	VASC	253
7	SCC	628
8	UNK	0

Hình B.3: Bảng phân bố nhãn bộ ISIC-2019



Hình B.4: Biểu đồ phân bố nhãn bộ ISIC-2019

Từ bảng phân bố nhãn (bảng B.3 và hình B.4) của bộ dữ liệu ISIC-2019, chúng ta có thể thấy rằng các loại bệnh da phân bố không đồng đều. Loại bệnh Melanocytic nevus (NV) có số lượng ảnh nhiều nhất (12,875 ảnh), chiếm hơn 50% tổng số ảnh trong bộ dữ liệu. Trong khi đó, loại bệnh Dermatofibroma (DF) và Vascular Lesion (VASC) có số lượng ảnh rất thấp, chỉ chiếm khoảng 1% tổng số ảnh.

Các loại bệnh ung thư da như Melanoma (MEL), Basal cell carcinoma (BCC), và Squamous cell carcinoma (SCC) cũng có số lượng ảnh khá cao, tuy nhiên vẫn thấp hơn số lượng ảnh của loại bệnh NV. Điều này thể hiện rằng các loại bệnh lành tính (như NV) thường phổ biến hơn các loại bệnh ung thư da, và có thể là do các loại bệnh này được chẩn đoán và phát hiện sớm hơn.

Ngoài ra, bộ dữ liệu ISIC-2019 còn có một lớp None of the others (UNK) với số lượng ảnh bằng 0, nghĩa là không có ảnh được gán nhãn vào lớp này. Lớp này thường được sử dụng để xử lý các ảnh không thuộc vào bất kỳ loại bệnh nào trong bộ dữ liệu.

Tập kiểm tra của bộ dữ liệu này có 8,238 ảnh, tuy nhiên nhãn của chúng không được công bố. Điều này nhằm mục đích đánh giá tính độc lập của các mô hình phân loại được huấn luyện trên tập dữ liệu huấn luyện.

Tóm lại, ta có thể có vài nhận xét như sau:

- Dữ liệu có sự mất cân bằng rất lớn.

- Bớt hắc tố (NV) là bệnh phổ biến nhất vì sự xuất hiện rất nhiều của nó.
- Bệnh ít gặp nhất trong bộ dữ liệu là U xơ da (DF) và Tổn thương mạch máu (VASC).
- Nhân UNK không cần thiết trong bộ dữ liệu.

### ***B.2.1.2 Phân bố vị trí của các bệnh***

Bộ dữ liệu ISIC-2019 còn cung cấp thông tin về các vùng da bị bệnh, bao gồm:

- Anterior torso (Thân trước)
- Head/neck (Đầu/cổ)
- Lateral torso (Thân bên)
- Lower extremity (Chi dưới)
- Oral/genital (Miệng/bộ phận sinh dục)
- Palms/soles (Lòng bàn tay/lòng bàn chân)
- Posterior torso (Thân sau)
- Unknow (Không xác định)
- Upper extremity (Chi trên)

Phân bố bệnh theo vị trí được thể hiện như hình B.5. Qua đó, có thể thấy:

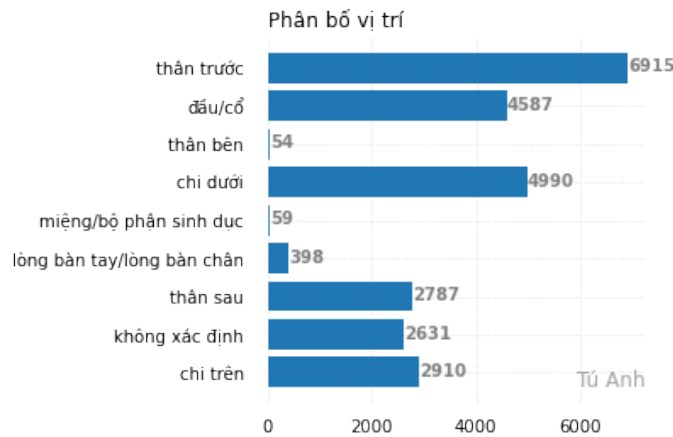
- Các bệnh về da thường tập trung ở thân trước, đầu/cổ và chi dưới.
- Có một lượng lớn dữ liệu không xác định được vị trí bệnh (2.6 nghìn ảnh).
- Không có nhiều dữ liệu bệnh ở vùng thân bên, miệng và bộ phận sinh dục.

Đối với mỗi bệnh thì có những đặc thù về vị trí khác nhau. Bộ dữ liệu ISIC-2019 cũng đã thể hiện được một phần nào luận điểm đó qua biểu đồ hình B.6. Ta có thể thấy:

- Dày sừng quan hoá thường tập trung ở đầu và cổ.
- U xơ da thường ở các chi (trên và dưới).
- Đa số bệnh còn lại thường ở thân trước, đầu cổ hay các chi.

Bộ dữ liệu ISIC-2019 cung cấp thông tin về vị trí của các vùng da bị bệnh. Dựa vào sự phân bố bệnh theo vị trí, có thể rút ra một số nhận xét quan trọng như sau:

- Các bệnh da liễu thường tập trung ở thân trước, đầu/cổ và chi dưới.



Hình B.5: Biểu đồ phân bố ảnh bệnh theo vị trí của bộ ISIC-2019

- Một lượng lớn dữ liệu (khoảng 2.6 nghìn ảnh) không xác định được vị trí bệnh.
- Dữ liệu về các vùng như thân bên, miệng và bộ phận sinh dục rất hạn chế.

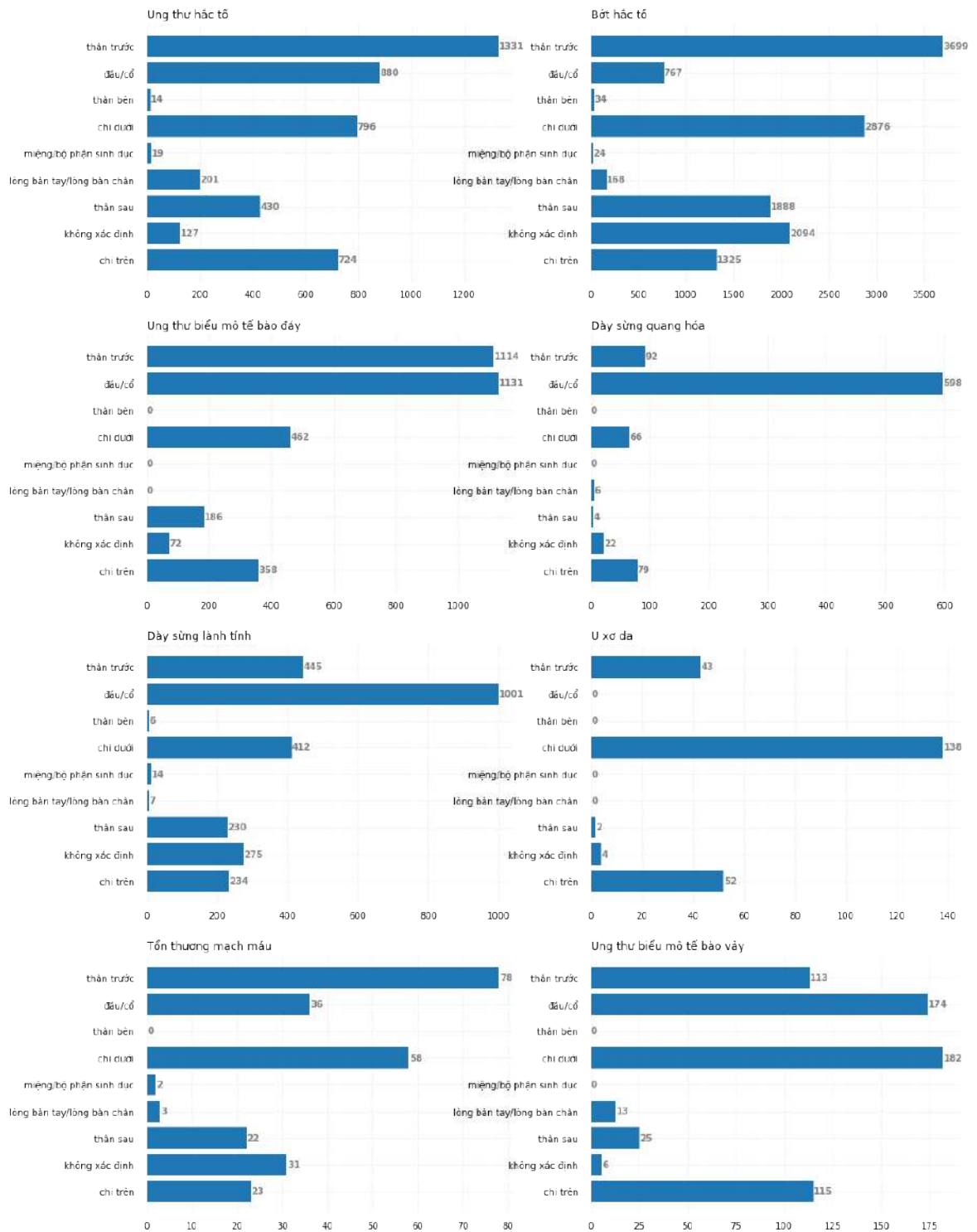
Bộ dữ liệu này cũng cung cấp thông tin chi tiết về tần suất của các loại ung thư da theo từng vị trí trên cơ thể. Cụ thể:

- **Đầu/cổ:** Bệnh ung thư biểu mô tế bào đáy (BCC) chiếm tỷ lệ cao nhất với 24.66%, tiếp theo là dày sừng lành tính (21.82%) và ung thư hắc tố (MEL) chiếm 19.18%. U xơ da (DF) có tỷ lệ thấp nhất với 0%, và tổn thương mạch máu (VASC) chiếm 0.78%.
- **Thân trước:** Bớt hắc tố (NV) chiếm ưu thế rõ rệt với 53.49%, cao gần gấp ba lần so với ung thư hắc tố (19.25%). Tổng cộng, hai bệnh liên quan đến mất sắc tố da (Bớt hắc tố và ung thư hắc tố) chiếm 72.74% tổng số ảnh.
- **Thân bên:** Tương tự thân trước, Bớt hắc tố chiếm phần lớn số ảnh với 62.96%, trong khi ung thư hắc tố chiếm 25.93%. Các bệnh khác chỉ chiếm một tỷ lệ rất nhỏ.
- **Chi dưới:** Bớt hắc tố vẫn chiếm tỷ lệ cao nhất với 57.64%, tiếp theo là ung thư hắc tố với 15.95% và ung thư biểu mô tế bào đáy với 9.26%. Các bệnh khác như dày sừng quang hóa, dày sừng lành tính, u xơ da và tổn thương mạch máu chiếm tỷ lệ thấp hơn nhiều.
- **Miệng và bộ phận sinh dục:** Bớt hắc tố chiếm 40.68% số ảnh, ung thư hắc tố chiếm 32.2%, trong khi các bệnh khác chiếm tỷ lệ nhỏ hơn. Tổn thương mạch máu chỉ chiếm 3.39%.
- **Lòng bàn tay/lòng bàn chân:** Ung thư hắc tố chiếm ưu thế với 50.5%, theo sau là bớt hắc tố với 42.21%. Các bệnh khác chiếm tỷ lệ dưới 5%.

- **Thân sau:** Bớt hắc tố là bệnh phổ biến nhất với 67.74%, theo sau là ung thư hắc tố với 15.43%. Các bệnh còn lại chiếm tỷ lệ rất thấp, dưới 1%.
- **Chi trên:** Bớt hắc tố chiếm 45.53%, trong khi ung thư hắc tố chiếm 24.88%. Dày sừng lành tính xếp thứ ba với 8.04%.

Nhìn chung, sự phân bố bệnh trên cơ thể có một số xu hướng rõ rệt. Ví dụ, ung thư hắc tố thường xuất hiện nhiều nhất ở đầu/cổ và chi trên, trong khi bớt hắc tố thường phổ biến ở thân trước và chi dưới. Ung thư biểu mô tế bào đáy thường tập trung ở đầu/cổ và thân trước, do bệnh này hay xuất hiện ở những khu vực tiếp xúc nhiều với ánh nắng mặt trời, như mặt và môi. Bệnh dày sừng quang hóa cũng tập trung chủ yếu ở đầu/cổ.

Phân tích này giúp làm rõ mối liên hệ giữa vị trí bệnh và loại bệnh, qua đó hỗ trợ các bác sĩ và hệ thống chẩn đoán tự động trong việc phát hiện sớm và chính xác các bệnh ung thư da.



Hình B.6: Biểu đồ phân bố ảnh bệnh theo vị trí của bộ ISIC-2019

# PHỤ LỤC C | *WN Labeller* - Công cụ quản lý dữ liệu hình ảnh và hỗ trợ chẩn đoán bệnh da liễu

Trong lĩnh vực chẩn đoán bệnh lý da liễu, việc xử lý và quản lý dữ liệu hình ảnh đóng vai trò quan trọng, đặc biệt khi số lượng ảnh da liễu cần phân tích ngày càng tăng. Để đáp ứng nhu cầu này, công cụ *WN Labeller* đã được phát triển, tích hợp công nghệ trí tuệ nhân tạo (AI) nhằm hỗ trợ các bác sĩ và cơ sở y tế trong việc quản lý, gán nhãn và tự động nhận diện vùng da tổn thương. Đây là một hệ thống toàn diện, giúp tối ưu hóa quá trình chẩn đoán và nâng cao hiệu quả điều trị bệnh lý da liễu.

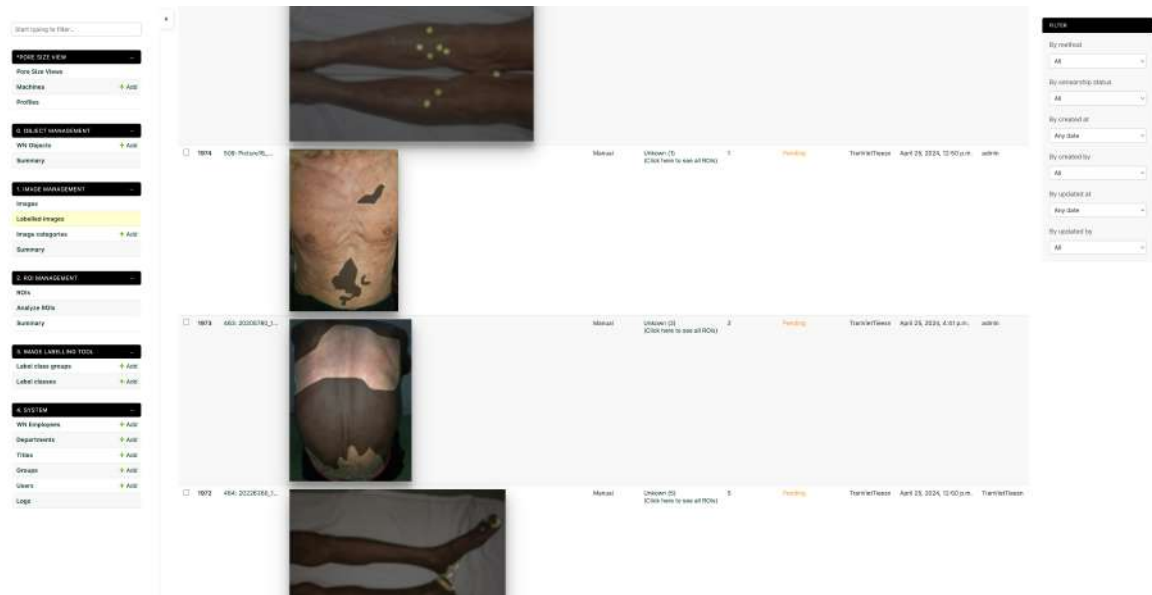
Công cụ *WN Labeller* không chỉ giúp cải thiện hiệu suất làm việc của bác sĩ mà còn tăng cường độ chính xác trong chẩn đoán thông qua việc tự động nhận diện và phân tích các tổn thương trên da. Hệ thống này có tính linh hoạt cao, dễ dàng triển khai trong nhiều môi trường y tế khác nhau, từ phòng khám nhỏ đến các bệnh viện lớn.

## C.1 Quản lý dữ liệu hình ảnh

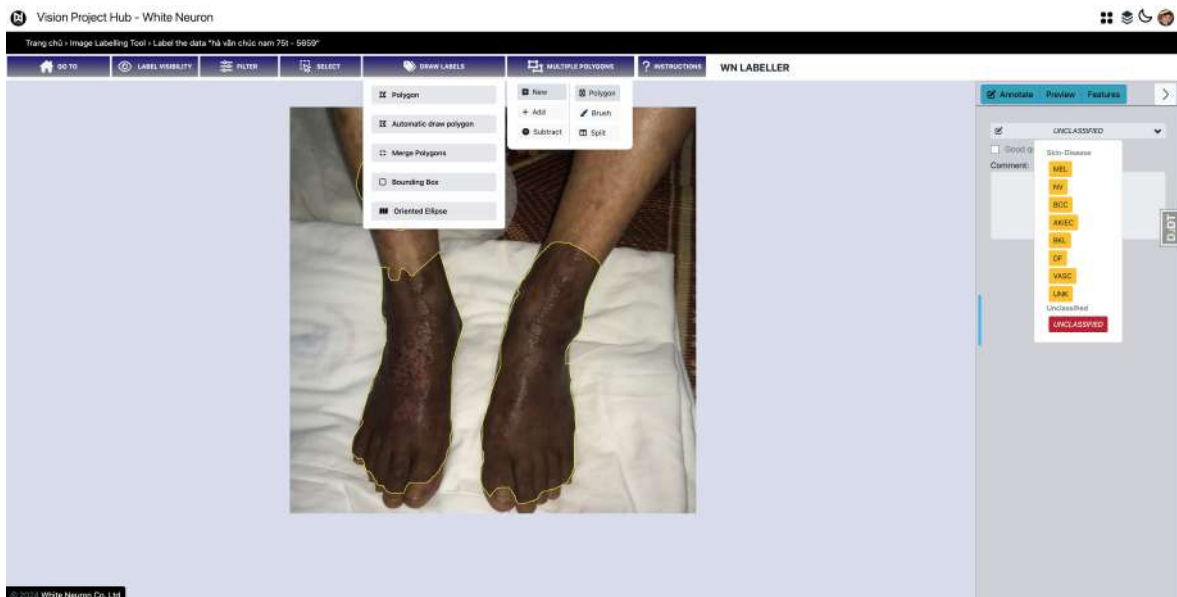
Tính năng quan trọng đầu tiên của *WN Labeller* là khả năng quản lý hiệu quả dữ liệu hình ảnh da liễu. Các hình ảnh được lưu trữ và sắp xếp theo hệ thống, giúp người dùng dễ dàng truy cập và quản lý một khối lượng lớn dữ liệu. Các hình ảnh da liễu được tổ chức theo các danh mục rõ ràng, bao gồm thông tin về loại bệnh, thời gian chụp, và các thông tin liên quan khác (hình C.1). Điều này không chỉ giúp theo dõi tình trạng của bệnh nhân mà còn hỗ trợ quá trình nghiên cứu và chẩn đoán chính xác.

## C.2 Gán nhãn hình ảnh

Một trong những thách thức lớn nhất trong việc xử lý dữ liệu y tế là gán nhãn chính xác cho các hình ảnh. *WN Labeller* cung cấp một công cụ gán nhãn thông minh, cho phép các chuyên viên y tế dễ dàng thêm các nhãn và chú thích liên quan đến tình trạng tổn thương da (hình C.2). Các nhãn này được sử dụng để tổ chức dữ liệu có hệ thống, giúp việc tìm kiếm và phân tích sau này dễ dàng hơn. Hơn nữa, công cụ hỗ trợ cả việc gán nhãn thủ công lẫn tự động, giảm thiểu thời gian cần thiết để xử lý các khối lượng dữ liệu lớn và đảm bảo tính nhất quán của các nhãn.



Hình C.1: Màn quản lý hình ảnh



Hình C.2: Màn gắn nhãn hình ảnh

### C.3 Tự động nhận diện vùng tổn thương da liễu bằng AI

Tính năng nổi bật nhất của *WN Labeller* là khả năng tự động nhận diện các vùng tổn thương da dựa trên công nghệ trí tuệ nhân tạo (AI). Bằng cách sử dụng các thuật toán học sâu và mạng nơ-ron tích chập (CNN), hệ thống có thể phân tích hình ảnh đầu vào và xác định chính xác các vùng da bị tổn thương (hình ??). Điều này giúp giảm tải công việc cho các bác sĩ và tăng tốc quá trình chẩn đoán. AI sẽ cô lập và làm nổi bật các vùng quan trọng trên hình ảnh, giúp bác sĩ dễ dàng theo dõi và đánh giá tình trạng bệnh lý.



---

Image: 2143: Picture43\_...

---

Method: Skin Lesion Detection

---

Labels: lesion (17)  
(Click here to see all ROIs)

---

Preview:



(click to zoom)

Heatmap:



(click to zoom)

Hình C.3: Chức năng tự động nhận diện vùng tổn thương da

## C.4 Hiển thị và phân tích hình ảnh da liễu

*WN Labeller* cung cấp một bộ công cụ mạnh mẽ để hiển thị và phân tích chi tiết các hình ảnh da liễu:

- **Hiển thị hình ảnh sắc nét:** Hệ thống hỗ trợ hiển thị hình ảnh với độ phân giải cao, giúp các bác sĩ dễ dàng quan sát và phân tích chi tiết từng vùng tổn thương trên da.
- **Biểu đồ và sự kiện:** Bên cạnh việc hiển thị hình ảnh, hệ thống còn cho phép vẽ biểu đồ và ghi nhận các sự kiện quan trọng trong quá trình phân tích hình ảnh. Điều này giúp theo dõi sự tiến triển của các vùng tổn thương theo thời gian.
- **Phóng to/thu nhỏ:** Hệ thống cung cấp khả năng phóng to hoặc thu nhỏ hình ảnh, giúp các bác sĩ dễ dàng xem xét chi tiết hơn ở các vùng tổn thương nhỏ hoặc khó nhận biết.
- **Làm nổi bật vùng tổn thương:** Hệ thống hỗ trợ làm nổi bật các vùng da bị tổn thương, giúp bác sĩ dễ dàng xác định và phân tích các khu vực quan trọng cần quan tâm.

## C.5 Quản trị hệ thống

*WN Labeller* cung cấp các công cụ quản trị toàn diện nhằm hỗ trợ các cơ sở y tế trong việc quản lý dữ liệu bệnh nhân và hình ảnh liên quan:

- **Quản lý bệnh án:** Hệ thống cho phép quản lý bệnh án và các thông tin liên quan đến quá trình chẩn đoán và điều trị của từng bệnh nhân, từ đó giúp theo dõi tiến trình bệnh lý (hình C.4).
- **Quản lý hình ảnh và sự kiện:** Các hình ảnh và sự kiện liên quan đến quá trình chẩn đoán và điều trị được lưu trữ và sắp xếp một cách hệ thống, giúp bác sĩ dễ dàng truy xuất và phân tích khi cần.

## C.6 Kết luận

*WN Labeller* là một công cụ không thể thiếu trong lĩnh vực da liễu, đặc biệt đối với các cơ sở y tế cần quản lý và xử lý một khối lượng lớn hình ảnh da liễu. Với khả năng quản lý dữ liệu, gán nhãn thông minh, và nhận diện tự động vùng tổn thương, hệ thống này đã mang lại sự hỗ trợ hiệu quả cho các bác sĩ trong quá trình chẩn đoán và điều trị bệnh. *WN Labeller* không chỉ cải thiện hiệu suất làm việc mà còn giúp tăng cường chất lượng dịch vụ y tế, từ đó nâng cao hiệu quả điều trị cho bệnh nhân.

Vision Project Hub - White Neuron

Trang chủ - Project management - Projects - PRJ00 - Skin Disease Detection

Change project PRJ00 - Skin Disease Detection [Dashboard](#) [Export](#) [View all files](#)

Project	Project configurations	Project team	Data groups	Label groups								
ID	GROUP NAME	DESCRIPTION	STATUS	# DATA	# NEW DATA	NGÀY TẠO	NGÀY CHẤM DỨT	Áp dụng	Đã xóa	Tạo bởi	Link API ID	Liên hệ người khác
125	hà văn châu nam 75x	skin photo-hà văn châu nam 75x	<a href="#">Data</a> (2024-09-17 08:03:54)	5	0	09:42 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
126	nguyễn tuấn minh nam 40x - s40x	skin photo-nguyễn tuấn minh nam 40x - s40x	<a href="#">Data</a> (2024-09-17 08:03:54)	19	0	09:42 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
127	huỳnh quốc trung nam 42x	skin photo-huỳnh quốc trung nam 42x	<a href="#">Data</a> (2024-09-17 08:03:54)	4	0	09:47 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
128	bùi túy bích nữ 40x s40x - s40x	skin photo-bùi túy bích nữ 40x s40x - s40x	<a href="#">Data</a> (2024-09-17 08:03:54)	7	0	09:49 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
129	nguyễn thàn anh nữ 22x	skin photo-nguyễn thàn anh nữ 22x	<a href="#">Data</a> (2024-09-17 08:03:54)	3	0	09:50 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
130	hà ngọc lệ thị trung nữ 70x	skin photo-hà ngọc lệ thị trung nữ 70x	<a href="#">Data</a> (2024-09-17 08:03:54)	7	2	09:50 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
131	phạm minh hiếu nam 40x	skin photo-phạm minh hiếu nam 40x	<a href="#">Data</a> (2024-09-17 08:03:54)	1	0	09:50 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
132	trần trung thành nam 34x	skin photo-trần trung thành nam 34x	<a href="#">Data</a> (2024-09-17 08:03:54)	6	0	09:50 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
133	trần thị chuẩn nữ 82x	skin photo-trần thị chuẩn nữ 82x	<a href="#">Data</a> (2024-09-17 08:03:54)	6	0	09:50 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>
134	trần ngọc khánh nữ 30x	skin photo-trần ngọc khánh nữ 30x	<a href="#">Data</a> (2024-09-17 08:03:54)	2	0	09:50 Ngày 23 tháng 2 năm 2024	10-03 Ngày 17 tháng 9 năm 2024	<input type="checkbox"/>	<input type="checkbox"/>	admin	TransNetTeam	<input type="checkbox"/>

White Neuron Co., Ltd. [Lưu và thêm mới](#) [Lưu và tiếp tục chỉnh sửa](#) [Lưu lại](#)

Hình C.4: Quản lý bệnh án