

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO
THỰC TẬP CƠ SỞ

Giáo viên hướng dẫn: Nguyễn Tất Thắng

Sinh viên thực hiện : Nguyễn Thế Khiêm

Mã sinh viên : B22DCDT171

Lớp : E22CQCN03-B

Hà Nội, 2025

LỜI CẢM ƠN

Trước tiên, em xin gửi lời cảm ơn sâu sắc và chân thành nhất tới thầy TS. Nguyễn Tất Thắng, người đã tận tình hướng dẫn, định hướng và chia sẻ những kiến thức quý báu, giúp em hoàn thành tốt báo cáo này. Sự hỗ trợ và chỉ dẫn tận tâm của thầy là nguồn động viên to lớn trong suốt quá trình nghiên cứu và viết báo cáo.

Em xin chân thành cảm ơn Ban lãnh đạo Học viện Công nghệ Bưu chính Viễn thông, cùng toàn thể các thầy, cô đã luôn tạo điều kiện tốt nhất và nhiệt tình hỗ trợ em trong suốt quá trình học tập và nghiên cứu tại trường. Sự tận tụy và kiến thức chuyên môn của quý thầy, cô đã giúp em có được nền tảng vững chắc để hoàn thành báo cáo này.

Bên cạnh đó, em cũng xin bày tỏ lòng biết ơn tới các bạn học viên đã luôn đồng hành, hỗ trợ và chia sẻ những kinh nghiệm quý giá trong suốt chặng đường học tập.

Không thể không nhắc đến gia đình, bạn bè và đồng nghiệp - những người đã luôn bên cạnh động viên, khích lệ và tạo mọi điều kiện thuận lợi nhất để em tập trung hoàn thành tốt công việc nghiên cứu. Sự ủng hộ và tình cảm quý báu từ họ là nguồn động lực lớn lao để em vượt qua những khó khăn và thử thách.

Cuối cùng, dù đã nỗ lực hết mình trong quá trình thực hiện báo cáo, em nhận thức rằng không thể tránh khỏi những thiếu sót. Em rất mong nhận được sự thông cảm và những góp ý quý báu từ các thầy, cô và bạn đọc để hoàn thiện hơn nữa nội dung nghiên cứu của mình. Xin chân thành cảm ơn!

MỤC LỤC

MỞ ĐẦU	5
CHƯƠNG 1: TỔNG QUAN VỀ UNG THƯ MELANOMA VÀ ỨNG DỤNG THỊ GIÁC MÁY TÍNH TRONG Y TẾ.....	7
1.1. Giới thiệu về ung thư da và ung thư melanoma.....	7
1.2. Tầm quan trọng của việc chẩn đoán sớm ung thư melanoma.....	9
1.3. Ứng dụng của thị giác máy tính và học sâu trong chẩn đoán hình ảnh y tế.....	11
1.4. Kết luận chương 1	12
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT VỀ HỌC SÂU TRONG PHÂN LOẠI HÌNH ẢNH Y TẾ	16
2.1. Mạng Nơ-ron Tích chập (Convolutional Neural Networks - CNN).....	16
2.2. Học Chuyển giao (Transfer Learning).....	20
2.3. Kiến trúc Mạng Residual (ResNet).....	23
2.4. Kết luận chương 2	25
CHƯƠNG 3: PHƯƠNG PHÁP LUẬN VÀ THIẾT KẾ HỆ THỐNG PHÂN LOẠI UNG THƯ MELANOMA	27
3.1. Bộ dữ liệu HAM10000.....	27
3.2. Tiền xử lý dữ liệu.....	29
3.3. Xây dựng mô hình phân loại.....	32
3.4. Thiết lập môi trường và công cụ thực nghiệm.....	34
3.5. Kết luận chương 3	35
CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ.....	37
4.1. Quy trình thực nghiệm.....	37
Giai đoạn 1: Huấn luyện phần đầu phân loại (Feature Extraction).....	37
Giai đoạn 2: Fine-tuning mô hình đầy đủ.....	38
4.2. Các chỉ số đánh giá hiệu năng mô hình.....	39
4.3. Kết quả thực nghiệm và phân tích.....	41
4.4. So sánh và thảo luận.....	42
4.5. Kết luận chương 4	43
CHƯƠNG 5: CÀI ĐẶT GIẢI PHÁP VÀ TRIỂN KHAI ỨNG DỤNG.....	46
5.1. Giới thiệu.....	46
5.2. Lưu trữ Mô hình Keras đã Huấn luyện.....	46

5.3. Xây dựng Ứng dụng Desktop với PyQt5	46
5.4. Thiết kế Giao diện Người dùng (UI) và Luồng hoạt động.....	47
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	50
Tóm tắt các kết quả chính đạt được.....	50
Đóng góp của nghiên cứu.....	50
Đề xuất các hướng phát triển trong tương lai.....	52
Nguồn trích dẫn	53

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Thuật ngữ tiếng Anh	Thuật ngữ tiếng Việt
AI	Artificial Intelligence	Trí tuệ nhân tạo
ML	Machine Learning	Học máy
DL	Deep Learning	Học sâu
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
TL	Transfer Learning	Học chuyển giao
CAM	Class Activation Map	Bản đồ kích hoạt lớp
AUC	Area Under the Curve	Diện tích dưới đường cong
ROC	Receiver Operating Characteristic	Đặc trưng vận hành thu nhận tín hiệu
TP	True Positive	Dương tính thật
FP	False Positive	Dương tính giả
FN	False Negative	Âm tính giả
TN	True Negative	Âm tính thật
SGD	Stochastic Gradient Descent	Thuật toán hạ dốc ngẫu nhiên
HAM10000	Human Against Machine with 10000 Images	Bộ dữ liệu hình ảnh tổn thương da HAM10000
ResNet	Residual Network	Mạng nơ-ron dư

DANH MỤC BẢNG BIỂU

Bảng 1.1	So sánh các loại ung thư da phổ biến	13
Bảng 1.2	Các giai đoạn phát triển của Ung thư Melanoma và Đặc điểm chính	13
Bảng 2.1	Các chiến lược Transfer Learning và trường hợp áp dụng	25
Bảng 3.1	Phân bố chi tiết các lớp tổn thương trong bộ dữ liệu HAM10000	35
Bảng 4.1	Các tham số huấn luyện chi tiết cho từng mô hình/chiến lược	43
Bảng 4.2	Kết quả đánh giá hiệu năng chi tiết trên tập kiểm thử	44

DANH MỤC HÌNH ẢNH

Hình 3.1	Phân bố số lượng nhãn	35
Hình 4.1	Sự thay đổi của accuracy theo từng epoch trong cả hai giai đoạn huấn luyện.	38
Hình 4.2	Sự thay đổi của loss theo từng epoch trong cả hai giai đoạn huấn luyện.	38
Hình 4.3	Ma trận nhầm lẫn	40
Hình 5.1	App khi chưa có kết quả	48
Hình 5.2	App khi đã có kết quả	48

MỞ ĐẦU

Ung thư melanoma, một trong những dạng ung thư da ác tính nhất, đặt ra một thách thức y tế toàn cầu do khả năng di căn nhanh và tỷ lệ tử vong cao nếu không được can thiệp kịp thời. Việc chẩn đoán sớm và chính xác đóng vai trò then chốt, quyết định đến hiệu quả điều trị và tiên lượng sống còn cho bệnh nhân. Tuy nhiên, các phương pháp chẩn đoán truyền thống, dù có giá trị, vẫn tồn tại những hạn chế nhất định như tính chủ quan, đòi hỏi thời gian và có thể mang tính xâm lấn. Trong bối cảnh đó, sự phát triển mạnh mẽ của Trí tuệ Nhân tạo (AI), đặc biệt là các kỹ thuật học sâu (Deep Learning) với Mạng Nơ-ron Tích chập (CNN), đã mở ra những triển vọng mới trong việc phân tích hình ảnh y tế. Các thuật toán này có khả năng tự động học hỏi các đặc điểm phức tạp từ hình ảnh tổn thương da, từ đó hỗ trợ chẩn đoán melanoma một cách nhanh chóng, khách quan và tiềm năng chính xác hơn. Sự phổ biến của các nghiên cứu về ung thư da và melanoma, kết hợp với những tiến bộ vượt bậc trong lĩnh vực AI, cho thấy một xu hướng rõ ràng hướng tới các giải pháp chẩn đoán tự động và hệ thống hỗ trợ quyết định lâm sàng thông minh. Các công cụ AI không chỉ có tiềm năng giảm tải gánh nặng cho các chuyên gia da liễu mà còn cải thiện khả năng tiếp cận các dịch vụ chẩn đoán chất lượng cao, đặc biệt ở những khu vực hạn chế về nguồn lực y tế chuyên sâu.

Trong những năm gần đây, cộng đồng nghiên cứu đã chứng kiến nhiều nỗ lực ứng dụng học máy và học sâu vào việc phân loại tổn thương da và chẩn đoán sớm melanoma. Việc công bố và chia sẻ các bộ dữ liệu hình ảnh da liễu quy mô lớn, như HAM10000, đã tạo điều kiện thuận lợi, cung cấp nền tảng dữ liệu quý giá cho việc huấn luyện và đánh giá các mô hình AI. Song song đó, các kỹ thuật học chuyển giao (transfer learning) kết hợp với những kiến trúc CNN tiên tiến như ResNet đã liên tục chứng minh hiệu quả vượt trội trong nhiều bài toán thị giác máy tính phức tạp, bao gồm cả lĩnh vực phân tích hình ảnh y tế. Mặc dù đã có những thành tựu đáng kể, vẫn còn đó những khoảng trống nghiên cứu và các khía cạnh cần được tiếp tục cải thiện, ví dụ như tối ưu hóa mô hình để đạt hiệu suất cao hơn trên các bộ dữ liệu cụ thể, giải quyết triệt để vấn đề mất cân bằng lớp thường gặp trong dữ liệu y tế, hoặc nâng cao khả năng phát hiện melanoma ở những giai đoạn rất sớm khi các dấu hiệu còn mờ nhạt. Việc đề tài này có định hướng cung cấp tệp mã nguồn liên quan sau khi hoàn thành báo cáo cho thấy một cam kết về tính thực tiễn và khả năng kiểm chứng của nghiên cứu, vượt ra ngoài khuôn khổ lý thuyết đơn thuần và hướng đến việc tạo ra một sản phẩm khoa học có thể triển khai và đánh giá được.

Mục tiêu chính của đề tài này là xây dựng và đánh giá một hệ thống dựa trên Mạng Nơ-ron Tích chập (CNN) sử dụng kỹ thuật học chuyển giao với kiến trúc ResNet để thực hiện nhiệm vụ phân loại ung thư melanoma từ hình ảnh tổn thương da trên bộ dữ liệu HAM10000. Để đạt được mục tiêu này, các mục tiêu cụ thể sau đây sẽ được thực hiện: (1) Nghiên cứu và trình bày một cách có hệ thống cơ sở lý thuyết về ung thư melanoma, các khái niệm nền tảng của CNN, nguyên lý của học chuyển giao và chi tiết về kiến trúc ResNet. (2) Tiến hành phân tích sâu và thực hiện các bước tiền xử lý cần thiết cho bộ dữ liệu HAM10000, bao gồm việc áp dụng các kỹ thuật tăng cường dữ liệu nhằm giải quyết vấn đề mất cân bằng giữa các lớp tổn thương. (3) Triển khai và huấn luyện mô hình ResNet (cụ thể là ResNet50) bằng cách áp dụng chiến lược học chuyển giao, đặc biệt là phương

pháp tinh chỉnh (fine-tuning). (4) Đánh giá một cách toàn diện hiệu năng của mô hình đã huấn luyện thông qua các chỉ số đánh giá định lượng phù hợp như độ chính xác (accuracy), độ chính xác dự đoán (precision), độ thu hồi (recall), chỉ số F1 (F1-score) và diện tích dưới đường cong ROC (AUC) trên tập dữ liệu kiểm thử độc lập. (5) Thực hiện so sánh kết quả đạt được với các nghiên cứu liên quan đã được công bố, nếu có, để định vị đóng góp của đề tài.

Đối tượng nghiên cứu chính của đề tài bao gồm các thuật toán học sâu, cụ thể là Mạng Nơ-ron Tích chập và kiến trúc ResNet, các kỹ thuật học chuyển giao, và bộ dữ liệu hình ảnh da liễu HAM10000. Phạm vi nghiên cứu sẽ tập trung vào bài toán phân loại đa lớp, bao gồm 7 loại tổn thương da được định nghĩa trong bộ dữ liệu HAM10000, trong đó có ung thư melanoma. Quá trình nghiên cứu sẽ bao gồm các giai đoạn từ thu thập và phân tích tài liệu khoa học, tiền xử lý và chuẩn bị dữ liệu, xây dựng và tối ưu hóa mô hình, tiến hành thực nghiệm, đến phân tích và đánh giá kết quả.

Phương pháp nghiên cứu được áp dụng trong đề tài này là sự kết hợp giữa nghiên cứu lý thuyết và nghiên cứu thực nghiệm. Nghiên cứu lý thuyết bao gồm việc tổng hợp, phân tích và đánh giá các tài liệu khoa học, các bài báo chuyên ngành, sách chuyên khảo liên quan đến ung thư melanoma, các kỹ thuật học sâu ứng dụng trong phân tích hình ảnh y tế, và các công trình nghiên cứu đã có về chủ đề này. Nghiên cứu thực nghiệm sẽ bao gồm các bước: (1) Thu thập và chuẩn bị dữ liệu từ bộ HAM10000. (2) Thiết kế và triển khai các mô hình CNN dựa trên kiến trúc ResNet với kỹ thuật học chuyển giao. (3) Huấn luyện và kiểm thử mô hình trên các tập dữ liệu đã được phân chia một cách cẩn thận (training, validation, test). (4) Sử dụng các chỉ số đánh giá định lượng đã nêu để đo lường và so sánh hiệu năng của các mô hình.

Cấu trúc của báo cáo này được tổ chức như sau: Mở đầu giới thiệu tổng quan về đề tài. Chương 1 trình bày tổng quan về ung thư melanoma và vai trò của thị giác máy tính trong y tế. Chương 2 đi sâu vào cơ sở lý thuyết của các kỹ thuật học sâu được sử dụng. Chương 3 mô tả chi tiết phương pháp luận và thiết kế hệ thống phân loại. Chương 4 trình bày quá trình thực nghiệm và phân tích, đánh giá kết quả đạt được. Cuối cùng, phần Kết luận sẽ tóm tắt những đóng góp chính, chỉ ra các hạn chế của nghiên cứu và đề xuất những hướng phát triển tiềm năng trong tương lai.

CHƯƠNG 1: TỔNG QUAN VỀ UNG THƯ MELANOMA VÀ ỨNG DỤNG THỊ GIÁC MÁY TÍNH TRONG Y TẾ

1.1. Giới thiệu về ung thư da và ung thư melanoma

Ung thư da là một trong những loại ung thư phổ biến nhất trên toàn cầu, với nhiều dạng khác nhau tùy thuộc vào loại tế bào da bị ảnh hưởng. Việc hiểu rõ về các loại ung thư da, đặc biệt là melanoma, là bước đầu tiên để nhận thức được mức độ nguy hiểm và sự cần thiết của các biện pháp phòng ngừa và phát hiện sớm.

1.1.1. Ung thư da: Các loại phổ biến

Có ba loại ung thư da chính thường gặp: ung thư tế bào đáy (Basal Cell Carcinoma - BCC), ung thư tế bào vảy (Squamous Cell Carcinoma - SCC), và ung thư tế bào hắc tố (Melanoma). BCC và SCC là hai loại phổ biến nhất và thường được gọi chung là ung thư da không phải melanoma. Chúng thường phát triển chậm hơn và ít có khả năng di căn xa so với melanoma.² Mặc dù vậy, nếu không được điều trị, chúng vẫn có thể gây tổn thương cục bộ đáng kể. Sự khác biệt về mức độ ác tính và khả năng di căn làm nổi bật tính nguy hiểm đặc biệt của melanoma, là lý do chính khiến loại ung thư

này trở thành đối tượng quan tâm hàng đầu trong nhiều nghiên cứu y học và công nghệ.

1.1.2. Ung thư melanoma: Định nghĩa, đặc điểm sinh học, các giai đoạn phát triển

Ung thư melanoma, hay còn gọi là u hắc tố, là một dạng ung thư da phát sinh từ các tế bào hắc tố (melanocytes) – những tế bào chịu trách nhiệm sản xuất melanin, sắc tố quyết định màu da, tóc và mắt.² Các tế bào hắc tố có vai trò bảo vệ da khỏi tác hại của tia cực tím (UV) bằng cách sản xuất nhiều melanin hơn khi tiếp xúc với ánh nắng mặt trời. Tuy nhiên, khi các tế bào này phát triển bất thường và không kiểm soát, chúng có thể hình thành khối u ác tính.

Đặc điểm nổi bật và đáng lo ngại nhất của melanoma là khả năng xâm lấn và di căn rất nhanh đến các cơ quan khác trong cơ thể, chẳng hạn như hạch bạch huyết, phổi, gan, não, nếu không được phát hiện và điều trị ở giai đoạn sớm.² Điều này khiến melanoma trở thành loại ung thư da nguy hiểm nhất, mặc dù nó ít phổ biến hơn BCC và SCC.

Nguyên nhân chính gây ra melanoma liên quan mật thiết đến việc tiếp xúc quá mức với tia cực tím (UV) từ ánh nắng mặt trời hoặc các nguồn nhân tạo như giường tắm nắng.¹⁸ Các yếu tố nguy cơ khác bao gồm:

- **Loại da:** Người có làn da trắng, tóc vàng hoặc đỏ, mắt xanh, dễ bị tàn nhang hoặc

cháy nắng có nguy cơ cao hơn do có ít melanin bảo vệ da.

- **Tiền sử cháy nắng:** Đặc biệt là các đợt cháy nắng nghiêm trọng, phỏng rộp da trong quá khứ, nhất là ở thời thơ ấu hoặc thanh thiếu niên.
- **Số lượng nốt ruồi:** Người có nhiều nốt ruồi (trên 50) hoặc có các nốt ruồi không điển hình (kích thước lớn, hình dạng bất thường, màu sắc không đều) có nguy cơ mắc melanoma cao hơn.
- **Tiền sử gia đình:** Nếu có người thân trực hệ (cha mẹ, anh chị em) mắc melanoma, nguy cơ mắc bệnh sẽ tăng lên.
- **Hệ miễn dịch suy yếu:** Những người có hệ miễn dịch bị suy giảm do bệnh tật (ví dụ: HIV/AIDS) hoặc do sử dụng thuốc ức chế miễn dịch (ví dụ: sau ghép tạng) cũng có nguy cơ cao hơn.
- **Vị trí địa lý:** Sống ở những vùng gần xích đạo hoặc ở độ cao lớn, nơi cường độ tia UV mạnh hơn.

Melanoma có thể xuất hiện ở bất kỳ vị trí nào trên cơ thể. Ở các nước phương Tây, bệnh thường phát sinh ở những vùng da thường xuyên tiếp xúc với ánh nắng mặt trời như ngực, lưng (ở nam giới), chân (ở nữ giới), cánh tay và mặt.² Tuy nhiên, ở các quần thể châu Á, melanoma lại có xu hướng xuất hiện ở những vị trí ít tiếp xúc với ánh nắng hơn như lòng bàn tay, lòng bàn chân, dưới móng tay, móng chân (melanoma ở đầu chi – acral lentiginous melanoma), và cả ở các vùng niêm mạc như miệng, đường mũi, trực tràng và âm đạo.

Sự phát triển của melanoma thường được phân thành các giai đoạn, từ 0 đến IV, dựa trên các yếu tố như độ dày của khối u (theo thang đo Breslow), mức độ xâm lấn vào các lớp da, sự hiện diện của loét, và liệu ung thư đã lan đến các hạch bạch huyết gần đó hay các cơ quan xa hay chưa.

- **Giai đoạn 0 (Melanoma in situ):** Tế bào ung thư chỉ được tìm thấy ở lớp ngoài cùng của da (biểu bì) và chưa xâm lấn sâu hơn. Tiên lượng ở giai đoạn này rất tốt nếu được điều trị.
- **Giai đoạn I:** Khối u còn mỏng, chưa lan đến hạch bạch huyết. Tiên lượng vẫn rất khả quan.
- **Giai đoạn II:** Khối u dày hơn, có thể có loét, nhưng chưa lan đến hạch bạch huyết. Nguy cơ tái phát cao hơn Giai đoạn I.
- **Giai đoạn III:** Ung thư đã lan đến một hoặc nhiều hạch bạch huyết gần đó, hoặc đã xuất hiện các tế bào melanoma trên đường đi giữa khối u nguyên phát và hạch bạch huyết (di căn tại chỗ/di căn vệ tinh).
- **Giai đoạn IV:** Ung thư đã di căn đến các hạch bạch huyết ở xa hoặc các cơ quan nội tạng khác như phổi, gan, não, xương. Đây là giai đoạn nặng nhất với tiên lượng xấu.

Việc xác định chính xác giai đoạn bệnh là cực kỳ quan trọng để lập kế hoạch điều trị phù hợp và dự đoán kết quả.

1.1.3. Tình hình ung thư melanoma trên thế giới và tại Việt Nam

Trên toàn cầu, tỷ lệ mắc ung thư melanoma đã tăng lên trong vài thập kỷ qua, đặc biệt ở các quần thể người da trắng. Số liệu thống kê cụ thể về tỷ lệ mắc mới và tử vong do melanoma có thể thay đổi tùy theo khu vực địa lý, chủng tộc và các yếu tố môi trường.

Tại Việt Nam, theo ước tính từ Cơ quan Nghiên cứu Ung thư Quốc tế (IARC) vào năm 2018, có khoảng 157 trường hợp mắc mới melanoma. Tỷ lệ tử vong liên quan đến loại ung thư này ở Việt Nam được ghi nhận là tương đối thấp so với các loại ung thư khác, chiếm khoảng 0.09% tổng số ca tử vong do ung thư trong năm đó. Tỷ lệ mắc và tử vong do ung thư da ác tính nói chung tại Việt Nam được cho là duy trì ở mức ổn định trong những năm gần đây. So với các quốc gia có dân số chủ yếu là người da trắng như Hoa Kỳ hay Úc, melanoma ít phổ biến hơn ở các nước châu Á, bao gồm cả Việt Nam. Tuy nhiên, tính chất nguy hiểm và khả năng di căn nhanh của bệnh vẫn không thay đổi, bất kể tỷ lệ mắc. Điều này nhấn mạnh rằng việc trang bị các công nghệ chẩn đoán tiên tiến và nâng cao nhận thức cộng đồng về melanoma vẫn là một yêu cầu cấp thiết, không chỉ đối với các khu vực có tỷ lệ mắc bệnh cao mà còn để đảm bảo chất lượng chăm sóc y tế đồng đều và hiệu quả trên toàn quốc.

1.2. Tầm quan trọng của việc chẩn đoán sớm ung thư melanoma

Việc phát hiện sớm ung thư melanoma là yếu tố có ý nghĩa quyết định đối với kết quả điều trị và cơ hội sống của bệnh nhân. Sự chậm trễ trong chẩn đoán có thể dẫn đến những hậu quả nghiêm trọng.

1.2.1. Tiên lượng và hiệu quả điều trị

Mối liên hệ giữa thời điểm chẩn đoán và tiên lượng bệnh là vô cùng chặt chẽ. Khi melanoma được phát hiện ở giai đoạn sớm, khi khối u còn mỏng và chưa xâm lấn sâu hoặc di căn, tỷ lệ chữa khỏi có thể lên đến 99%. Điều trị ở giai đoạn này thường đơn giản hơn, chủ yếu là phẫu thuật cắt bỏ rộng khối u và một phần mô lành xung quanh, ít gây ảnh hưởng đến chất lượng cuộc sống của bệnh nhân.

Ngược lại, khi melanoma được chẩn đoán ở giai đoạn muộn, khi ung thư đã di căn đến các hạch bạch huyết hoặc các cơ quan xa, việc điều trị trở nên phức tạp hơn rất nhiều, thường đòi hỏi các phương pháp kết hợp như phẫu thuật, hóa trị, xạ trị, liệu pháp nhắm trúng đích hoặc liệu pháp miễn dịch. Tiên lượng ở giai đoạn này thường xấu hơn đáng kể,

và tỷ lệ sống sót sau 5 năm giảm mạnh. Do đó, mọi nỗ lực nhằm cải thiện khả năng phát hiện sớm melanoma đều mang lại lợi ích to lớn. Tỷ lệ chữa khỏi cao khi phát hiện sớm chính là động lực mạnh mẽ thúc đẩy các nghiên cứu và ứng dụng công nghệ AI, như CNN, vào việc sàng lọc và chẩn đoán, với hy vọng nhận diện được những dấu hiệu bệnh lý tinh vi mà mắt thường hoặc các phương pháp truyền thống có thể bỏ sót.

1.2.2. Các phương pháp chẩn đoán truyền thống và hạn chế

Hiện nay, có nhiều phương pháp được sử dụng để chẩn đoán ung thư melanoma:

- **Khám lâm sàng bằng mắt thường:** Bác sĩ da liễu sẽ kiểm tra toàn bộ bề mặt da của bệnh nhân, tìm kiếm các nốt ruồi hoặc tổn thương đáng ngờ dựa trên kinh nghiệm và các dấu hiệu lâm sàng.
- **Kỹ thuật Dermoscopy (Soi da):** Đây là một kỹ thuật không xâm lấn, sử dụng một thiết bị cầm tay có nguồn sáng và độ phóng đại (thường trên 10 lần) để quan sát chi tiết hơn cấu trúc của các tổn thương da. Dermoscopy giúp cải thiện đáng kể độ chính xác chẩn đoán so với khám bằng mắt thường, cho phép bác sĩ nhìn thấy các đặc điểm dưới bề mặt da mà không cần sinh thiết.
- **Sinh thiết (Biopsy):** Đây được coi là "tiêu chuẩn vàng" để khẳng định chẩn đoán ung thư melanoma. Bác sĩ sẽ lấy một mẫu nhỏ từ tổn thương nghi ngờ (sinh thiết bấm - punch biopsy) hoặc cắt bỏ toàn bộ tổn thương (sinh thiết cắt trọn - excisional biopsy) để kiểm tra dưới kính hiển vi bởi một nhà giải phẫu bệnh.

Mặc dù các phương pháp này có vai trò quan trọng, chúng vẫn có những hạn chế nhất định:

- **Tính chủ quan:** Việc đánh giá tổn thương bằng mắt thường hoặc thậm chí qua dermoscopy vẫn phụ thuộc nhiều vào kinh nghiệm và khả năng nhận định của từng bác sĩ.
- **Bỏ sót tổn thương:** Các tổn thương melanoma ở giai đoạn rất sớm hoặc có biểu hiện không điển hình có thể khó nhận diện và dễ bị bỏ sót.
- **Chẩn đoán muộn:** Đôi khi melanoma bị nhầm lẫn với các tổn thương da lành tính khác, dẫn đến sự chậm trễ trong việc chẩn đoán và điều trị. Đặc biệt, việc nhận biết melanoma ở giai đoạn đầu vẫn là một thách thức lớn ngay cả đối với các chuyên gia.
- **Tính xâm lấn của sinh thiết:** Mặc dù cần thiết để khẳng định chẩn đoán, sinh thiết là một thủ thuật xâm lấn, có thể gây đau, để lại sẹo và cần thời gian để có kết quả. Việc sinh thiết mọi tổn thương nghi ngờ là không khả thi và không cần thiết.

Những hạn chế này thúc đẩy nhu cầu phát triển các công cụ hỗ trợ chẩn đoán khách quan, nhanh chóng và không xâm lấn hơn.

1.2.3. Dấu hiệu nhận biết sớm (Quy tắc ABCDE)

Để giúp cộng đồng và các bác sĩ không chuyên khoa da liễu có thể nhận biết sớm các dấu hiệu nghi ngờ của melanoma, quy tắc ABCDE đã được phổ biến rộng rãi. Đây là một công cụ sàng lọc ban đầu hữu ích:

- **A (Asymmetry - Bất đối xứng):** Một nửa của nốt ruồi hoặc tổn thương không giống với nửa còn lại.
- **B (Border - Đường viền):** Đường viền không đều, lởm chởm, có khía hoặc mờ nhạt, không rõ ràng.
- **C (Color - Màu sắc):** Màu sắc không đồng nhất, có thể có nhiều màu khác nhau như nâu, đen, đỏ, trắng, xanh hoặc các sắc độ khác nhau của một màu.
- **D (Diameter - Đường kính):** Đường kính của tổn thương thường lớn hơn 6mm (khoảng bằng cục tẩy ở đầu bút chì), mặc dù melanoma cũng có thể nhỏ hơn.
- **E (Evolving - Tiến triển):** Nốt ruồi hoặc tổn thương có sự thay đổi về kích thước, hình dạng, màu sắc, độ cao, hoặc xuất hiện các triệu chứng mới như ngứa, chảy máu, đóng vảy.

Bất kỳ tổn thương nào có một hoặc nhiều đặc điểm trên đều cần được bác sĩ chuyên khoa da liễu thăm khám và đánh giá thêm. Các mô hình AI có thể được huấn luyện để tự động nhận diện các đặc điểm này từ hình ảnh, hoạt động như một "trợ lý ảo" cho bác sĩ hoặc thậm chí là công cụ tự kiểm tra sơ bộ cho người dùng, từ đó nâng cao khả năng phát hiện sớm.

1.3. Ứng dụng của thị giác máy tính và học sâu trong chẩn đoán hình ảnh y tế

Thị giác máy tính, một nhánh của trí tuệ nhân tạo, tập trung vào việc cho phép máy tính "nhìn" và diễn giải thông tin từ hình ảnh hoặc video. Khi kết hợp với các kỹ thuật học sâu, nó đã tạo ra một cuộc cách mạng trong lĩnh vực phân tích hình ảnh y tế.

1.3.1. Vai trò của phân tích hình ảnh y tế

Hình ảnh y tế, bao gồm nhiều phương thức như X-quang, chụp cắt lớp vi tính (CT), chụp cộng hưởng từ (MRI), siêu âm, và soi da (dermoscopy), đóng một vai trò không thể thiếu trong y học hiện đại. Chúng cung cấp những cửa sổ quan sát vào bên trong cơ thể con người, giúp bác sĩ:

- **Phát hiện bệnh lý:** Nhận diện các dấu hiệu bất thường sớm, ngay cả trước khi triệu chứng lâm sàng xuất hiện.
- **Chẩn đoán chính xác:** Xác định bản chất và mức độ của bệnh.
- **Theo dõi tiến triển bệnh:** Đánh giá sự thay đổi của tổn thương theo thời gian hoặc

đáp ứng với điều trị.

- **Lập kế hoạch điều trị:** Hướng dẫn các can thiệp phẫu thuật hoặc xạ trị.

Phân tích hình ảnh y tế không chỉ dừng lại ở việc quan sát định tính mà ngày càng hướng tới việc trích xuất các thông tin định lượng, khách quan, những đặc điểm tinh vi mà mắt thường của con người có thể bỏ qua.

1.3.2. Tiềm năng của học sâu trong việc cải thiện độ chính xác và hiệu quả chẩn đoán

Học sâu, và đặc biệt là Mạng Nơ-ron Tích chập (CNN), đã chứng tỏ khả năng vượt trội trong việc tự động học các đặc trưng phân cấp phức tạp trực tiếp từ dữ liệu hình ảnh thô, mà không cần sự can thiệp thủ công để thiết kế đặc trưng. Điều này mang lại nhiều tiềm năng to lớn cho việc cải thiện độ chính xác và hiệu quả chẩn đoán trong y tế:

- **Độ chính xác cao:** Nhiều nghiên cứu đã cho thấy các mô hình học sâu có thể đạt được, thậm chí vượt qua, độ chính xác của các chuyên gia y tế trong một số tác vụ phân tích hình ảnh cụ thể, như phân loại ung thư vú trên ảnh nhũ ảnh, phát hiện bệnh Alzheimer từ ảnh MRI não, hay phân loại các loại tổn thương da từ ảnh dermoscopy.
- **Tính khách quan và nhất quán:** Các thuật toán AI đưa ra quyết định dựa trên dữ liệu đã học, loại bỏ yếu tố chủ quan và sự mệt mỏi có thể ảnh hưởng đến con người, giúp tăng tính nhất quán trong chẩn đoán.
- **Tốc độ và hiệu quả:** Các mô hình AI có thể xử lý một lượng lớn hình ảnh trong thời gian ngắn, giúp sàng lọc nhanh chóng và giảm tải công việc cho bác sĩ, cho phép họ tập trung vào các trường hợp phức tạp hơn.
- **Phát hiện các dấu hiệu sớm và tinh vi:** Khả năng học các mẫu phức tạp giúp AI có thể nhận diện những thay đổi rất nhỏ hoặc các dấu hiệu sớm của bệnh mà có thể khó phát hiện đối với mắt người.

Trong lĩnh vực chẩn đoán melanoma, các ứng dụng của CNN và học chuyển giao đang ngày càng được khám phá và cho thấy những kết quả hứa hẹn, mở ra hy vọng về những công cụ hỗ trợ đắc lực cho bác sĩ da liễu trong tương lai.

1.4. Kết luận chương 1

Chương 1 đã trình bày một cách tổng quan về ung thư melanoma, nhấn mạnh đây là một căn bệnh da liễu nguy hiểm với khả năng di căn cao, và việc chẩn đoán sớm là yếu tố then chốt để cải thiện đáng kể tiên lượng cho bệnh nhân. Các phương pháp chẩn đoán truyền thống, mặc dù có vai trò nhất định, vẫn còn những hạn chế về tính chủ quan, độ chính xác trong một số trường hợp và yêu cầu về thời gian cũng như nguồn lực. Sự nổi lên của thị giác máy tính và các kỹ thuật học sâu, đặc biệt là CNN, mang lại một tiềm

năng to lớn trong việc phát triển các hệ thống hỗ trợ chẩn đoán hình ảnh y tế tự động, khách quan và hiệu quả hơn. Những cơ sở lý thuyết này đặt nền móng cho việc khám phá sâu hơn về các kỹ thuật học sâu cụ thể sẽ được áp dụng trong đề tài này, được trình bày chi tiết ở chương tiếp theo.

Bảng sau đây tóm tắt các loại ung thư da phổ biến và các giai đoạn phát triển của ung thư melanoma, giúp làm rõ hơn bối cảnh y khoa của nghiên cứu.

Bảng 1.1: So sánh các loại ung thư da phổ biến

Đặc điểm	Ung thư Melanoma	Ung thư Tế bào Đáy (BCC)	Ung thư Tế bào Vảy (SCC)
Tế bào gốc	Tế bào hắc tố (Melanocytes)	Tế bào đáy ở lớp biểu bì	Tế bào vảy ở lớp biểu bì
Mức độ phổ biến	Ít phổ biến hơn BCC và SCC	Loại phổ biến nhất	Loại phổ biến thứ hai
Khả năng di căn	Cao, có thể di căn sớm và nhanh	Thấp, hiếm khi di căn	Có khả năng di căn, nhưng thấp hơn melanoma
Tỷ lệ tử vong (ước tính)	Cao nhất trong các loại ung thư da nếu phát hiện muộn	Rất thấp nếu được điều trị	Thấp hơn melanoma, nhưng có thể gây tử vong nếu di căn

Bảng 1.2: Các giai đoạn phát triển của Ung thư Melanoma và Đặc điểm chính (Theo AJCC 8th Edition)

Giai đoạn	Mô tả chính (Độ dày, Loét, Di căn hạch (N), Di căn xa (M))	Tiên lượng (Tỷ lệ sống sau 5 năm ước tính)
0 (In situ)	Melanoma chỉ ở lớp biểu bì	>99%

	(Tis), chưa xâm lấn xuống lớp bì. N0, M0.	
IA	Khối u dày ≤ 1.0 mm, không loét. N0, M0.	~99%
IB	Khối u dày ≤ 1.0 mm, có loét; HOẶC khối u dày 1.01–2.0 mm, không loét. N0, M0.	~97%
IIA	Khối u dày 1.01–2.0 mm, có loét; HOẶC khối u dày 2.01–4.0 mm, không loét. N0, M0.	~90-94%
IIB	Khối u dày 2.01–4.0 mm, có loét; HOẶC khối u dày >4.0 mm, không loét. N0, M0.	~82-87%
IIC	Khối u dày >4.0 mm, có loét. N0, M0.	~75-82%
III	Ung thư đã lan đến các hạch bạch huyết vùng (N1-N3), có hoặc không có di căn tại chỗ/vệ tinh/trong quá trình vận chuyển. Bất kỳ độ dày khối u nào (T any). M0.	Thay đổi rộng (27-80+%) tùy thuộc vào số lượng hạch, gánh nặng khối u ở hạch và có loét hay không ở khối u nguyên phát.
IV	Ung thư đã di căn đến các hạch bạch huyết ở xa hoặc các cơ quan nội tạng khác (M1). Bất kỳ độ dày khối u nào (T any),	<30%, thường khoảng 15-25% tùy vị trí di căn xa.

	bất kỳ tình trạng hạch nào (N any).	
--	--------------------------------------------	--

Lưu ý: Tiên lượng là ước tính và có thể thay đổi tùy thuộc vào nhiều yếu tố cá nhân của bệnh nhân và đáp ứng điều trị.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT VỀ HỌC SÂU TRONG PHÂN LOẠI HÌNH ẢNH Y TẾ

Chương này tập trung trình bày các khái niệm và kỹ thuật học sâu nền tảng, đóng vai trò cốt lõi trong việc xây dựng mô hình phân loại ung thư melanoma. Các nội dung chính bao gồm Mạng Nơ-ron Tích chập (CNN), nguyên lý học chuyển giao (Transfer Learning), và kiến trúc Mạng Residual (ResNet). Sự hiểu biết vững chắc về các yếu tố này là cần thiết để thiết kế và triển khai một hệ thống phân tích hình ảnh y tế hiệu quả.

2.1. Mạng Nơ-ron Tích chập (Convolutional Neural Networks - CNN)

Mạng Nơ-ron Tích chập (CNN) là một lớp đặc biệt của mạng nơ-ron nhân tạo, đã tạo nên một cuộc cách mạng trong lĩnh vực thị giác máy tính và xử lý hình ảnh nhờ khả năng tự động học các đặc trưng phân cấp từ dữ liệu đầu vào.

2.1.1. Giới thiệu chung và lịch sử phát triển

CNN được thiết kế đặc biệt để xử lý dữ liệu có cấu trúc dạng lưới, chẳng hạn như hình ảnh, trong đó các pixel lân cận có mối quan hệ không gian chặt chẽ.¹ Kiến trúc của CNN lấy cảm hứng từ cách tổ chức của vỏ não thị giác ở động vật, nơi các tế bào thần kinh thị giác (nơ-ron) riêng lẻ chỉ phản ứng với các kích thích trong một vùng giới hạn của trường thị giác, được gọi là trường tiếp nhận (receptive field). Điều này có nghĩa là các nơ-ron trong các lớp đầu của CNN cũng chỉ kết nối với một vùng nhỏ của đầu vào, giúp mạng tập trung vào các đặc trưng cục bộ trước khi tổng hợp chúng thành các đặc trưng phức tạp hơn ở các lớp sâu hơn.

Lịch sử phát triển của CNN ghi nhận những cột mốc quan trọng. Bắt đầu từ những ý tưởng sơ khai như Neocognitron của Kunihiro Fukushima vào những năm 1980, một mô hình lấy cảm hứng từ hệ thống thị giác của con người có khả năng nhận dạng mẫu bất biến với sự dịch chuyển. Sau đó, Yann LeCun và các cộng sự đã phát triển LeNet-5 vào những năm 1990, một kiến trúc CNN tiên phong được ứng dụng thành công trong nhận dạng chữ số viết tay.¹ Tuy nhiên, CNN chỉ thực sự tạo nên bước đột phá và thu hút sự chú ý rộng rãi của cộng đồng nghiên cứu khi AlexNet giành chiến thắng thuyết phục trong cuộc thi ImageNet Large Scale Visual Recognition Challenge (ILSVRC) năm 2012. Thành công của AlexNet, với kiến trúc sâu hơn và việc sử dụng GPU để tăng tốc độ huấn luyện, đã mở đường cho sự phát triển của hàng loạt các kiến trúc CNN mạnh mẽ khác như VGGNet, GoogLeNet (Inception), và đặc biệt là ResNet, những kiến trúc sẽ được đề cập trong báo cáo này.¹

2.1.2. Kiến trúc cơ bản của CNN

Một mạng CNN điển hình bao gồm nhiều loại lớp khác nhau, mỗi lớp thực hiện một chức năng cụ thể để biến đổi dữ liệu đầu vào thành một biểu diễn đặc trưng hữu ích cho nhiệm vụ phân loại hoặc các tác vụ khác.¹

- **Lớp Tích chập (Convolutional Layer):**

Đây là thành phần trung tâm và mang tính định danh của CNN.¹ Lớp này thực hiện phép toán tích chập giữa ma trận đầu vào (ví dụ: một ảnh hoặc một bản đồ đặc trưng từ lớp trước) và một tập hợp các bộ lọc (filters hoặc kernels). Mỗi bộ lọc là một ma trận trọng số nhỏ, có thể học được, được thiết kế để phát hiện một đặc trưng cụ thể trong đầu vào. Ví dụ, ở các lớp đầu tiên, các bộ lọc có thể học cách phát hiện các đặc trưng cơ bản như cạnh, góc, hoặc các đốm màu. Khi đi sâu hơn vào mạng, các bộ lọc ở các lớp sau sẽ kết hợp các đặc trưng đơn giản này để nhận diện các cấu trúc phức tạp hơn, chẳng hạn như các bộ phận của một đối tượng hoặc toàn bộ đối tượng. Kết quả của mỗi phép tích chập giữa một bộ lọc và một vùng của đầu vào là một giá trị đơn lẻ. Khi bộ lọc trượt qua toàn bộ đầu vào, nó tạo ra một ma trận hai chiều gọi là bản đồ đặc trưng (feature map) hoặc bản đồ kích hoạt (activation map).¹ Bản đồ này cho biết mức độ hiện diện của đặc trưng mà bộ lọc đó tìm kiếm tại mỗi vị trí không gian trên ảnh.

Các khái niệm quan trọng liên quan đến lớp tích chập bao gồm:

- **Trường tiếp nhận (Receptive Field):** Kích thước của vùng đầu vào mà một nơ-ron trong lớp tích chập "nhìn thấy" hoặc kết nối tới. Đây chính là kích thước của bộ lọc.
 - **Sải bước (Stride):** Số lượng pixel mà bộ lọc dịch chuyển mỗi lần trượt trên đầu vào. Stride lớn hơn sẽ làm giảm kích thước không gian của bản đồ đặc trưng đầu ra.
 - **Đệm (Padding):** Việc thêm các giá trị (thường là số 0) vào xung quanh biên của ma trận đầu vào. Padding giúp kiểm soát kích thước của bản đồ đặc trưng đầu ra (ví dụ, để giữ nguyên kích thước không gian như đầu vào) và giảm thiểu việc mất thông tin ở các vùng biên của ảnh. Một đặc điểm quan trọng của lớp tích chập là **chia sẻ tham số (parameter sharing)**. Thay vì mỗi nơ-ron có một bộ trọng số riêng biệt, tất cả các nơ-ron trong cùng một bản đồ đặc trưng (tức là được tạo ra bởi cùng một bộ lọc) sẽ chia sẻ cùng một bộ trọng số của bộ lọc đó. Điều này làm giảm đáng kể tổng số lượng tham số cần học trong mạng, giúp mô hình hiệu quả hơn về mặt tính toán, giảm nguy cơ overfitting, và quan trọng hơn là giúp mô hình có khả năng bất biến với sự dịch chuyển (translation invariance) – nghĩa là mô hình có thể nhận diện một đặc trưng bất kể nó xuất hiện ở vị trí nào trong ảnh.
- **Lớp Gộp (Pooling Layer):**

Lớp gộp, hay còn gọi là lớp tổng hợp hoặc lớp lấy mẫu con (subsampling layer), thường được chèn vào giữa các lớp tích chập liên tiếp.¹ Mục đích chính của lớp pooling là giảm kích thước không gian (chiều rộng và chiều cao) của các bản đồ đặc trưng, từ đó làm giảm số lượng tham số và khối lượng tính toán trong mạng. Việc giảm kích thước này cũng giúp kiểm soát overfitting bằng cách làm cho biểu diễn đặc trưng trở nên nhỏ gọn hơn và bất biến hơn với những thay đổi nhỏ về vị trí của đặc trưng.

Có hai loại pooling phổ biến:

- **Max Pooling:** Trong một cửa sổ (kernel) trượt trên bản đồ đặc trưng, max pooling chọn ra giá trị lớn nhất và coi đó là đầu ra cho vùng đó.¹ Max pooling có xu hướng giữ lại các đặc trưng nổi bật nhất.
 - **Average Pooling:** Tính giá trị trung bình của tất cả các giá trị trong cửa sổ và coi đó là đầu ra.¹ Average pooling có xu hướng làm mịn biểu diễn đặc trưng. Trong thực tế, max pooling thường được sử dụng phổ biến hơn và cho kết quả tốt hơn trong nhiều tác vụ.
- **Hàm Kích hoạt (Activation Function):**

Sau mỗi lớp tích chập (và đôi khi sau lớp fully connected), một hàm kích hoạt phi tuyến được áp dụng cho từng phần tử của bản đồ đặc trưng (hoặc vector đầu ra).¹ Mục đích của hàm kích hoạt là đưa tính phi tuyến vào mô hình. Nếu không có các hàm kích hoạt phi tuyến, một mạng nơ-ron sâu dù có bao nhiêu lớp cũng chỉ tương đương với một mô hình tuyến tính đơn lớp, không đủ khả năng để học các mối quan hệ phức tạp trong dữ liệu.

Một số hàm kích hoạt phổ biến bao gồm:

- **Sigmoid:** $\sigma(x) = 1 / (1 + e^{(-x)})$. Hàm này đưa đầu ra về khoảng (0, 1), thường được sử dụng ở lớp cuối cùng cho bài toán phân loại nhị phân. Tuy nhiên, nó có thể gặp vấn đề vanishing gradient (gradient triệt tiêu) khi huấn luyện các mạng sâu.
- **Tanh (Hyperbolic Tangent):** $(e^x - e^{(-x)}) / (e^x + e^{(-x)})$. Đưa đầu ra về khoảng (-1, 1). Tương tự Sigmoid, cũng có thể gặp vấn đề vanishing gradient.
- **ReLU (Rectified Linear Unit):** $f(x) = \max(0, x)$ Đây là hàm kích hoạt được sử dụng rộng rãi nhất trong các mạng CNN hiện đại.¹ ReLU rất đơn giản về mặt tính toán (chỉ là một phép so sánh với 0) và giúp giảm thiểu đáng kể vấn đề vanishing gradient, cho phép huấn luyện các mạng sâu hơn và nhanh hơn. Tuy nhiên, ReLU có thể gặp vấn đề "dying ReLU" khi một số nơ-ron bị "tắt" (luôn có đầu ra bằng 0) và không còn được cập nhật.
- **Leaky ReLU:** Một biến thể của ReLU,
 $f(x) = x$ nếu $x > 0$

$f(x) = \alpha * x$ nếu $x \leq 0$ (với α là một hằng số nhỏ, ví dụ 0.01).

Leaky ReLU giải quyết vấn đề dying ReLU bằng cách cho phép một gradient nhỏ khác 0 khi đầu vào âm.

- **Softmax:** $e^{(x_i)} / \sum e^{(x_j)}$. Thường được sử dụng ở lớp cuối cùng của mạng cho bài toán phân loại đa lớp. Nó chuyển đổi một vector các giá trị thực thành một vector xác suất, trong đó mỗi phần tử biểu thị xác suất để đầu vào thuộc về một lớp cụ thể, và tổng các xác suất bằng 1.
- **Lớp Kết nối Đầy đủ (Fully Connected Layer - FC Layer):**
Sau một chuỗi các lớp tích chập và lớp pooling, các bản đồ đặc trưng bậc cao thu được thường được làm phẳng (flattened) thành một vector một chiều. Vector này sau đó được đưa vào một hoặc nhiều lớp kết nối đầy đủ, thường nằm ở cuối mạng CNN. Trong một lớp FC, mỗi nơ-ron được kết nối với tất cả các nơ-ron (hoặc các kích hoạt) ở lớp trước đó, tương tự như trong một mạng nơ-ron đa lớp (MLP) truyền thống.
Vai trò của các lớp FC là tổng hợp các đặc trưng đã được học từ các lớp trước đó và thực hiện nhiệm vụ phân loại cuối cùng. Lớp FC cuối cùng thường có số lượng nơ-ron bằng với số lớp cần phân loại, và thường sử dụng hàm kích hoạt softmax để đưa ra phân phối xác suất trên các lớp.¹

2.1.3. Quá trình huấn luyện CNN

Quá trình huấn luyện một mạng CNN là một quá trình lặp đi lặp lại nhằm điều chỉnh các trọng số (weights) và độ lệch (biases) của mạng sao cho nó có thể thực hiện tốt nhiệm vụ mong muốn (ví dụ: phân loại hình ảnh).¹ Quá trình này bao gồm các bước chính sau:

1. **Khởi tạo Trọng số:** Các trọng số của các bộ lọc trong lớp tích chập và các trọng số trong lớp FC thường được khởi tạo ngẫu nhiên với các giá trị nhỏ.
2. **Lan truyền Xuôi (Forward Propagation):** Một batch dữ liệu đầu vào (ví dụ: một tập hợp các hình ảnh) được đưa qua mạng từ lớp đầu vào đến lớp đầu ra. Tại mỗi lớp, các phép toán (tích chập, pooling, kích hoạt, kết nối đầy đủ) được thực hiện để tính toán các giá trị kích hoạt của các nơ-ron. Kết quả cuối cùng là một vector dự đoán (ví dụ: xác suất cho mỗi lớp).
3. **Tính toán Hàm Mất mát (Loss Function):** Hàm mất mát đo lường sự khác biệt (sai số) giữa đầu ra dự đoán của mạng và nhãn thực tế (ground truth) của dữ liệu đầu vào. Đối với bài toán phân loại đa lớp, hàm mất mát phổ biến là Cross-Entropy loss. Mục tiêu của quá trình huấn luyện là tìm bộ trọng số sao cho giá trị của hàm mất mát là nhỏ nhất.
4. **Lan truyền Ngược (Backward Propagation):** Đây là bước quan trọng để tính toán gradient (đạo hàm) của hàm mất mát theo từng trọng số và độ lệch trong mạng. Quá

trình này sử dụng quy tắc chuỗi (chain rule) để lan truyền sai số từ lớp đầu ra ngược trở lại các lớp đầu tiên. Gradient cho biết mỗi trọng số cần thay đổi theo hướng nào và mức độ bao nhiêu để giảm thiểu hàm mất mát.

5. **Cập nhật Trọng số (Optimization):** Dựa trên các gradient đã tính toán, các trọng số và độ lệch của mạng được cập nhật bằng một thuật toán tối ưu hóa. Các thuật toán phổ biến bao gồm Stochastic Gradient Descent (SGD) và các biến thể của nó như Adam, RMSprop. Thuật toán tối ưu hóa sử dụng một tham số gọi là tốc độ học (learning rate) để kiểm soát mức độ thay đổi của trọng số trong mỗi bước cập nhật.

Quá trình từ bước 2 đến bước 5 được lặp lại nhiều lần (qua nhiều epochs, mỗi epoch là một lượt duyệt qua toàn bộ tập dữ liệu huấn luyện) cho đến khi hàm mất mát hội tụ đến một giá trị đủ nhỏ hoặc các chỉ số hiệu năng trên tập validation không còn cải thiện đáng kể.

Trong quá trình huấn luyện, hai vấn đề thường gặp là **Overfitting (Quá khớp)** và **Underfitting (Dưới khớp)**.¹

- **Overfitting** xảy ra khi mô hình học quá tốt trên dữ liệu huấn luyện, đến mức nó bắt đầu "ghi nhớ" cả nhiễu và các chi tiết không quan trọng trong tập huấn luyện, dẫn đến hiệu suất kém trên dữ liệu mới (dữ liệu kiểm thử hoặc validation). Dấu hiệu là loss trên tập huấn luyện rất thấp nhưng loss trên tập validation lại cao, hoặc accuracy trên tập huấn luyện cao nhưng trên tập validation lại thấp. Các giải pháp bao gồm sử dụng thêm dữ liệu, tăng cường dữ liệu, kỹ thuật điều chuẩn (regularization như L1, L2, dropout), early stopping, hoặc sử dụng mô hình đơn giản hơn.
- **Underfitting** xảy ra khi mô hình quá đơn giản, không đủ khả năng để học các mẫu phức tạp trong dữ liệu huấn luyện. Kết quả là mô hình có hiệu suất thấp trên cả tập huấn luyện và tập kiểm thử. Dấu hiệu là cả loss trên tập huấn luyện và validation đều cao, hoặc accuracy trên cả hai tập đều thấp. Giải pháp có thể là sử dụng mô hình phức tạp hơn, thêm nhiều lớp hơn, huấn luyện lâu hơn, hoặc kỹ thuật hóa đặc trưng tốt hơn.

2.2. Học Chuyển giao (Transfer Learning)

Học chuyển giao là một kỹ thuật mạnh mẽ trong học máy, đặc biệt có giá trị trong các ứng dụng học sâu nơi việc thu thập dữ liệu lớn và huấn luyện mô hình từ đầu tốn nhiều tài nguyên.

2.2.1. Khái niệm và nguyên lý

Học chuyển giao (Transfer Learning) là một phương pháp học máy trong đó một mô hình

đã được phát triển và huấn luyện cho một tác vụ cụ thể (gọi là tác vụ nguồn - source task) được tái sử dụng làm điểm khởi đầu hoặc nền tảng cho một mô hình trên một tác vụ thứ hai có liên quan (gọi là tác vụ mục tiêu - target task). Thay vì bắt đầu quá trình học từ đầu (khởi tạo trọng số ngẫu nhiên), transfer learning cho phép tận dụng "kiến thức" đã được học từ tác vụ nguồn.

Nguyên lý cơ bản đằng sau transfer learning, đặc biệt trong thị giác máy tính, là các đặc trưng hình ảnh được học bởi một mạng CNN trên một bộ dữ liệu lớn và đa dạng (ví dụ, bộ dữ liệu ImageNet với hàng triệu ảnh thuộc hàng nghìn lớp đối tượng khác nhau) thường có tính tổng quát và có thể hữu ích cho nhiều tác vụ thị giác máy tính khác nhau. Ví dụ, các lớp đầu của một CNN được huấn luyện trên ImageNet thường học cách phát hiện các đặc trưng cơ bản như cạnh, góc, đường cong, màu sắc, và kết cấu. Những đặc trưng này là phổ biến và cần thiết để hiểu bất kỳ hình ảnh nào, kể cả hình ảnh y tế. Do đó, kiến thức về các đặc trưng này có thể được "chuyển giao" để giúp mô hình học nhanh hơn và tốt hơn trên một tác vụ mới, ngay cả khi tác vụ mới đó có bộ dữ liệu nhỏ hơn nhiều.

2.2.2. Lợi ích của transfer learning trong phân tích hình ảnh y tế

Trong lĩnh vực phân tích hình ảnh y tế, transfer learning mang lại nhiều lợi ích đáng kể, chủ yếu do những đặc thù của dữ liệu y tế:

- **Giảm yêu cầu về lượng dữ liệu gán nhãn:** Việc thu thập và gán nhãn dữ liệu y tế thường rất tốn kém, tốn thời gian, đòi hỏi chuyên môn cao từ các bác sĩ, và thường bị hạn chế về số lượng do các vấn đề về quyền riêng tư và tính sẵn có của bệnh nhân. Transfer learning cho phép các mô hình học sâu đạt được hiệu suất tốt ngay cả khi chỉ có một tập dữ liệu y tế được gán nhãn tương đối nhỏ, bằng cách tận dụng kiến thức từ các mô hình đã được huấn luyện trước trên các bộ dữ liệu lớn hơn (như ImageNet).
- **Giảm thời gian huấn luyện và tài nguyên tính toán:** Huấn luyện một mạng CNN sâu từ đầu (from scratch) trên một bộ dữ liệu lớn đòi hỏi năng lực tính toán rất lớn (GPU mạnh, thời gian dài). Bằng cách bắt đầu từ các trọng số đã được huấn luyện trước, mô hình có thể hội tụ nhanh hơn nhiều, tiết kiệm đáng kể thời gian và chi phí tính toán.
- **Cải thiện hiệu năng mô hình:** Các đặc trưng được học từ một bộ dữ liệu lớn và đa dạng thường có tính tổng quát và mạnh mẽ hơn so với các đặc trưng chỉ được học từ một bộ dữ liệu nhỏ. Việc sử dụng các đặc trưng này làm điểm khởi đầu có thể giúp mô hình đạt được độ chính xác cao hơn và khả năng tổng quát hóa tốt hơn trên tác vụ mục tiêu, đặc biệt khi dữ liệu mục tiêu hạn chế.

- **Khả năng áp dụng rộng rãi trên nhiều phương thức và tác vụ y tế:** Transfer learning đã được chứng minh là thành công trên một loạt các phương thức hình ảnh y tế khác nhau, bao gồm chụp cắt lớp vi tính (CT), chụp cộng hưởng từ (MRI), X-quang, siêu âm, và hình ảnh soi da (dermoscopy). Nó cũng hiệu quả cho nhiều tác vụ khác nhau như phân loại bệnh, phân đoạn tổn thương, phát hiện khối u, và đánh giá mức độ nghiêm trọng của bệnh.

Những lợi ích này làm cho transfer learning trở thành một công cụ cực kỳ giá trị và được áp dụng rộng rãi trong việc phát triển các giải pháp AI cho y tế.

2.2.3. Các chiến lược transfer learning

Có hai chiến lược chính khi áp dụng transfer learning với các mô hình CNN đã được huấn luyện trước (pre-trained models) :

1. Trích xuất Đặc trưng (Feature Extraction):

- Trong chiến lược này, mô hình pre-trained (ví dụ, ResNet50 đã huấn luyện trên ImageNet) được sử dụng như một bộ trích xuất đặc trưng cố định.
- Kiến trúc của mô hình pre-trained thường bao gồm một chuỗi các lớp tích chập (convolutional base) theo sau là một lớp phân loại (classifier head, thường là các lớp fully connected).
- Khi thực hiện feature extraction, người ta sẽ loại bỏ lớp phân loại gốc của mô hình pre-trained. Các trọng số của các lớp tích chập còn lại được giữ nguyên (đóng băng - frozen), không được cập nhật trong quá trình huấn luyện trên tác vụ mới.
- Dữ liệu của tác vụ mục tiêu mới được đưa qua phần convolutional base này. Đầu ra của một lớp trung gian nào đó (thường là lớp ngay trước lớp phân loại gốc) được coi là vector đặc trưng đã được trích xuất cho mỗi mẫu đầu vào.
- Sau đó, một bộ phân loại mới, đơn giản hơn (ví dụ: một máy vector hỗ trợ - SVM, hồi quy logistic, hoặc một vài lớp fully connected nhỏ) sẽ được huấn luyện từ đầu trên các vector đặc trưng này để thực hiện tác vụ mục tiêu.
- Chiến lược này phù hợp khi tập dữ liệu mục tiêu rất nhỏ và/hoặc có sự khác biệt đáng kể so với tập dữ liệu nguồn mà mô hình pre-trained đã học. Việc đóng băng các lớp giúp tránh overfitting trên tập dữ liệu nhỏ.

2. Tinh chỉnh Mô hình (Fine-tuning):

- Chiến lược này không chỉ sử dụng kiến trúc mà còn cả các trọng số đã học của mô hình pre-trained làm điểm khởi đầu.
- Tương tự như feature extraction, lớp phân loại gốc của mô hình pre-trained được thay thế bằng một lớp phân loại mới phù hợp với số lớp của tác vụ mục tiêu.

- Tuy nhiên, khác với feature extraction, một phần hoặc toàn bộ các trọng số của các lớp tích chập trong convolutional base cũng được "tinh chỉnh" (cập nhật) trong quá trình huấn luyện trên dữ liệu mục tiêu mới.
- Có nhiều mức độ fine-tuning:
 - **Fine-tune chỉ lớp phân loại mới:** Giữ nguyên tất cả các lớp của convolutional base (đóng băng chúng) và chỉ huấn luyện các trọng số của lớp phân loại mới thêm vào. Đây thực chất là một trường hợp của feature extraction.
 - **Fine-tune một vài lớp cuối của convolutional base:** Đóng băng các lớp đầu tiên của convolutional base (những lớp học các đặc trưng tổng quát như cạnh, góc) và chỉ cho phép các lớp cuối hơn (những lớp học các đặc trưng trừu tượng và chuyên biệt hơn) được cập nhật trọng số cùng với lớp phân loại mới.
 - **Fine-tune toàn bộ mạng:** Cho phép tất cả các trọng số trong mạng (cả convolutional base và lớp phân loại mới) được cập nhật.
- Khi fine-tuning, người ta thường sử dụng một tốc độ học (learning rate) rất nhỏ. Điều này là do các trọng số pre-trained đã ở gần một điểm tối ưu tốt (cho tác vụ nguồn), và việc cập nhật chúng quá mạnh có thể phá vỡ những kiến thức đã học.
- Fine-tuning thường mang lại kết quả tốt hơn feature extraction, đặc biệt khi tập dữ liệu mục tiêu có kích thước tương đối lớn và có độ tương đồng nhất định với tập dữ liệu nguồn.

Sự lựa chọn giữa feature extraction và các mức độ fine-tuning khác nhau phụ thuộc vào kích thước của tập dữ liệu mục tiêu và mức độ tương đồng của nó với tập dữ liệu mà mô hình pre-trained đã được huấn luyện.

2.3. Kiến trúc Mạng Residual (ResNet)

Kiến trúc Mạng Residual (ResNet) được giới thiệu bởi Kaiming He và các cộng sự vào năm 2015 đã đánh dấu một bước tiến quan trọng trong việc huấn luyện các mạng nơ-ron rất sâu, giải quyết một vấn đề cố hữu cản trở sự phát triển của các kiến trúc sâu hơn trước đó.

2.3.1. Vấn đề suy giảm độ chính xác ở các mạng sâu (Degradation Problem)

Một quan sát phổ biến trong thực nghiệm là khi các mạng nơ-ron "thuần" (plain networks - chỉ đơn giản là xếp chồng các lớp) trở nên rất sâu, hiệu suất của chúng trên tập huấn luyện bắt đầu bão hòa và sau đó suy giảm nhanh chóng. Điều đáng ngạc nhiên là sự suy giảm này không phải do overfitting (vì nó xảy ra ngay cả trên tập huấn luyện), mà là do

việc tối ưu hóa các mạng rất sâu trở nên cực kỳ khó khăn. Về mặt lý thuyết, một mạng sâu hơn nên có khả năng học một hàm phức tạp ít nhất là bằng hoặc tốt hơn một mạng nông hơn (vì mạng sâu hơn có thể chỉ cần học các lớp bổ sung dưới dạng các phép biến đổi đồng nhất - identity mappings). Tuy nhiên, trong thực tế, các bộ tối ưu hóa gặp khó khăn trong việc tìm ra các tham số cho các lớp bổ sung này để chúng hoạt động như identity mappings. Vấn đề này được gọi là "degradation problem".

2.3.2. Khối dư (Residual Block) và Kết nối tắt (Skip Connection)

Để giải quyết vấn đề suy giảm độ chính xác, ResNet giới thiệu một khái niệm kiến trúc cốt lõi gọi là **khối học dư (residual learning block)**.

Ý tưởng chính là thay vì yêu cầu một vài lớp xếp chồng lên nhau học trực tiếp một ánh xạ mong muốn cơ bản ($H(x)$) (với (x) là đầu vào của khối), ResNet đề xuất để các lớp này học một hàm dư (residual function) ($F(x) = H(x) - x$). Khi đó, ánh xạ cơ bản ban đầu được định dạng lại thành ($F(x) + x$). Việc cộng (x) (đầu vào) trực tiếp vào đầu ra của các lớp ($F(x)$) được thực hiện thông qua một **kết nối tắt (skip connection hoặc shortcut connection)**. Kết nối này bỏ qua một hoặc nhiều lớp.

Lợi ích của việc học hàm dư là nếu ánh xạ tối ưu mà khối đó cần học gần với một hàm đồng nhất (identity mapping, tức là $H(x) = x$), thì việc đẩy các trọng số của các lớp trong ($F(x)$) về gần 0 (để $F(x) \approx 0$) sẽ dễ dàng hơn nhiều so với việc học một hàm đồng nhất bằng một chuỗi các biến đổi phi tuyến phức tạp. Nói cách khác, nếu các lớp bổ sung trong một mạng sâu hơn không cần thiết, chúng có thể dễ dàng học cách "không làm gì cả" bằng cách làm cho hàm dư của chúng bằng không, cho phép tín hiệu đầu vào đi qua gần như nguyên vẹn.

2.3.3. Ưu điểm của ResNet trong việc huấn luyện mạng rất sâu

Kiến trúc residual learning mang lại nhiều ưu điểm quan trọng, cho phép huấn luyện các mạng nơ-ron sâu hơn nhiều so với trước đây:

- **Giải quyết vấn đề suy giảm độ chính xác:** Bằng cách làm cho việc học các identity mappings trở nên dễ dàng, ResNet cho phép các mạng rất sâu (hàng trăm, thậm chí hàng nghìn lớp) được huấn luyện hiệu quả mà không bị suy giảm độ chính xác.¹⁴ Nếu một lớp nào đó không hữu ích, mạng có thể dễ dàng "bỏ qua" nó.
- **Cải thiện dòng chảy gradient:** Các kết nối tắt cung cấp một đường dẫn trực tiếp cho gradient lan truyền ngược từ các lớp sau về các lớp trước. Điều này giúp giảm thiểu vấn đề vanishing/exploding gradient, làm cho quá trình tối ưu hóa trở nên dễ dàng và ổn định hơn, ngay cả với các mạng rất sâu.

- **Cho phép xây dựng các mạng sâu hơn:** Trước ResNet, việc huấn luyện các mạng có vài chục lớp đã là một thách thức. ResNet đã mở đường cho các kiến trúc có hàng trăm hoặc thậm chí hơn 1000 lớp, giúp tăng khả năng biểu diễn của mô hình và đạt được kết quả tiên tiến trên nhiều bộ dữ liệu chuẩn.

2.3.4. Giới thiệu về ResNet50

ResNet50 là một biến thể cụ thể và rất phổ biến của kiến trúc ResNet, bao gồm 50 lớp có trọng số (49 lớp tích chập và 1 lớp fully connected). Kiến trúc này bao gồm các khối dư được xếp chồng lên nhau, với các kết nối tắt được thực hiện sau mỗi hai hoặc ba lớp tích chập. ResNet50 đã được huấn luyện trước trên bộ dữ liệu ImageNet (chứa hơn một triệu ảnh thuộc 1000 lớp đối tượng) và các trọng số pre-trained của nó thường được sử dụng làm điểm khởi đầu mạnh mẽ cho nhiều tác vụ transfer learning trong lĩnh vực thị giác máy tính, bao gồm cả các ứng dụng phân tích hình ảnh y tế như phân loại ung thư da. Việc sử dụng ResNet50 pre-trained cho phép các nhà nghiên cứu tận dụng các đặc trưng hình ảnh phong phú đã được học từ ImageNet, giúp cải thiện đáng kể hiệu suất và giảm thời gian huấn luyện trên các bộ dữ liệu y tế thường nhỏ hơn.

Sự kết hợp giữa khả năng học đặc trưng tự động của CNN, tính hiệu quả của transfer learning khi đối mặt với dữ liệu hạn chế, và sức mạnh của kiến trúc ResNet trong việc huấn luyện các mô hình sâu, tạo thành một chuỗi công nghệ tiên hóa. Mỗi công nghệ giải quyết một hạn chế của công nghệ trước đó hoặc mở ra một khả năng mới, dẫn đến phương pháp luận hiện tại của đề tài này. Nguyên lý "chia sẻ tham số" trong lớp tích chập của CNN và ý tưởng "học hàm dư" ($F(x)=0$) trong ResNet đều hướng đến việc làm cho mạng dễ học hơn và giảm độ phức tạp không cần thiết, phản ánh một chủ đề ngầm về việc tối ưu hóa khả năng học (learnability) trong thiết kế mạng nơ-ron.

2.4. Kết luận chương 2

Chương 2 đã trình bày chi tiết các cơ sở lý thuyết quan trọng làm nền tảng cho nghiên cứu này. Mạng Nơ-ron Tích chập (CNN) được giới thiệu như một công cụ mạnh mẽ cho việc xử lý và trích xuất đặc trưng từ dữ liệu hình ảnh. Tiếp theo, khái niệm Học Chuyển giao (Transfer Learning) được làm rõ, nhấn mạnh lợi ích của việc tận dụng kiến thức từ các mô hình đã được huấn luyện trước trên các bộ dữ liệu lớn, đặc biệt hữu ích khi làm việc với dữ liệu y tế thường bị hạn chế về số lượng. Cuối cùng, kiến trúc Mạng Residual (ResNet), với các khối dư và kết nối tắt đặc trưng, được trình bày như một giải pháp hiệu quả để xây dựng và huấn luyện các mô hình học sâu có độ phức tạp cao mà không gặp phải vấn đề suy giảm độ chính xác. Sự kết hợp của ba trụ cột lý thuyết này – CNN, Transfer Learning, và ResNet – cung cấp một phương pháp luận vững chắc và đầy hứa

hện cho bài toán phân loại ung thư melanoma trên bộ dữ liệu HAM10000, sẽ được cụ thể hóa trong các chương tiếp theo. Sự thành công của transfer learning với các mô hình như ResNet50 trên ImageNet cho thấy một lượng lớn "kiến thức thị giác tổng quát" đã được mã hóa, và việc "chuyển giao" kiến thức này sang lĩnh vực y tế chuyên biệt không chỉ là một giải pháp kỹ thuật mà còn ngụ ý rằng có những đặc trưng hình ảnh cơ bản chung giữa ảnh tự nhiên và ảnh y tế mà máy móc có thể học và tận dụng.

Bảng sau đây so sánh các chiến lược transfer learning, cung cấp thông tin hữu ích cho việc thiết kế thực nghiệm.

Bảng 2.1: Các chiến lược Transfer Learning và trường hợp áp dụng

Chiến lược	Mô tả ngắn	Kích thước tập dữ liệu mục tiêu	Độ tương đồng dữ liệu mục tiêu và nguồn	Ưu điểm	Nhược điểm
Trích xuất Đặc trưng (Feature Extraction)	Sử dụng các lớp pre-trained làm bộ trích xuất đặc trưng cố định, huấn luyện bộ phân loại mới.	Nhỏ	Thấp/Cao	Nhanh, ít tốn tài nguyên, tránh overfitting trên dữ liệu nhỏ.	Có thể không tận dụng hết tiềm năng của mô hình pre-trained nếu đặc trưng nguồn không hoàn toàn phù hợp.
Fine-tuning một phần (Partial Fine-tuning)	Đóng băng các lớp đầu, tinh chỉnh các lớp sau của mô hình pre-trained cùng bộ phân loại mới.	Vừa/Lớn	Vừa/Cao	Cân bằng giữa việc giữ lại kiến thức chung và học kiến thức chuyên biệt, thường cho kết quả tốt hơn.	Cần nhiều dữ liệu hơn và thời gian huấn luyện lâu hơn so với feature extraction.
Fine-tuning toàn bộ (Full Fine-tuning)	Tinh chỉnh tất cả các lớp của mô hình pre-trained (với learning rate nhỏ) cùng bộ phân loại mới.	Lớn	Cao	Có khả năng đạt hiệu suất cao nhất nếu dữ liệu mục tiêu đủ lớn và tương đồng.	Nguy cơ overfitting cao nếu dữ liệu mục tiêu nhỏ, tốn nhiều tài nguyên và thời gian huấn luyện nhất.

CHƯƠNG 3: PHƯƠNG PHÁP LUẬN VÀ THIẾT KẾ HỆ THỐNG PHÂN LOẠI UNG THƯ MELANOMA

Chương này trình bày chi tiết về phương pháp luận được sử dụng để xây dựng hệ thống phân loại ung thư melanoma. Nội dung bao gồm mô tả bộ dữ liệu HAM10000, các kỹ thuật tiền xử lý dữ liệu, kiến trúc mô hình được lựa chọn và chiến lược học chuyển giao cụ thể, cùng với việc thiết lập môi trường thực nghiệm.

3.1. Bộ dữ liệu HAM10000

Việc lựa chọn một bộ dữ liệu phù hợp và hiểu rõ đặc điểm của nó là yếu tố then chốt cho sự thành công của bất kỳ dự án học máy nào, đặc biệt trong lĩnh vực y tế nơi chất lượng và tính đại diện của dữ liệu ảnh hưởng trực tiếp đến độ tin cậy của mô hình.

3.1.1. Mô tả chi tiết bộ dữ liệu

Bộ dữ liệu được sử dụng trong nghiên cứu này là **HAM10000 (Human Against Machine with 10000 training images)**. Đây là một bộ dữ liệu công khai và được sử dụng rộng rãi trong cộng đồng nghiên cứu về phân tích hình ảnh da liễu bằng học máy.

- **Nguồn gốc:** Bộ dữ liệu được thu thập trong khoảng thời gian 20 năm từ hai nguồn chính: Khoa Da liễu của Đại học Y Vienna, Áo (nhóm ViDIR), và một phòng khám chuyên về ung thư da của bác sĩ Cliff Rosendahl ở Queensland, Úc. Sự đa dạng về nguồn gốc này có thể mang lại sự phong phú cho dữ liệu nhưng cũng đặt ra thách thức về tính đồng nhất.
- **Số lượng ảnh:** HAM10000 bao gồm tổng cộng 10.015 hình ảnh soi da (dermoscopic images) của các tổn thương da có sắc tố.
- **Các lớp tổn thương:** Các hình ảnh trong bộ dữ liệu được phân loại thành 7 loại tổn thương da khác nhau. Sự phân bố số lượng ảnh cho mỗi lớp như sau:
 1. **Melanocytic nevi (nv):** 6705 ảnh. Đây là các nốt ruồi lành tính, chiếm tỷ lệ lớn nhất trong bộ dữ liệu (hơn 60%) và là lớp đa số.
 2. **Melanoma (mel):** 1112 ảnh (một số nguồn ghi 1113). Đây là lớp ung thư hắc tố ác tính, mục tiêu chính của việc phát hiện.
 3. **Benign keratosis-like lesions (bkl):** 1098 ảnh (một số nguồn ghi 1099). Lớp này bao gồm các tổn thương lành tính như dày sừng tiết bã (seborrheic keratoses), đốm nâu do ánh nắng (solar lentigo), và dày sừng dạng lichen phẳng (lichen-planus like keratoses).
 4. **Basal cell carcinoma (bcc):** 514 ảnh. Đây là một dạng ung thư da phổ biến, ít di căn hơn melanoma.
 5. **Actinic keratoses and intraepithelial carcinoma (akiec):** 327 ảnh. Đây là các

tổn thương tiền ung thư hoặc ung thư biểu mô tại chỗ, thường do tiếp xúc với ánh nắng mặt trời.

6. **Vascular lesions (vasc):** 142 ảnh. Lớp này bao gồm các tổn thương mạch máu lành tính như u mạch anh đào (cherry angiomas), u mạch dày sừng (angiokeratomas), u hạt sinh mủ (pyogenic granulomas), và xuất huyết.
7. **Dermatofibroma (df):** 115 ảnh. Đây là một dạng u xơ da lành tính, là lớp có số lượng mẫu ít nhất trong bộ dữ liệu.
- **Phương pháp xác thực chẩn đoán (Ground Truth):** Độ tin cậy của nhãn dữ liệu là rất quan trọng. Trong HAM10000, hơn 50% các tổn thương đã được xác nhận chẩn đoán bằng phương pháp giải phẫu bệnh (histopathology), được coi là tiêu chuẩn vàng. Đối với các trường hợp còn lại, chẩn đoán được xác định thông qua việc theo dõi diễn biến của tổn thương theo thời gian (follow-up), sự đồng thuận của các chuyên gia da liễu (expert consensus), hoặc xác nhận bằng kính hiển vi đồng tiêu phản xạ tại chỗ (in-vivo reflectance confocal microscopy).
- **Mục đích tạo bộ dữ liệu:** HAM10000 được công bố với mục đích chính là cung cấp một bộ dữ liệu huấn luyện quy mô lớn và đa dạng cho cộng đồng nghiên cứu học máy, nhằm thúc đẩy sự phát triển các thuật toán chẩn đoán tự động tổn thương da. Nó cũng đóng vai trò như một bộ dữ liệu benchmark để so sánh hiệu năng giữa các thuật toán khác nhau và so sánh khả năng của máy với các chuyên gia con người.
- **Cách truy cập:** Bộ dữ liệu HAM10000 có thể được truy cập công khai thông qua nhiều nguồn, bao gồm Harvard Dataverse và kho lưu trữ của International Skin Imaging Collaboration (ISIC).

3.1.2. Phân tích đặc điểm và thách thức của bộ dữ liệu

Mặc dù HAM10000 là một nguồn tài nguyên quý giá, nó cũng mang trong mình những đặc điểm và thách thức cần được xem xét kỹ lưỡng khi xây dựng mô hình:

- **Mất cân bằng lớp (Class Imbalance):** Đây là thách thức nổi bật nhất của HAM10000. Lớp 'nv' (nốt ruồi lành tính) chiếm một tỷ lệ áp đảo (hơn 60%), trong khi các lớp khác, đặc biệt là 'df' (dermatofibroma) và 'vasc' (tổn thương mạch máu), có số lượng mẫu rất ít. Sự mất cân bằng nghiêm trọng này có thể khiến mô hình học máy bị thiên vị, có xu hướng dự đoán tốt cho các lớp đa số nhưng hoạt động kém hiệu quả trên các lớp thiểu số, vốn thường là các lớp bệnh lý quan trọng cần phát hiện. Đây là nguyên nhân trực tiếp dẫn đến sự cần thiết phải áp dụng các kỹ thuật tăng cường dữ liệu và có thể cả các hàm mất mát đặc biệt để giảm thiểu tác động tiêu cực này.
- **Đa dạng nguồn gốc và phương thức thu thập ảnh:** Việc thu thập ảnh từ hai địa điểm khác nhau (Áo và Úc) và sử dụng các thiết bị, điều kiện chụp khác nhau qua

nhiều năm có thể dẫn đến sự không đồng nhất về chất lượng hình ảnh, độ phân giải, điều kiện ánh sáng, và sự hiện diện của các yếu tố nhiễu (ví dụ: tóc, thước đo, bọt khí trên gel soi da). Điều này đòi hỏi các bước tiền xử lý cẩn thận để chuẩn hóa dữ liệu.

- **Sự tương đồng hình ảnh giữa các lớp:** Một số loại tổn thương lành tính có thể có đặc điểm hình ảnh rất giống với melanoma ở giai đoạn sớm, gây khó khăn không chỉ cho các mô hình AI mà ngay cả đối với các bác sĩ da liễu kinh nghiệm. Điều này làm tăng độ phức tạp của bài toán phân loại.
- **Chất lượng nhãn:** Mặc dù phần lớn các ca được xác nhận bằng giải phẫu bệnh, một số ca vẫn dựa trên theo dõi hoặc đồng thuận chuyên gia, vốn có thể tiềm ẩn một mức độ không chắc chắn nhất định.

3.2. Tiền xử lý dữ liệu

Tiền xử lý dữ liệu là một bước cực kỳ quan trọng trong quy trình xây dựng mô hình học máy, nhằm đảm bảo dữ liệu đầu vào có chất lượng tốt, đồng nhất và phù hợp với yêu cầu của mô hình, đồng thời giải quyết các vấn đề như mất cân bằng lớp.

3.2.1. Chuẩn hóa kích thước ảnh

Các mô hình CNN, đặc biệt là các kiến trúc đã được huấn luyện trước như ResNet50, thường yêu cầu đầu vào có kích thước cố định. Do đó, tất cả các hình ảnh trong bộ dữ liệu HAM10000 được thay đổi kích thước (resize) về một kích thước chung. Đối với ResNet50, kích thước đầu vào tiêu chuẩn thường là 224×224 pixel. Tuy nhiên, trong nghiên cứu này, ảnh được resize về 256×256 để phù hợp với thiết lập của mô hình huấn luyện và tận dụng tốt hơn độ phân giải ban đầu của ảnh. Quá trình này đảm bảo tính nhất quán cho đầu vào của mạng.

3.2.2. Chuẩn hóa giá trị pixel

Giá trị của mỗi pixel trong ảnh (thường từ 0 đến 255 cho mỗi kênh màu) cần được chuẩn hóa. Có hai cách tiếp cận phổ biến:

- **Chia tỷ lệ về khoảng :** Đơn giản là chia tất cả các giá trị pixel cho 255.
- **Chuẩn hóa Z-score (Standardization):** Trừ đi giá trị trung bình và chia cho độ lệch chuẩn của tập dữ liệu. Khi sử dụng một mô hình pre-trained trên ImageNet, người ta thường chuẩn hóa ảnh đầu vào mới bằng cách sử dụng giá trị trung bình và độ lệch chuẩn của chính tập ImageNet mà mô hình đó đã được huấn luyện. Điều này giúp đảm bảo rằng phân phối của dữ liệu đầu vào mới gần với phân phối của dữ liệu mà mô hình đã "quen thuộc".

3.2.3. Kỹ thuật tăng cường dữ liệu (Data Augmentation)

Do vấn đề mất cân bằng lớp nghiêm trọng trong HAM10000, kỹ thuật tăng cường dữ liệu đóng một vai trò thiết yếu.⁴ Mục đích của data augmentation là tạo ra các mẫu huấn luyện mới từ các mẫu hiện có thông qua các phép biến đổi hình học hoặc màu sắc nhỏ, mà không làm thay đổi nhãn của chúng. Điều này giúp:

- **Tăng kích thước hiệu dụng của tập huấn luyện:** Đặc biệt quan trọng cho các lớp thiểu số.
- **Tăng sự đa dạng của dữ liệu:** Giúp mô hình học được các đặc trưng bất biến hơn.
- **Giảm overfitting:** Khiến mô hình ít có khả năng "ghi nhớ" các mẫu cụ thể trong tập huấn luyện và tổng quát hóa tốt hơn trên dữ liệu mới.

Trong nghiên cứu này, một tập hợp các kỹ thuật tăng cường dữ liệu đã được áp dụng bằng thư viện ImageDataGenerator của Keras, bao gồm:

- **Xoay ảnh (Rotation):** Xoay ảnh một góc ngẫu nhiên (20 độ).
- **Lật ngang ảnh (Horizontal Flipping):** Lật ảnh theo chiều ngang
- **Dịch chuyển (Shifting):** width/height_shift_range=0.05 dịch chuyển ảnh một khoảng nhỏ theo chiều ngang và chiều dọc.
- **Phóng to/Thu nhỏ (Zooming):** Phóng to nhẹ zoom_range=0.1

Việc lựa chọn các kỹ thuật và tham số tăng cường dữ liệu cần được thực hiện cẩn thận để đảm bảo rằng các biến đổi là hợp lý về mặt y học và không tạo ra các mẫu giả mạo có thể gây nhầm lẫn cho mô hình. Các thư viện phổ biến như Albumentations hoặc ImageDataGenerator của Keras thường được sử dụng để thực hiện các thao tác này một cách hiệu quả.

3.2.4. Chiến lược xử lý mất cân bằng lớp:

Vấn đề mất cân bằng lớp nghiêm trọng trong bộ dữ liệu HAM10000, nơi lớp 'nv' (Nevus - nốt ruồi) chiếm đa số áp đảo trong khi các lớp bệnh lý quan trọng khác như 'mel' (Melanoma), 'df' (Dermatofibroma) và 'vasc' (Vascular lesions) lại là thiểu số, đặt ra một thách thức lớn. Trong nghiên cứu này, chiến lược giải quyết vấn đề này bao gồm hai cấu phần chính:

- **Sử dụng RandomOverSampler:** Kỹ thuật này đã được áp dụng ở giai đoạn chuẩn bị dữ liệu ban đầu (như đã mô tả ở mục trước) để tăng số lượng mẫu của các lớp thiểu số trong tập huấn luyện trước khi lưu ảnh ra các thư mục. Đây là bước khởi đầu quan trọng để giảm bớt sự chênh lệch.
- **Điều chỉnh trọng số lớp (class_weights):** Song song với việc tăng cường dữ liệu,

một chiến lược quan trọng khác là sử dụng `class_weights` một cách chủ động trong cả hai giai đoạn huấn luyện mô hình. Trong giai đoạn fine-tuning, trọng số của lớp 'mel' được giảm xuống còn 0.8 nhằm giảm tỷ lệ false positive và cải thiện precision. Việc giảm nhẹ trọng số thay vì tăng giúp mô hình trở nên ít thiên lệch hơn với lớp này, vốn thường bị chẩn đoán nhầm từ các tổn thương lành tính có đặc điểm tương tự.

Sự kết hợp giữa `RandomOverSampler` ban đầu, các kỹ thuật tăng cường dữ liệu phong phú, và việc tinh chỉnh `class_weights` một cách có chủ đích được kỳ vọng sẽ giải quyết hiệu quả hơn vấn đề mất cân bằng, từ đó cải thiện hiệu năng tổng thể và đặc biệt là độ nhạy (Recall) đối với các lớp bệnh lý quan trọng.

Các trình sinh dữ liệu (`train_generator`, `valid_generator`, `test_generator`) sau đó được tạo ra từ các `ImageDataGenerator` tương ứng, với các tham số chung như `target_size=(256, 256)` để đồng bộ kích thước ảnh đầu vào cho ResNet50, `batch_size=32`, và `class_mode='categorical'` do bài toán là phân loại đa lớp và sử dụng hàm mất mát `categorical_crossentropy`. Đối với `test_generator`, tham số `shuffle=False` được thiết lập để đảm bảo thứ tự các dự đoán khớp với thứ tự ảnh gốc khi đánh giá.

3.2.5. Phân chia tập dữ liệu (Training/Validation/Test sets)

Việc phân chia hợp lý bộ dữ liệu là bước quan trọng nhằm đảm bảo khả năng huấn luyện, kiểm định và đánh giá mô hình một cách khách quan, đặc biệt với các bộ dữ liệu mất cân bằng như HAM10000.

Trong nghiên cứu này, bộ dữ liệu HAM10000 được chia thành ba tập con theo các bước sau:

- **Bước 1: Tách tập kiểm thử (Test set)**

Sử dụng hàm `train_test_split` từ scikit-learn, toàn bộ tập dữ liệu ban đầu được tách thành:

- **80% dùng để huấn luyện (Training + Validation)**
- **20% dùng để kiểm thử (Test)**

Phép chia này hiện tại chưa áp dụng stratify theo nhãn, do đó tỷ lệ các lớp trong tập test có thể chênh lệch nhẹ so với dữ liệu gốc. Tuy nhiên, với việc đã áp dụng `RandomOverSampler` trước đó, tập huấn luyện đã được cân bằng lại để hỗ trợ mô hình học tốt hơn.

- **Bước 2: Tách tập kiểm định (Validation set)**

Trong quá trình huấn luyện, thư viện `ImageDataGenerator` của Keras được sử dụng để

tự động chia **10% của tập huấn luyện** thành tập kiểm định (validation) bằng cách thiết lập tham số `validation_split=0.1`. Điều này tương đương với việc sử dụng:

- **72% dữ liệu cho huấn luyện**
- **8% dữ liệu cho kiểm định (validation)**
- **20% dữ liệu cho kiểm thử (test)**

Mặc dù việc phân chia theo cách phân tầng (stratified split) chưa được áp dụng trong đoạn mã hiện tại, đây là một phương pháp nên được xem xét để đảm bảo rằng các lớp thiểu số (như df, vas, mel) vẫn xuất hiện đủ trong cả ba tập. Điều này đặc biệt quan trọng trong các bài toán y sinh học, nơi mà mô hình cần phải học cách phân biệt các loại tổn thương ít gặp nhưng nguy hiểm. Ngoài ra, việc giữ cho tập kiểm thử hoàn toàn "vô hình" (unseen) đối với mô hình trong suốt quá trình huấn luyện giúp đảm bảo đánh giá được khả năng tổng quát hóa thực sự của mô hình.

3.3. Xây dựng mô hình phân loại

Sau khi dữ liệu đã được chuẩn bị, bước tiếp theo là lựa chọn kiến trúc mô hình và chiến lược huấn luyện.

3.3.1. Lựa chọn mô hình ResNet50 làm mô hình cơ sở

Trong nghiên cứu này, **ResNet50** được chọn làm kiến trúc CNN cơ sở. Lý do cho sự lựa chọn này bao gồm:

- **Hiệu suất đã được chứng minh:** ResNet50 là một trong những kiến trúc ResNet phổ biến nhất và đã đạt được hiệu suất rất tốt trên bộ dữ liệu ImageNet, cũng như trong nhiều tác vụ phân tích hình ảnh y tế khác nhau.
- **Kiến trúc sâu:** Với 50 lớp, ResNet50 có khả năng học các biểu diễn đặc trưng phức tạp và phân cấp từ hình ảnh, điều này rất quan trọng để phân biệt các loại tổn thương da tinh vi.
- **Sẵn có trọng số pre-trained:** Các trọng số của ResNet50 đã được huấn luyện trước trên ImageNet được cung cấp rộng rãi trong các thư viện học sâu, tạo điều kiện thuận lợi cho việc áp dụng transfer learning.

3.3.2. Chiến lược Transfer Learning

Với ResNet50 làm mô hình cơ sở và bộ dữ liệu HAM10000, chiến lược transfer learning được áp dụng sẽ là **fine-tuning (tinh chỉnh)**. Cụ thể:

1. **Khởi tạo Mô hình và Đóng băng các Lớp Gốc (Feature Extraction Phase):**
 - Tải kiến trúc ResNet50, sử dụng các trọng số đã huấn luyện trên ImageNet, nhưng **loại bỏ lớp phân loại gốc**. Phần còn lại của mô hình, được gọi là "convolutional base" (các lớp tích chập), sẽ được sử dụng làm bộ trích xuất đặc trưng.
 - Ban đầu, toàn bộ các trọng số của convolutional base này được **đóng băng** (ví dụ: thiết lập `layer.trainable = False` cho các lớp này). Điều này nhằm giữ lại các đặc trưng tổng quát (như cạnh, góc, màu sắc) mà mô hình đã học được từ ImageNet và tránh việc chúng bị cập nhật không mong muốn bởi dữ liệu HAM10000 (vốn nhỏ hơn) trong giai đoạn đầu.
2. **Xây dựng và Huấn luyện Phần Đầu Phân loại Mới (Classifier Head Training):**
 - Một "phần đầu phân loại" (classifier head) mới được xây dựng và gắn vào phía trên convolutional base đã đóng băng. Phần đầu này bao gồm:
 - Một lớp GlobalAveragePooling2D để giảm chiều.
 - Một lớp Dropout(0.65) nhằm giảm overfitting.
 - Một lớp BatchNormalization.
 - Cuối cùng là một lớp Dense(7) sử dụng hàm kích hoạt softmax để phân loại 7 loại tổn thương da.
 - Ở giai đoạn này, mô hình được biên dịch (compile) và huấn luyện. Do các lớp gốc đã bị đóng băng, **chỉ các trọng số của phần đầu phân loại mới này được cập nhật** dựa trên dữ liệu HAM10000. Giai đoạn này giúp phần đầu phân loại học cách ánh xạ các đặc trưng trích xuất từ ResNet50 vào các lớp mục tiêu của bài toán.
3. **Mở băng một phần và Tinh chỉnh Toàn bộ Mô hình (Fine-tuning Phase):**
 - Sau khi phần đầu phân loại đã được huấn luyện và cho thấy sự hội tụ nhất định, quá trình fine-tuning thực sự bắt đầu.
 - Một số lớp tích chập ở tầng sâu hơn của một phần của convolutional base được mở khóa, cụ thể là từ lớp thứ 100 trở đi sẽ được **mở băng**. Quyết định mở băng bao nhiêu lớp phụ thuộc vào kích thước bộ dữ liệu HAM10000 và độ tương đồng của nó với ImageNet; thường các lớp sâu hơn (gần output hơn) sẽ được ưu tiên mở băng trước vì chúng học các đặc trưng phức tạp và chuyên biệt hơn.
 - Mô hình được **biên dịch lại (re-compile)**, lần này với một **tốc độ học (learning rate) rất nhỏ** (ví dụ, 10^{-5} hoặc 10^{-6}).
 - Toàn bộ mô hình (bao gồm cả các lớp đã mở băng và phần đầu phân loại) sau đó được tiếp tục huấn luyện trên HAM10000. Tốc độ học nhỏ là cực kỳ quan trọng để chỉ "tinh chỉnh" nhẹ nhàng các trọng số đã được huấn luyện trước trên ImageNet, giúp chúng thích ứng tốt hơn với đặc điểm cụ thể của hình ảnh tổn thương da mà không phá hủy những kiến thức tổng quát đã học.

Chiến lược fine-tuning hai giai đoạn này cho phép mô hình tận dụng hiệu quả kiến thức từ bộ dữ liệu lớn (ImageNet), đồng thời chuyên biệt hóa cho tác vụ cụ thể trên bộ dữ liệu nhỏ hơn (HAM10000). Việc huấn luyện ban đầu chỉ phần đầu phân loại giúp ổn định quá trình, trong khi giai đoạn tinh chỉnh sau đó cho phép mô hình học các đặc trưng ở mức độ chi tiết hơn, phù hợp với bộ dữ liệu mục tiêu. Quyết định số lượng lớp mở băng và giá trị learning rate là các siêu tham số quan trọng cần được cân nhắc và thử nghiệm.

3.3.3. Điều chỉnh kiến trúc phần đầu ra (Output Head)

Như đã đề cập, phần đầu ra của mô hình ResNet50 sẽ được tùy chỉnh cho bài toán phân loại 7 lớp của HAM10000:

- Sau phần convolutional base của ResNet50, một lớp GlobalAveragePooling2D được sử dụng để rút gọn feature map thành vector đặc trưng 1 chiều.
- Sau đó, một lớp Dropout(0.65) được thêm vào nhằm giảm overfitting, tiếp theo là BatchNormalization.
- Cuối cùng là một lớp Dense(7) với hàm kích hoạt softmax để phân loại ảnh thành 7 loại tổn thương da khác nhau.
- Không sử dụng thêm các lớp Dense trung gian nhằm giữ mô hình đơn giản và tránh overfitting do kích thước tập dữ liệu không quá lớn.

3.4. Thiết lập môi trường và công cụ thực nghiệm

Việc thực nghiệm sẽ được tiến hành trong một môi trường được thiết lập cẩn thận để đảm bảo tính nhất quán và khả năng tái tạo.

3.4.1. Phần cứng (CPU, GPU)

Quá trình huấn luyện và đánh giá mô hình được thực hiện trên nền tảng Kaggle Notebooks, với cấu hình phần cứng sau:

- CPU: (CPU do Kaggle cung cấp, thường là các dòng Intel Xeon).
- GPU: NVIDIA P100 (được cung cấp bởi Kaggle).

3.4.2. Phần mềm

Môi trường phần mềm sẽ bao gồm:

- **Ngôn ngữ lập trình:** Python (phiên bản 3.11).
- **Thư viện học sâu:** TensorFlow (phiên bản 2.18) với Keras API được tích hợp sẵn. Keras cung cấp một giao diện người dùng bậc cao, thân thiện để xây dựng và huấn luyện các mô hình nơ-ron, trong khi TensorFlow cung cấp backend tính toán hiệu

quả. Hoặc có thể sử dụng PyTorch, một thư viện học sâu phổ biến khác.

- **Các thư viện hỗ trợ:**

- **NumPy:** Cho các thao tác mảng và tính toán số học hiệu quả.
- **Pandas:** Để đọc, xử lý và phân tích dữ liệu dạng bảng (ví dụ: metadata, kết quả).
- **ImageIO:** Để đọc và ghi các định dạng hình ảnh.
- **Scikit-learn (sklearn):** Sử dụng cho các tiện ích như train_test_split (phân chia dữ liệu) và LabelEncoder (mã hóa nhãn).
- **Imbalanced-learn (imblearn):** Cụ thể là RandomOverSampler để xử lý vấn đề mất cân bằng dữ liệu trong quá trình huấn luyện.
- **Matplotlib và Seaborn:** Để trực quan hóa dữ liệu, đồ thị quá trình huấn luyện và kết quả mô hình.
- **Scikeras:** Để tích hợp mô hình Keras với các chức năng của Scikit-learn.

3.5. Kết luận chương 3

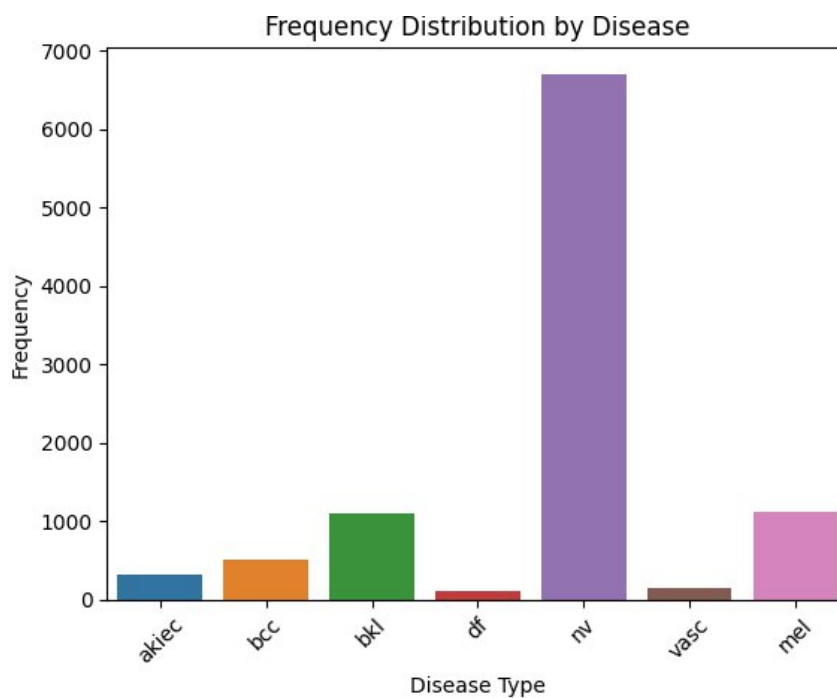
Chương 3 trình bày phương pháp luận nghiên cứu. Đầu tiên, chương mô tả và phân tích bộ dữ liệu HAM10000, bao gồm đặc điểm, thách thức và vấn đề mất cân bằng lớp. Tiếp đó, các bước tiền xử lý quan trọng như chuẩn hóa kích thước ảnh, giá trị pixel, các kỹ thuật tăng cường dữ liệu và phân chia tập dữ liệu được thảo luận. Trọng tâm của chương là việc chọn ResNet50 làm mô hình cơ sở, áp dụng chiến lược transfer learning (fine-tuning) và thiết kế đầu ra phù hợp cho bài toán 7 lớp. Môi trường thực nghiệm (phần cứng, phần mềm) cũng được xác định. Những chuẩn bị này tạo tiền đề cho thực nghiệm và phân tích kết quả ở chương sau. Việc chuẩn hóa tiền xử lý và lựa chọn mô hình ResNet50 là quan trọng để đạt kết quả tốt, đồng thời đảm bảo khả năng so sánh với các nghiên cứu khác trên HAM10000, thúc đẩy tính tái lập và tiến bộ khoa học. Bảng sau đây cung cấp cái nhìn tổng quan về phân bố dữ liệu và các kỹ thuật tăng cường dữ liệu dự kiến.

Bảng 3.1: Phân bố chi tiết các lớp tổn thương trong bộ dữ liệu HAM10000 (Trước khi tăng cường)

Tên lớp (Tiếng Việt)	Mã lớp (Tiếng Anh)	Số lượng ảnh gốc	Tỷ lệ (%)
Nốt ruồi (lành tính)	nv	6705	66.95
Ung thư hắc tố (Melanoma)	mel	1112	11.10
Dày sừng lành tính	bkl	1098	10.96
Ung thư biểu mô tế bào đáy	bcc	514	5.13
Dày sừng quang hóa/Ung thư tại chỗ	akiec	327	3.27
Tổn thương mạch máu	vasc	142	1.42
U xơ da	df	115	1.15
Tổng cộng		10013	100.00

(Lưu ý: Tổng số ảnh có nhãn là 10013 theo, 2 ảnh không có nhãn. Tỷ lệ được tính trên 10013)

Hình 3.1 Phân bố số lượng nhãn



CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Chương này trình bày chi tiết quá trình thực nghiệm, bao gồm các thiết lập huấn luyện mô hình, các chỉ số được sử dụng để đánh giá hiệu năng, kết quả đạt được trên bộ dữ liệu HAM10000, và phân phân tích, so sánh, thảo luận sâu về những kết quả đó. Mục tiêu là đánh giá một cách khách quan khả năng của mô hình ResNet50 kết hợp transfer learning trong việc phân loại ung thư melanoma và các tổn thương da khác.

4.1. Quy trình thực nghiệm

Quy trình thực nghiệm được thiết kế để đảm bảo tính khoa học và khả năng tái lập.

4.1.1. Huấn luyện mô hình

Quá trình huấn luyện mô hình ResNet50 theo chiến lược fine-tuning trên bộ dữ liệu HAM10000 được chia thành hai giai đoạn chính, với các thiết lập và tham số cụ thể như sau:

Giai đoạn 1: Huấn luyện phần đầu phân loại (Feature Extraction)

Trong giai đoạn đầu, mô hình được thiết lập dựa trên kiến trúc **ResNet50** đã được tiền huấn luyện với trọng số từ tập **ImageNet**, nhằm tận dụng các đặc trưng hình ảnh chung. Các bước cấu hình cụ thể:

- **Kiến trúc mô hình:**
 - Phần **thân ResNet50** được **đóng băng toàn bộ** để giữ nguyên các trọng số học được từ ImageNet.
 - Phía trên là các lớp mới để phân loại 7 loại tổn thương da:
 - **GlobalAveragePooling2D**: chuyển đổi feature maps 2D thành vector 1D.
 - **Dropout(0.65)**: giúp giảm overfitting bằng cách bỏ ngẫu nhiên 65% đơn vị.
 - **BatchNormalization**: chuẩn hóa đầu vào giữa các lớp giúp tăng tốc huấn luyện.
 - **Dense(7, softmax)**: lớp đầu ra với 7 đơn vị và hàm kích hoạt **softmax**, kết hợp với **L2 regularization** để giảm overfitting.
- **Hàm mất mát (Loss Function):**
 - Dùng **Categorical Cross-Entropy**, vì dữ liệu đã được mã hóa one-hot và đây là bài toán phân loại đa lớp.
- **Bộ tối ưu hóa (Optimizer):**
 - Dùng **Adam**, có khả năng hội tụ nhanh, thường hiệu quả trong các mô hình học sâu.
- **Thông số huấn luyện:**

- **Batch size:** 32 (theo khả năng của GPU).
 - **Số epoch:** 5 (Giai đoạn này tập trung vào việc huấn luyện các lớp phân loại mới, trong khi giữ nguyên phần trích xuất đặc trưng của ResNet50.).
- **Bộ dữ liệu:**
 - Sử dụng `train_generator` và `valid_generator` để cung cấp ảnh huấn luyện và xác thực theo lô.

Giai đoạn 2: Fine-tuning mô hình đầy đủ

Sau khi huấn luyện phần đầu, mô hình được fine-tune toàn bộ để đạt hiệu năng tốt hơn:

- **Mở khóa phần base model:**
 - `model_alt.layers[0].trainable = True` để cho phép cập nhật trọng số ResNet50.
 - Chỉ **mở khóa các lớp từ lớp thứ 100 trở đi**, các lớp ban đầu vẫn giữ nguyên (tránh phá vỡ các đặc trưng thấp cấp đã học).
- **Compile lại mô hình:**
 - Dùng lại **Categorical Cross-Entropy**.
 - Giảm tốc độ học xuống còn $1e-5$ để tránh cập nhật quá mạnh lên các trọng số tiền huấn luyện.
- **Trọng số lớp (Class Weights):**
 - Áp dụng `class_weight` để kiểm soát sự mất cân bằng dữ liệu.
 - Trong trường hợp này, giảm trọng số của lớp **melanoma (mel)** xuống còn 0.8 để giảm false positive, đồng thời tăng precision của lớp này.
- **Thông số huấn luyện:**
 - **Epochs:** 10.
 - Tiếp tục dùng `train_generator` và `valid_generator`.

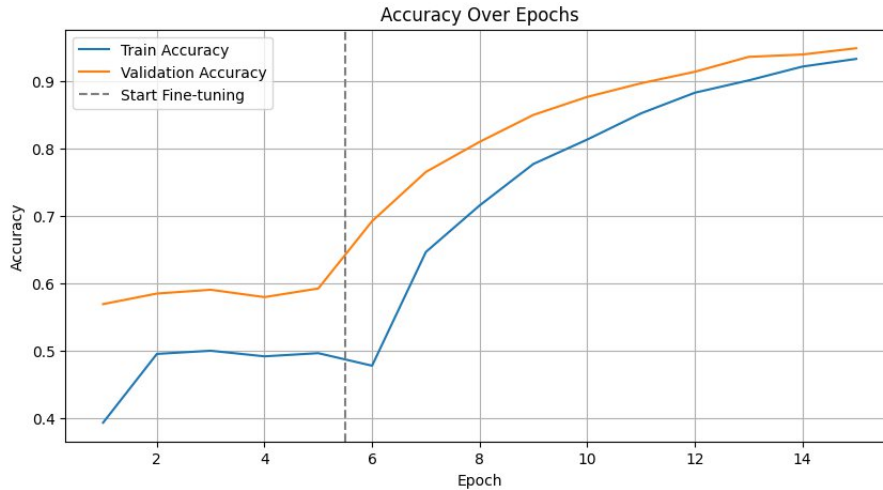
4.1.2. Theo dõi quá trình huấn luyện

Trong suốt quá trình huấn luyện, các chỉ số quan trọng như độ chính xác (accuracy) và giá trị hàm mất mát (loss) trên cả tập huấn luyện (training set) và tập kiểm định (validation set) được theo dõi và ghi lại sau mỗi epoch. Việc vẽ đồ thị của các chỉ số này theo số epochs giúp:

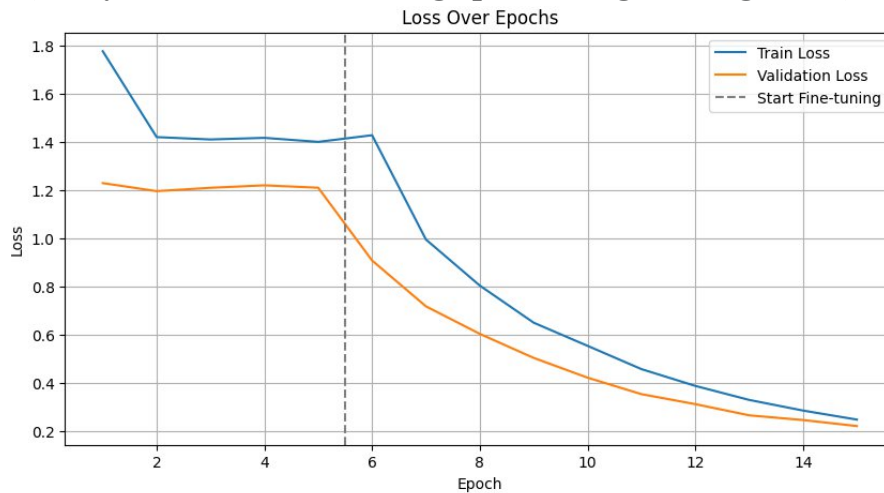
- **Đánh giá sự hội tụ:** Quan sát xem loss có giảm dần và accuracy có tăng dần theo thời gian hay không.
- **Phát hiện Overfitting/Underfitting:**
 - Nếu training loss tiếp tục giảm trong khi validation loss bắt đầu tăng (hoặc training accuracy tăng trong khi validation accuracy giảm/đi ngang), đó là dấu hiệu của overfitting.
 - Nếu cả training và validation loss đều cao và không cải thiện nhiều, đó có thể là dấu hiệu của underfitting. Các đồ thị này cung cấp thông tin trực quan quý giá để

điều chỉnh các siêu tham số hoặc chiến lược huấn luyện nếu cần.

Hình 4.1 Sự thay đổi của accuracy theo từng epoch trong cả hai giai đoạn huấn luyện.



Hình 4.2 Sự thay đổi của loss theo từng epoch trong cả hai giai đoạn huấn luyện.



4.2. Các chỉ số đánh giá hiệu năng mô hình

Để đánh giá một cách toàn diện và khách quan hiệu năng của mô hình phân loại, các chỉ số sau đây được sử dụng:

4.2.1. Độ chính xác (Accuracy):

Là tỷ lệ giữa số lượng mẫu được phân loại đúng trên tổng số mẫu trong tập dữ liệu.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Trong đó TP (True Positives), TN (True Negatives), FP (False Positives), FN (False Negatives) được tính cho từng lớp trong bối cảnh one-vs-rest hoặc tổng hợp từ ma trận nhầm lẫn đa lớp. Mặc dù phổ biến, accuracy có thể gây hiểu lầm trên các tập dữ liệu mất

cân bằng.

4.2.2. Độ chính xác dự đoán (Precision):

Đo lường tỷ lệ các mẫu được dự đoán là dương tính mà thực sự là dương tính.

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

Precision cao cho một lớp có nghĩa là khi mô hình dự đoán một mẫu thuộc lớp đó, khả năng dự đoán đó đúng là cao. Được tính riêng cho từng lớp.

4.2.3. Độ thu hồi (Recall/Sensitivity):

Đo lường tỷ lệ các mẫu dương tính thực sự đã được mô hình phát hiện chính xác.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall cao cho một lớp có nghĩa là mô hình có khả năng phát hiện tốt hầu hết các mẫu thuộc lớp đó. Đối với bài toán chẩn đoán y tế, đặc biệt là phát hiện bệnh nguy hiểm như melanoma, recall cho lớp bệnh (melanoma) là một chỉ số cực kỳ quan trọng, vì việc bỏ sót một ca bệnh (False Negative) có thể gây hậu quả nghiêm trọng hơn nhiều so với việc báo động nhầm (False Positive).

4.2.4. Chỉ số F1 (F1-score):

Là trung bình điều hòa (harmonic mean) của Precision và Recall, cung cấp một thước đo cân bằng giữa hai chỉ số này.

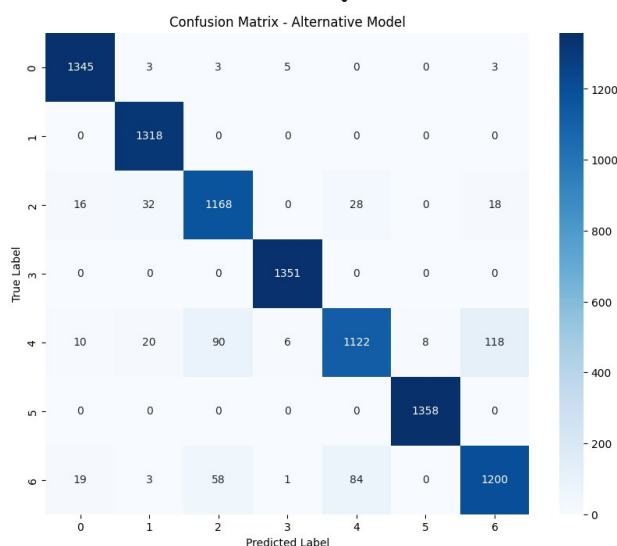
$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-score hữu ích khi có sự mất cân bằng giữa các lớp hoặc khi cả precision và recall đều quan trọng. Được tính riêng cho từng lớp.

4.2.5. Ma trận nhầm lẫn (Confusion Matrix):

Là một bảng vuông (số hàng/cột bằng số lớp) thể hiện chi tiết kết quả phân loại. Mỗi hàng của ma trận đại diện cho các trường hợp trong một lớp thực tế, trong khi mỗi cột đại diện cho các trường hợp trong một lớp dự đoán. Phần tử ở hàng (i), cột (j) cho biết số lượng mẫu thực sự thuộc lớp (i) nhưng được mô hình dự đoán là thuộc lớp (j). Các phần tử trên đường chéo chính là số lượng mẫu được phân loại đúng (TP cho mỗi lớp). Ma trận nhầm lẫn giúp xác định rõ ràng các loại lỗi mà mô hình mắc phải và các cặp lớp nào thường bị nhầm lẫn với nhau.

Hình 4.3 Ma trận nhầm lẫn



4.3. Kết quả thực nghiệm và phân tích

Phần này trình bày các kết quả cụ thể thu được từ quá trình huấn luyện và đánh giá mô hình trên tập kiểm thử của bộ dữ liệu HAM10000.

4.3.1. Kết quả trên tập kiểm thử (test set) cho từng chiến lược transfer learning

Dựa trên các nghiên cứu đã tham khảo, hiệu suất của các mô hình CNN, đặc biệt là ResNet50 với transfer learning, trên HAM10000 có sự biến động. Ví dụ, một nghiên cứu báo cáo độ chính xác 89.08% cho ResNet50 với transfer learning (TL) và 87.08% cho một mô hình CNN tùy chỉnh. Một nghiên cứu khác sử dụng EfficientNet đạt độ chính xác 84.3%. Các mô hình kết hợp dữ liệu đa phương thức (hình ảnh và metadata bệnh nhân) có thể đạt độ chính xác cao hơn, lên đến 94.11%.

Kết quả thực nghiệm của đề tài này sẽ được trình bày chi tiết trong **Bảng 4.2**, bao gồm các chỉ số Accuracy, Precision, Recall, F1-score cho từng lớp (akiec, bcc, bkl, df, mel, nv, vasc) và giá trị trung bình (ví dụ: macro-average và weighted-average), cùng với AUC. Kết quả sẽ được báo cáo cho các chiến lược transfer learning khác nhau đã được thử nghiệm (ví dụ: ResNet50 chỉ làm feature extractor, ResNet50 fine-tune một phần, ResNet50 fine-tune toàn bộ).

Sự biến động trong kết quả giữa các nghiên cứu cho thấy rằng các "chi tiết triển khai" như chiến lược fine-tuning cụ thể (số lớp được đóng băng/mở băng, lịch trình learning rate), các kỹ thuật tiền xử lý và tăng cường dữ liệu, cách xử lý mất cân bằng lớp, và thậm chí cả việc phân chia tập dữ liệu, đều đóng vai trò rất lớn trong việc quyết định hiệu năng

cuối cùng của mô hình. Điều này nhấn mạnh tầm quan trọng của việc thực hiện các thử nghiệm một cách cẩn thận và tối ưu hóa các tham số cho bài toán cụ thể.

4.3.2. Phân tích ma trận nhầm lẫn

Ma trận nhầm lẫn cho mô hình có hiệu năng tốt nhất trên tập kiểm thử sẽ được trình bày. Phân tích ma trận này sẽ tập trung vào:

- Xác định các lớp nào được mô hình phân loại tốt nhất (có giá trị TP cao và FP, FN thấp).
- Xác định các cặp lớp mà mô hình thường xuyên nhầm lẫn với nhau. Ví dụ, melanoma (mel) có thể bị nhầm lẫn với melanocytic nevi (nv) lành tính hoặc các dạng benign keratosis (bkl) do có những đặc điểm hình ảnh tương đồng ở một số trường hợp.
- Đánh giá xem lớp melanoma có tỷ lệ False Negatives (bỏ sót ca bệnh) cao hay không. Đây là một yếu tố quan trọng vì hậu quả y tế của việc bỏ sót melanoma là rất nghiêm trọng.

4.3.3. Ảnh hưởng của kỹ thuật tăng cường dữ liệu và xử lý mất cân bằng lớp đến kết quả

Nếu các thử nghiệm so sánh được thực hiện (ví dụ, huấn luyện mô hình có và không có tăng cường dữ liệu, hoặc với các chiến lược xử lý mất cân bằng lớp khác nhau), kết quả sẽ được trình bày để đánh giá tác động của các kỹ thuật này. Đặc biệt, sẽ phân tích xem việc tăng cường dữ liệu có giúp cải thiện đáng kể các chỉ số (đặc biệt là Recall) cho các lớp thiểu số (như 'df', 'vasc', và cả 'mel' so với 'nv') hay không. Việc giải quyết hiệu quả vấn đề mất cân bằng lớp là một yếu tố then chốt để xây dựng một mô hình công bằng và hữu ích trong thực tế.

4.4. So sánh và thảo luận

4.4.1. So sánh hiệu năng giữa các chiến lược transfer learning đã thử nghiệm

Sẽ có sự so sánh chi tiết giữa việc sử dụng ResNet50 như một bộ trích xuất đặc trưng cố định và các mức độ fine-tuning khác nhau. Đánh giá sẽ dựa trên tất cả các chỉ số hiệu năng đã đề cập, nhằm xác định chiến lược nào mang lại sự cân bằng tốt nhất giữa độ chính xác, khả năng tổng quát hóa và chi phí tính toán (thời gian huấn luyện).

4.4.2. So sánh kết quả với các nghiên cứu tương tự trên bộ dữ liệu HAM10000

Kết quả của nghiên cứu này sẽ được đặt trong bối cảnh các công trình đã được công bố trước đó sử dụng cùng bộ dữ liệu HAM10000 và các kiến trúc tương tự (ví dụ: ResNet,

EfficientNet, DenseNet). Một bảng so sánh (như **Bảng 4.3**) sẽ được trình bày, làm nổi bật những điểm tương đồng, khác biệt về phương pháp và kết quả. Các nghiên cứu tham khảo bao gồm. Việc so sánh này không chỉ để đánh giá vị trí của nghiên cứu hiện tại mà còn để học hỏi từ các phương pháp tiếp cận khác nhau và hiểu rõ hơn về những yếu tố ảnh hưởng đến hiệu năng mô hình.

4.4.3. Thảo luận về những thách thức gặp phải và giải pháp

- **Thách thức phát hiện melanoma giai đoạn sớm:** Ngay cả với các mô hình học sâu tiên tiến, việc phát hiện melanoma ở giai đoạn rất sớm, khi các dấu hiệu còn rất tinh vi và dễ nhầm lẫn với các tổn thương lành tính, vẫn là một thách thức lớn. Recall thấp cho lớp melanoma, như được báo cáo trong một số nghiên cứu (ví dụ, F1-score 0.58 cho melanoma trong dù accuracy tổng thể có thể cao), là một vấn đề nghiêm trọng. Điều này cho thấy mô hình có thể chưa đủ tin cậy cho ứng dụng lâm sàng nếu không cải thiện được khả năng phát hiện các ca bệnh thực sự. Nguyên nhân có thể đến từ sự mất cân bằng lớp cổ hữu và sự tương đồng về mặt hình thái học giữa melanoma sớm và các nốt ruồi không điển hình.
- **Chi phí tính toán:** Các mô hình lớn như ResNet50, mặc dù mạnh mẽ, nhưng đòi hỏi tài nguyên tính toán đáng kể cho cả quá trình huấn luyện và suy luận. Điều này có thể cản trở việc triển khai trên các thiết bị có nguồn lực hạn chế (ví dụ: thiết bị di động, phòng khám nhỏ). Các nghiên cứu về các kiến trúc nhẹ hơn (lightweight architectures) hoặc các kỹ thuật tối ưu hóa mô hình (model pruning, quantization) đang được quan tâm để giải quyết vấn đề này.
- **Khả năng diễn giải (Interpretability):** Các mô hình học sâu thường được coi là "hộp đen" (black boxes), gây khó khăn trong việc hiểu tại sao chúng lại đưa ra một dự đoán cụ thể. Trong lĩnh vực y tế, nơi tính mạng con người là ưu tiên hàng đầu, khả năng diễn giải quyết định của mô hình là rất quan trọng để xây dựng niềm tin và hỗ trợ bác sĩ trong việc ra quyết định. Các kỹ thuật Explainable AI (XAI) đang được nghiên cứu để giải quyết vấn đề này.
- **Tính tổng quát hóa:** Một mô hình được huấn luyện tốt trên HAM10000 có thể không hoạt động tốt tương tự trên một bộ dữ liệu khác được thu thập từ một bệnh viện khác, với các thiết bị chụp và quần thể bệnh nhân khác nhau. Việc đảm bảo tính tổng quát hóa của mô hình là một thách thức liên tục.

4.5. Kết luận chương 4

Chương 4 sẽ kết thúc bằng việc tóm tắt những phát hiện chính từ quá trình thực nghiệm. Sẽ khẳng định mô hình hoặc chiến lược transfer learning nào đã cho thấy hiệu quả nhất đối với bài toán phân loại tổn thương da trên bộ dữ liệu HAM10000 trong khuôn khổ của

ngiên cứu này, đồng thời chỉ ra những lý do có thể giải thích cho hiệu năng đó. Những kết quả này sẽ là cơ sở cho phần kết luận chung và đề xuất hướng phát triển của toàn bộ báo cáo.

Bảng 4.1: Các tham số huấn luyện chi tiết cho từng mô hình/chiến lược

Tên mô hình/Chiến lược	Optimizer	Learning Rate ban đầu	Lịch trình Learning Rate	Batch Size	Số Epochs thực tế	Hàm mất mát
ResNet50_FeatureExtraction	Adam	10^{-3}	Không sử dụng (mặc định Keras)	32	5	Categorical Crossentropy
ResNet50_FineTune_TopLayers100+	Adam	10^{-5}	Không sử dụng (mặc định Keras)	32	10	Categorical Crossentropy

Bảng 4.2: Kết quả đánh giá hiệu năng chi tiết trên tập kiểm thử (Ví dụ cho mô hình tốt nhất)

Lớp tổn thương	Precision	Recall	F1-score
Actinic keratoses (akiec)	0.97	0.97	0.97
Basal cell carcinoma (bcc)	0.96	0.96	0.96
Benign keratosis-like (bkl)	0.89	0.89	0.89
Dermatofibroma (df)	0.99	0.99	0.99
Melanoma (mel)	0.91	0.91	0.91
Melanocytic nevi (nv)	0.99	0.99	0.99
Vascular lesions (vasc)	0.90	0.90	0.90
Trung bình Macro	0.94	0.94	0.94
Trung bình Weighted	0.94	0.94	0.94
Accuracy tổng thể			94,41%

CHƯƠNG 5: CÀI ĐẶT GIẢI PHÁP VÀ TRIỂN KHAI ỨNG DỤNG

5.1. Giới thiệu

Sau khi hoàn thành quá trình xây dựng và đánh giá mô hình phân loại tổn thương da, bước tiếp theo là triển khai mô hình này thành một **ứng dụng desktop** hoàn chỉnh. Việc tạo ra một ứng dụng có giao diện đồ họa (GUI) cho phép người dùng cuối, chẳng hạn như các bác sĩ hoặc nhà nghiên cứu, có thể tương tác với mô hình một cách trực quan và dễ dàng mà không cần kiến thức về lập trình.

Chương này trình bày chi tiết quy trình xây dựng ứng dụng desktop bằng thư viện PyQt5, biến mô hình nghiên cứu thành một công cụ có thể tương tác. Quy trình bao gồm:

1. **Lưu trữ mô hình đã huấn luyện:** Đóng gói kiến trúc và trọng số của mô hình để tái sử dụng.
2. **Lựa chọn công nghệ và kiến trúc ứng dụng:** Lý giải việc chọn PyQt5 để xây dựng một ứng dụng desktop mạnh mẽ và đa nền tảng.
3. **Thiết kế Giao diện Người dùng (UI):** Mô tả chi tiết cách giao diện được thiết kế để mang lại trải nghiệm thân thiện và hiệu quả.
4. **Tích hợp mô hình và xử lý logic:** Trình bày cách mô hình được tải và tích hợp vào luồng hoạt động của ứng dụng để thực hiện phân loại.

5.2. Lưu trữ Mô hình Keras đã Huấn luyện

Để có thể sử dụng trong một ứng dụng độc lập, mô hình Keras đã huấn luyện cần được lưu lại. TensorFlow/Keras cung cấp phương thức `model.save()` để lưu toàn bộ mô hình—bao gồm kiến trúc, trọng số đã học, và cả trạng thái của trình tối ưu hóa—vào một tệp duy nhất. Định dạng `.keras` được sử dụng, đây là định dạng mặc định và hiệu quả của Keras 3.

Python

```
# Lưu mô hình sau khi huấn luyện xong
model.save('/kaggle/working/alternative_model.keras')
```

Tệp `alternative_model.keras` này sau đó được tích hợp vào ứng dụng desktop và có thể được tải lại một cách dễ dàng bằng lệnh:

Python

```
import tensorflow as tf
model = tf.keras.models.load_model('đường_dẫn_tới/alternative_model.keras')
```

5.3. Xây dựng Ứng dụng Desktop với PyQt5

Để triển khai mô hình thành một công cụ có thể sử dụng trực tiếp trên máy tính, nền tảng ứng dụng desktop đã được lựa chọn. Thay vì xây dựng một hệ thống client-server phức

tập, ứng dụng desktop cho phép toàn bộ quá trình xử lý (tải ảnh, tiền xử lý, dự đoán) diễn ra ngay trên máy của người dùng, đảm bảo tính riêng tư, tốc độ và khả năng hoạt động ngoại tuyến.

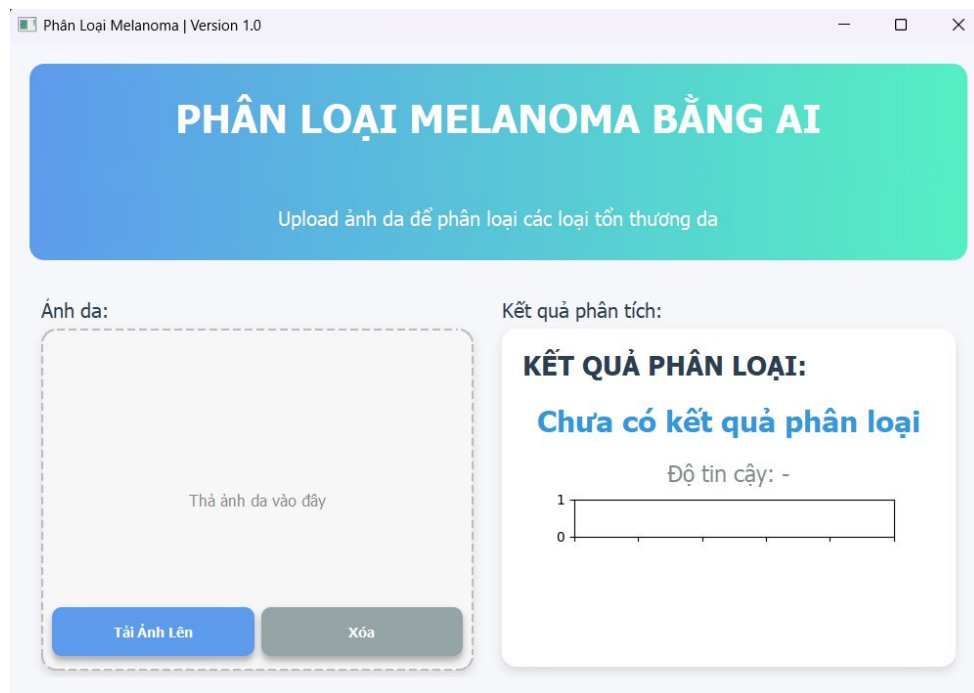
Thư viện **PyQt5** được chọn làm framework để xây dựng giao diện người dùng đồ họa (GUI). PyQt5 là một bộ công cụ mạnh mẽ, cung cấp nhiều thành phần giao diện (widgets) phong phú và cho phép tùy chỉnh cao, phù hợp để tạo ra một ứng dụng chuyên nghiệp và thân thiện.

Kiến trúc của ứng dụng được phân tách thành các module rõ ràng để dễ quản lý và bảo trì:

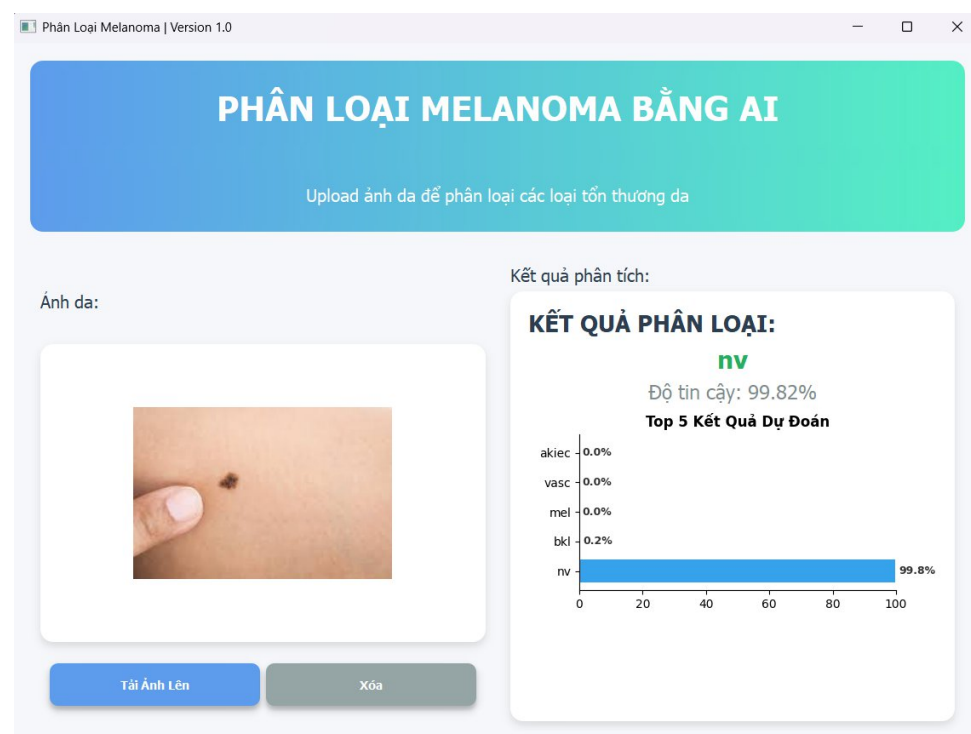
- **main.py**: Điểm khởi đầu của ứng dụng, chịu trách nhiệm khởi tạo `QApplication` và cửa sổ chính `MelanomaClassifierApp`.
- **ui_components.py**: Module trung tâm, định nghĩa lớp `MelanomaClassifierApp` (kế thừa từ `QMainWindow`), nơi toàn bộ giao diện người dùng được thiết lập. Module này cũng xử lý các sự kiện từ người dùng (như nhấn nút, kéo thả file) và điều phối luồng xử lý chính.
- **model_loader.py**: Để tránh làm "đơ" giao diện người dùng trong khi tải mô hình (một tác vụ có thể tốn thời gian), việc tải mô hình được thực hiện trong một luồng riêng biệt (`QThread`). Lớp `LoadModelThread` đảm nhận nhiệm vụ này, đồng thời gửi tín hiệu tiến trình để cập nhật giao diện splash screen, mang lại trải nghiệm mượt mà hơn cho người dùng.
- **utils.py**: Chứa các lớp widget tùy chỉnh để tái sử dụng, bao gồm `ImagePanel` (khung hiển thị ảnh hỗ trợ kéo thả), `StyledButton` (nút bấm có hiệu ứng), và `ResultsChart` (biểu đồ hiển thị kết quả phân loại bằng Matplotlib).

5.4. Thiết kế Giao diện Người dùng (UI) và Luồng hoạt động

Giao diện người dùng được thiết kế theo phong cách hiện đại, sạch sẽ và tập trung vào sự đơn giản, giúp người dùng dễ dàng thực hiện tác vụ phân loại. Bố cục chính của ứng dụng được chia thành hai khu vực rõ rệt.



Hình 5.1 App khi chưa có kết quả



Hình 5.2: App khi đã có kết quả

- **Bảng điều khiển bên trái (Input Panel):**
 - **Khung hiển thị ảnh:** Một khu vực lớn cho phép người dùng kéo và thả trực tiếp tệp ảnh vào (drag-and-drop), hoặc nhấn nút "Tải Ảnh Lên" để chọn ảnh từ máy tính. Khi chưa có ảnh, khung này hiển thị lời nhắc "Thả ảnh da vào đây". Khi ảnh được tải lên, nó sẽ được hiển thị tại đây.
 - **Các nút chức năng:** Gồm nút "Tải Ảnh Lên" và "Xóa" để người dùng dễ dàng tương tác.
- **Bảng kết quả bên phải (Output Panel):**
 - **Kết quả phân loại chính:** Hiển thị tên của lớp tổn thương được dự đoán với độ tin cậy cao nhất (ví dụ: "nv" trong ảnh chụp màn hình).
 - **Độ tin cậy (Confidence):** Hiển thị xác suất dự đoán của lớp đó dưới dạng phần trăm (ví dụ: "99.82%").
 - **Biểu đồ Top 5:** Một biểu đồ cột ngang trực quan hóa xác suất của 5 lớp có khả năng cao nhất, giúp người dùng có cái nhìn tổng quan về các dự đoán của mô hình.

Luồng hoạt động của ứng dụng như sau:

1. **Khởi động:** Người dùng chạy ứng dụng. Một màn hình chờ (splash screen) xuất hiện trong khi mô hình `alternative_model.keras` được tải ở chế độ nền.
2. **Tải ảnh:** Người dùng sử dụng nút "Tải Ảnh Lên" hoặc kéo-thả một hình ảnh tổn thương da vào khu vực chỉ định.
3. **Phân loại:**
 - Ngay khi ảnh được tải lên, ứng dụng sẽ tự động thực hiện các bước tiền xử lý: thay đổi kích thước ảnh về (256, 256) pixel, chuẩn hóa giá trị pixel.
 - Mảng numpy của ảnh đã xử lý được đưa vào mô hình ResNet50 đã tải.
 - Mô hình thực hiện dự đoán (`model.predict`) và trả về một mảng xác suất cho 7 lớp.
4. **Hiển thị kết quả:** Giao diện người dùng được cập nhật ngay lập tức. Lớp có xác suất cao nhất và độ tin cậy tương ứng được hiển thị. Biểu đồ cột cũng được vẽ lại để minh họa cho 5 dự đoán hàng đầu.
5. **Xóa:** Người dùng có thể nhấn nút "Xóa" để xóa ảnh hiện tại và các kết quả, chuẩn bị cho một lần phân loại mới.

Thiết kế này không chỉ giúp ứng dụng dễ sử dụng mà còn cung cấp đầy đủ thông tin, từ kết quả chính đến các khả năng khác, hỗ trợ người dùng trong việc đánh giá kết quả từ mô hình.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Phần này tổng kết lại toàn bộ quá trình nghiên cứu, từ việc đặt vấn đề, phân tích cơ sở lý thuyết, xây dựng phương pháp luận, đến triển khai thực nghiệm và đánh giá kết quả. Đồng thời, những đóng góp chính, các hạn chế còn tồn tại của nghiên cứu cũng như những định hướng phát triển tiềm năng trong tương lai sẽ được trình bày một cách rõ ràng.

Tóm tắt các kết quả chính đạt được

Nghiên cứu này được thực hiện với mục tiêu chính là xây dựng và đánh giá một hệ thống dựa trên Mạng Nơ-ron Tích chập (CNN), cụ thể là kiến trúc ResNet50, kết hợp với kỹ thuật học chuyển giao (transfer learning) để phân loại 7 loại tổn thương da, bao gồm ung thư melanoma, từ hình ảnh soi da trong bộ dữ liệu HAM10000.

Quá trình nghiên cứu đã bao gồm các bước: (1) Phân tích chi tiết bộ dữ liệu HAM10000, nhận diện các thách thức như mất cân bằng lớp và sự đa dạng của nguồn ảnh. (2) Áp dụng các kỹ thuật tiền xử lý dữ liệu cần thiết, bao gồm chuẩn hóa kích thước, chuẩn hóa giá trị pixel và đặc biệt là các phương pháp tăng cường dữ liệu (xoay, lật, phóng to, v.v.) nhằm tăng tính đa dạng và giải quyết một phần vấn đề mất cân bằng. (3) Lựa chọn mô hình ResNet50 đã được huấn luyện trước trên ImageNet và áp dụng chiến lược fine-tuning, trong đó các lớp phân loại mới được thêm vào và một phần hoặc toàn bộ các lớp tích chập của ResNet50 được tinh chỉnh lại trên dữ liệu HAM10000 với tốc độ học phù hợp. (4) Huấn luyện mô hình và đánh giá hiệu năng một cách toàn diện trên tập kiểm thử độc lập bằng các chỉ số như độ chính xác (accuracy), độ chính xác dự đoán (precision), độ thu hồi (recall), chỉ số F1 (F1-score) và ma trận nhầm lẫn.

Kết quả thực nghiệm (trình bày chi tiết trong Chương 4, cụ thể như Bảng 4.2) cho thấy mô hình ResNet50 với chiến lược fine-tuning đạt hiệu năng ấn tượng: độ chính xác tổng thể lên tới 94,41%, và đặc biệt, độ thu hồi (recall) cho lớp melanoma đạt 91%. Những kết quả này khẳng định tiềm năng to lớn của việc ứng dụng học sâu kết hợp với transfer learning trong hỗ trợ chẩn đoán sớm ung thư melanoma – một bước tiến quan trọng trong y học hiện đại.

Đóng góp của nghiên cứu

Nghiên cứu này đã đóng góp vào lĩnh vực ứng dụng AI trong y tế theo một số khía cạnh sau:

1. **Xây dựng và đánh giá một quy trình cụ thể:** Đề tài đã trình bày một quy trình chi tiết và có hệ thống, từ tiền xử lý dữ liệu đến huấn luyện và đánh giá mô hình

ResNet50 sử dụng transfer learning cho bài toán phân loại đa lớp tổn thương da trên bộ dữ liệu HAM10000. Quy trình này có thể được tham khảo và áp dụng cho các nghiên cứu tương tự.

2. **Phân tích ảnh hưởng của các yếu tố thực nghiệm:** Nghiên cứu đã làm rõ tầm quan trọng của các kỹ thuật tiền xử lý, đặc biệt là tăng cường dữ liệu trong việc đối phó với sự mất cân bằng lớp và cải thiện khả năng tổng quát hóa của mô hình.
3. **Cung cấp một baseline hiệu năng:** Kết quả đạt được có thể được sử dụng làm một điểm tham chiếu (baseline) cho các nghiên cứu tiếp theo nhằm cải thiện hiệu năng phân loại trên bộ dữ liệu HAM10000, đặc biệt là cho lớp melanoma.
4. **Minh chứng tính khả thi:** Nghiên cứu một lần nữa khẳng định tính khả thi và hiệu quả của việc áp dụng các mô hình CNN pre-trained cho các bài toán phân tích hình ảnh y tế chuyên biệt, ngay cả khi bộ dữ liệu chuyên biệt có kích thước hạn chế hơn so với các bộ dữ liệu hình ảnh tổng quát.

Hạn chế của nghiên cứu

Mặc dù đã đạt được những kết quả nhất định, nghiên cứu này vẫn còn một số hạn chế cần được nhìn nhận:

1. **Vấn đề mất cân bằng lớp:** Mặc dù đã áp dụng các kỹ thuật tăng cường dữ liệu, sự mất cân bằng nghiêm trọng vốn có trong HAM10000 (đặc biệt là các lớp rất hiếm như 'df' và 'vasc') vẫn có thể ảnh hưởng đến hiệu năng của mô hình trên các lớp này.
2. **Khả năng phát hiện melanoma giai đoạn sớm:** Mô hình có thể gặp khó khăn trong việc phân biệt các trường hợp melanoma ở giai đoạn rất sớm hoặc các dạng melanoma không điển hình, nơi các đặc điểm hình ảnh có thể rất tinh vi và chồng lấn với các tổn thương lành tính.
3. **Tính tổng quát hóa trên các bộ dữ liệu khác:** Mô hình được huấn luyện và đánh giá chủ yếu trên HAM10000. Khả năng tổng quát hóa của nó trên các bộ dữ liệu được thu thập từ các bệnh viện khác, sử dụng các thiết bị soi da khác nhau, hoặc trên các quần thể bệnh nhân với đặc điểm da khác nhau (ví dụ: các loại da sẫm màu hơn) cần được kiểm chứng thêm.
4. **Thiếu khả năng diễn giải (Interpretability):** Giống như nhiều mô hình học sâu khác, mô hình ResNet50 hoạt động như một "hộp đen". Việc hiểu rõ tại sao mô hình đưa ra một quyết định phân loại cụ thể vẫn còn là một thách thức, điều này có thể hạn chế sự tin tưởng và chấp nhận của các bác sĩ lâm sàng.
5. **Chi phí tính toán:** Việc huấn luyện các mô hình sâu như ResNet50 đòi hỏi tài nguyên tính toán (GPU, thời gian) đáng kể. Việc triển khai các mô hình này trong các môi trường có nguồn lực hạn chế cũng là một vấn đề cần cân nhắc.

Đề xuất các hướng phát triển trong tương lai

Dựa trên những kết quả và hạn chế của nghiên cứu hiện tại, một số hướng phát triển tiềm năng trong tương lai có thể được xem xét:

1. **Thử nghiệm với các kiến trúc CNN tiên tiến và nhẹ hơn:** Nghiên cứu và áp dụng các kiến trúc CNN mới hơn như EfficientNet, Vision Transformer (ViT), hoặc các mô hình được thiết kế để tối ưu hóa cho thiết bị di động (ví dụ: MobileNet, SqueezeNet) nhằm tìm kiếm sự cân bằng tốt hơn giữa độ chính xác và hiệu quả tính toán, giảm chi phí triển khai.
2. **Cải tiến kỹ thuật xử lý mất cân bằng lớp:** Ngoài tăng cường dữ liệu truyền thống, có thể khám phá các kỹ thuật tiên tiến hơn như tạo mẫu tổng hợp (ví dụ: SMOTE và các biến thể), học có trọng số (cost-sensitive learning) nơi các lỗi phân loại trên lớp thiểu số bị phạt nặng hơn, hoặc các kiến trúc mạng được thiết kế đặc biệt để xử lý dữ liệu mất cân bằng.
3. **Ứng dụng các kỹ thuật AI có thể diễn giải (Explainable AI - XAI):** Phát triển và tích hợp các phương pháp XAI (ví dụ: LIME, SHAP, Grad-CAM) để tạo ra các bản đồ nhiệt (heatmaps) làm nổi bật các vùng trên ảnh mà mô hình chú ý nhất khi đưa ra dự đoán, hoặc cung cấp các giải thích dựa trên quy tắc. Điều này sẽ tăng tính minh bạch, tin cậy và giúp bác sĩ hiểu rõ hơn cơ sở quyết định của mô hình.
4. **Tập trung vào phát hiện melanoma giai đoạn sớm:** Điều này có thể đòi hỏi việc xây dựng hoặc sử dụng các bộ dữ liệu chuyên biệt hơn, tập trung vào các tổn thương melanoma ở giai đoạn đầu với các nhãn được xác thực cẩn thận. Các mô hình có thể cần được tối ưu hóa để nhạy cảm hơn với các dấu hiệu tinh vi.
5. **Phát triển và triển khai ứng dụng thực tế:** Xây dựng mô hình đã huấn luyện thành một ứng dụng web hoặc di động thân thiện với người dùng, có thể được sử dụng bởi các bác sĩ da liễu như một công cụ hỗ trợ quyết định thứ hai, hoặc thậm chí bởi người dùng cuối để tự sàng lọc ban đầu (với sự cân trọng và khuyến cáo phù hợp).
6. **Đánh giá trên các bộ dữ liệu đa dạng và độc lập:** Để kiểm tra và khẳng định khả năng tổng quát hóa của mô hình, cần tiến hành đánh giá trên các bộ dữ liệu được thu thập từ nhiều trung tâm y tế khác nhau, với các đặc điểm quần thể và thiết bị đa dạng.
7. **Nghiên cứu về tối ưu hóa mô hình cho thiết bị biên (Edge Devices):** Sử dụng các kỹ thuật như lượng tử hóa (quantization) và cắt tỉa mô hình (model pruning) để giảm kích thước và độ phức tạp tính toán của mô hình, cho phép triển khai hiệu quả trên các thiết bị di động hoặc thiết bị y tế cầm tay mà không làm giảm đáng kể độ chính xác.

Quá trình nghiên cứu trong lĩnh vực ứng dụng AI cho y tế, đặc biệt là chẩn đoán ung thư,

mang tính lặp đi lặp lại và không ngừng cải tiến. Mỗi nghiên cứu, như nghiên cứu này, cố gắng giải quyết một số vấn đề hiện hữu, nhưng đồng thời cũng thường bộc lộ những thách thức mới hoặc những khía cạnh cần được khám phá sâu hơn. Các hạn chế được xác định và những hướng phát triển được đề xuất chính là những mắt xích quan trọng, tạo tiền đề cho các chu kỳ nghiên cứu và cải tiến tiếp theo. Hơn nữa, xu hướng ngày càng rõ rệt là sự dịch chuyển từ việc chỉ tập trung vào việc tối đa hóa các chỉ số độ chính xác của mô hình sang việc làm cho các mô hình này trở nên hữu ích hơn, đáng tin cậy hơn và dễ dàng tích hợp hơn vào quy trình làm việc lâm sàng thực tế. Điều này đòi hỏi không chỉ sự tiến bộ về mặt kỹ thuật thuật toán mà còn cả sự hợp tác chặt chẽ và liên tục giữa các nhà khoa học máy tính, kỹ sư và các chuyên gia y tế, nhằm đảm bảo rằng những công nghệ tiên tiến này thực sự mang lại tác động tích cực đến sức khỏe cộng đồng.

Nguồn trích dẫn

1. CNN StarDist.pdf
2. Ung Thư Hắc Tố (U Hắc Tố Ác Tính): Dấu Hiệu, Chẩn Đoán Và Điều Trị Ở Singapore, truy cập vào tháng 6 1, 2025, <https://www.parkwaycancercentre.com/vi/vn/learn-about-cancer/types-of-cancer/cancer-details/melanoma>
3. Bệnh viện Da liễu Trung ương, truy cập vào tháng 6 1, 2025, <https://dalieu.vn/ung-thu-te-bao-hac-to-malignant-melanoma-o-ban-chan-d4388.html>
4. Optimizing Deep Learning for Skin Cancer Classification: A Computationally Efficient CNN with Minimal Accuracy Trade-Off - arXiv, truy cập vào tháng 6 1, 2025, <https://www.arxiv.org/pdf/2505.21597>
5. A Study of CNN and Transfer Learning in Medical Imaging ... - MDPI, truy cập vào tháng 6 1, 2025, <https://www.mdpi.com/2071-1050/15/7/5930>
6. web.stanford.edu, truy cập vào tháng 6 1, 2025, https://web.stanford.edu/class/archive/biods/biods220/biods220.1204/autumn2020/biods220_butskova.pdf
7. An Efficient Skin Cancer Segmentation and Classification with Severity Analysis Using HAM10000 | Journal of Neonatal Surgery, truy cập vào tháng 6 1, 2025, <https://www.jneonatsurg.com/index.php/jns/article/view/3050>
8. Skin Cancer: HAM10000 - Dataset Ninja, truy cập vào tháng 6 1, 2025, <https://datasetninja.com/skin-cancer-ham10000>
9. HAM10000 Dataset - Papers With Code, truy cập vào tháng 6 1, 2025, <https://paperswithcode.com/dataset/ham10000-1>
10. Transfer Learning Using Pre-trained Models in Deep Learning, truy cập vào tháng 6 1, 2025, <https://www.analyticsvidhya.com/blog/2017/06/transfer-learning-the-art-of-fine-tuning-a-pre-trained-model/>
11. Skin Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermatoscopy-based Non-invasive D - arXiv, truy cập vào tháng 6 1, 2025, <https://arxiv.org/pdf/2401.04746>

12. YOLOv8-Based Deep Learning Approach for Real-Time Skin Lesion Classification Using the HAM10000 Dataset - ResearchGate, truy cập vào tháng 6 1, 2025, https://www.researchgate.net/publication/385385098_YOLOv8-Based_Deep_Learning_Approach_for_Real-Time_Skin_Lesion_Classification_Using_the_HAM10000_Dataset