

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



BÁO CÁO CUỐI KỲ

MÔN HỌC
XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH

ĐỀ TÀI
**ỨNG DỤNG MẠNG HỌC SÂU VÀO BÀI TOÁN PHÁT HIỆN U
HẮC TỐ ÁC TÍNH TRÊN BỘ DỮ LIỆU ISIC 2024.**

Nhóm sinh viên thực hiện

Nguyễn Dương Hải (250101015)
Cao Đức Trí (250101069)
Trịnh Tuấn Nam (250101046)

TP. Hồ Chí Minh, 11/2025
BẢNG PHÂN CÔNG, ĐÁNH GIÁ THÀNH VIÊN

Họ và tên	Phân công công việc	Hoàn thành
Trịnh Tuấn Nam	<ul style="list-style-type: none"> • Tìm kiếm, quan sát đánh giá dữ liệu • Phân chia dữ liệu • Tăng cường dữ liệu • Tiền xử lý dữ liệu 	100%
Cao Đức Trí	<ul style="list-style-type: none"> • Tìm hiểu độ đo pAUC • Mô hình EfficientNet-B3 • Phát triển mô hình phân lớp với EfficientNet-B3 version 1 	100%
Nguyễn Dương Hải	<ul style="list-style-type: none"> • Thử nghiệm các mô hình • Điều chỉnh tham số • Chuẩn bị code demo và visualize kết quả training mô hình 	100%

MỤC LỤC

PHẦN MỞ ĐẦU.....	3
CHƯƠNG 1: GIỚI THIỆU BÀI TOÁN VÀ DỮ LIỆU THỰC NGHIỆM	3
1.1. Tổng quan về bệnh ung thư hắc tố (Melanoma)	4
1.2. Mục tiêu nghiên cứu	4
1.3. Phát biểu bài toán.....	4
1.4. Đặc điểm bộ dữ liệu ISIC 2024	4
1.5. Các thách thức và Phương pháp đánh giá	5
CHƯƠNG 2: PHƯƠNG PHÁP GIẢI QUYẾT.....	5
2.1. Quy trình xử lý tổng quát.....	5
2.2. Tiền xử lý ảnh: Loại bỏ lông và tóc bằng Blackhat và Inpainting.....	5
2.3. Kiến trúc mạng Deep Learning: EfficientNet-B3.....	6
2.4. Chiến lược huấn luyện và xử lý mắt câm bằng.....	6
2.5. Hàm mắt mát và Tối ưu hóa ngưỡng	7
CHƯƠNG 3: KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ CHI TIẾT	7
3.1. Thiết lập môi trường thực nghiệm.....	7
3.2. Tiêu chí đánh giá đặc thù	8
3.3. Phân tích chi tiết quá trình tối ưu hóa qua 5 phiên bản	8
3.4. Tổng hợp ưu điểm và hạn chế của giải pháp đề xuất.....	10
CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	11
4.1. Kết luận.....	11
4.2. Bài học kinh nghiệm	12
4.3. Hướng phát triển trong tương lai	12
TÀI LIỆU THAM KHẢO	12

PHẦN MỞ ĐẦU

Trong lĩnh vực y tế, ung thư da, đặc biệt là u hắc tố (Melanoma), được đánh giá là một trong những loại ung thư nguy hiểm nhất nhưng hoàn toàn có tỷ lệ sống sót cao nếu được phát hiện và can thiệp ở giai đoạn sớm. Tuy nhiên, công tác chẩn đoán bằng mắt thường hiện nay vẫn gặp rất nhiều khó khăn do sự tương đồng lớn về đặc điểm hình thái giữa các tổn thương lành tính và ác tính.

Đồ án này được thực hiện nhằm xây dựng một hệ thống hỗ trợ chẩn đoán tự động sử dụng các kỹ thuật Xử lý ảnh và Học máy tiên tiến. Dữ liệu nghiên cứu được khai thác từ bộ dữ liệu ISIC 2024, bao gồm hơn 400.000 ảnh 3D Total Body Photography đã qua xử lý cắt (crop). Thách thức lớn nhất mà nhóm phải đối mặt là sự mất cân bằng dữ liệu cực kỳ nghiêm trọng, khi chỉ có khoảng 393 ảnh mang nhãn ác tính trong tổng số hơn 400.000 mẫu.

Để giải quyết vấn đề này, nhóm đã tập trung triển khai mô hình mạng học sâu kiến trúc EfficientNet kết hợp với các kỹ thuật xử lý dữ liệu chuyên sâu. Quy trình thực hiện bao gồm việc sử dụng phép biến đổi Blackhat để loại bỏ nhiễu lông tóc, áp dụng bộ lấy mẫu WeightedRandomSampler để cân bằng dữ liệu huấn luyện, và tối ưu hóa mô hình thông qua Test Time Augmentation (TTA) cùng tinh chỉnh ngưỡng cắt (Threshold Tuning).

Qua các thực nghiệm, báo cáo sẽ trình bày chi tiết quá trình cải thiện khả năng bắt bệnh (Recall) và đánh giá mô hình bằng độ đo pAUC, qua đó khẳng định hiệu quả của phương pháp trong việc hỗ trợ chẩn đoán lâm sàng.

CHƯƠNG 1: GIỚI THIỆU BÀI TOÁN VÀ DỮ LIỆU THỰC NGHIỆM

1.1. Tổng quan về bệnh ung thư hắc tố (Melanoma)

Ung thư da, đặc biệt là u hắc tố (Melanoma), được xem là một trong những loại ung thư nguy hiểm nhất hiện nay. Mặc dù có tính chất ác tính cao, nhưng bệnh nhân có tỷ lệ sống sót rất khả quan nếu bệnh được phát hiện sớm và can thiệp kịp thời.

Thách thức lớn nhất trong việc sàng lọc sớm nằm ở khâu chẩn đoán lâm sàng. Việc phân biệt giữa tổn thương lành tính và ác tính bằng mắt thường gặp rất nhiều khó khăn do sự tương đồng cao về đặc điểm hình thái. Điều này đòi hỏi các công cụ hỗ trợ chẩn đoán tự động để tăng độ chính xác và giảm thiểu sai sót chủ quan.

1.2. Mục tiêu nghiên cứu

Mục tiêu chính của đề tài là xây dựng một hệ thống hỗ trợ chẩn đoán tự động sử dụng kỹ thuật Xử lý ảnh và Học máy (Machine Learning). Hệ thống này đóng vai trò như một công cụ tham vấn, giúp các chuyên gia y tế nhanh chóng phân loại các vùng da bị tổn thương, từ đó tập trung nguồn lực vào các ca bệnh có nguy cơ cao.

1.3. Phát biểu bài toán

Bài toán được định nghĩa là một bài toán phân loại nhị phân dựa trên hình ảnh y tế:

- **Đầu vào (Input):** Hình ảnh 3D Total Body Photography đã được crop để tập trung vào vùng tổn thương.
- **Đầu ra (Output):** Dự đoán nhãn của tổn thương:
 - Nhãn 0 (Benign): Tổn thương lành tính.
 - Nhãn 1 (Malignant): U hắc tố ác tính (Melanoma).

1.4. Đặc điểm bộ dữ liệu ISIC 2024

Nhóm sử dụng dữ liệu từ cuộc thi ISIC 2024 – một trong những nguồn dữ liệu uy tín và lớn nhất trong lĩnh vực da liễu.

1.4.1. Quy mô và sự mất cân bằng

Bộ dữ liệu có quy mô rất lớn nhưng gặp phải hiện tượng lệch lớp (imbalance) cực kỳ nghiêm trọng:

- Tổng số mẫu: Hơn 400.000 ảnh.
- Số mẫu ác tính (Malignant): 393 ảnh.
- Số mẫu lành tính (Benign): Khoảng 400.000 ảnh.

Tỷ lệ giữa lớp ác tính và lành tính xấp xỉ 1:1000, đây là một trở ngại lớn cho việc huấn luyện mô hình, vì các mô hình deep learning thông thường có xu hướng "học vẹt" và đoán toàn bộ là lành tính để giảm thiểu hàm mất mát nhanh nhất.

1.4.2. Chiến lược phân chia dữ liệu

Để đảm bảo tính khách quan và khả năng tổng quát hóa của mô hình, dữ liệu được chia theo tỷ lệ:

- **Tập huấn luyện (70%):** Dùng để mô hình học các đặc trưng tổn thương.
- **Tập kiểm định (10%):** Dùng để tinh chỉnh các tham số và lựa chọn mô hình.
- **Tập kiểm tra (20%):** Dùng để đánh giá cuối cùng về hiệu năng thực tế.

Nhóm chú trọng việc chia tập sao cho tỷ lệ giữa hai lớp ở cả 3 tập đều đồng nhất với nhau.

1.5. Các thách thức và Phương pháp đánh giá

Do dữ liệu mất cân bằng mạnh, chỉ số độ chính xác (Accuracy) không còn phản ánh đúng chất lượng của mô hình. Thay vào đó, nhóm tập trung vào các tiêu chí:

- **Chỉ số pAUC (Partial AUC):** Tính diện tích dưới đường cong ROC với tỷ lệ dương tính giả (FPR) thấp ($FPR = 0.01$). Điều này đảm bảo mô hình hoạt động hiệu quả ngay cả khi yêu cầu về sai sót dương tính giả là cực thấp.
- **Cần kỹ thuật xử lý dữ liệu:** Áp dụng tăng cường dữ liệu (Data Augmentation) cho lớp ác tính để giúp mô hình nhận diện tốt hơn lớp dữ liệu hiếm này.

CHƯƠNG 2: PHƯƠNG PHÁP GIẢI QUYẾT

2.1. Quy trình xử lý tổng quát

Để giải quyết bài toán phát hiện tổn thương da, nhóm đã xây dựng một quy trình xử lý từ giai đoạn tiền xử lý ảnh, thiết kế kiến trúc mạng học sâu cho đến các chiến lược huấn luyện đặc thù cho dữ liệu mắt cáng bằng. Quy trình này đảm bảo mô hình không chỉ học được các đặc trưng hình thái học của tổn thương mà còn có khả năng tổng quát hóa tốt trên tập dữ liệu hiếm.

2.2. Tiền xử lý ảnh: Loại bỏ lông và tóc bằng Blackhat và Inpainting

Trong ảnh da liễu, sự xuất hiện của lông và tóc thường gây nhiễu, làm che lấp các đặc điểm quan trọng của tổn thương và ảnh hưởng đến hiệu năng của mô hình.

Nhóm đã áp dụng kết hợp phép biến đổi Blackhat và kỹ thuật nội suy (Inpainting) để giải quyết vấn đề này:

- **Bước 1:** Phép biến đổi Blackhat (Nhận diện sợi lông): Sử dụng phép đóng (Closing) trên ảnh gốc để lấp đầy các cấu trúc mảnh và tối.
 - Giá trị Blackhat được tính theo công thức: $\text{Blackhat} = \text{Closing} - \text{Ảnh_gốc}$.
 - Kết quả của bước này tạo ra một mặt nạ (mask) chứa các sợi lông đã được tách biệt hoàn toàn khỏi nền da dựa trên sự chênh lệch giá trị xám.
- **Bước 2:** Kỹ thuật Inpainting (Xử lý nội suy vùng trống): Sau khi xác định được vị trí sợi lông qua mặt nạ Blackhat, nhóm tiến hành thay thế các điểm ảnh tại vị trí đó bằng các giá trị điểm ảnh lân cận (nội suy).
 - Ví dụ kỹ thuật: Tại vị trí có sợi lông, giá trị xám từ ảnh gốc cực thấp (khoảng 9) sẽ được thay thế bằng giá trị sau phép closing (khoảng 200) để đồng nhất với vùng da xung quanh.
 - Công thức thực tế tại điểm ảnh: $\text{Blackhat}(C-I) = 200 - 9 = 191$ (giá trị này giúp định vị chính xác để bù đắp lại vùng da).
- **Hiệu quả:** Sự kết hợp này giúp trích xuất chính xác vị trí các sợi lông và "lấp đầy" chúng một cách tự nhiên. Kết quả là một vùng da sạch nhiễu, tạo điều kiện thuận lợi cho mạng CNN tập trung trích xuất các đặc trưng bệnh lý thay vì bị phân tâm bởi các cấu trúc không liên quan.

2.3. Kiến trúc mạng Deep Learning: EfficientNet-B3

Nhóm lựa chọn kiến trúc EfficientNet, cụ thể là phiên bản B3, làm mô hình cốt lõi.

- **Lý do lựa chọn:** Ảnh da liễu đòi hỏi mô hình có khả năng nhận diện tốt nhưng không quá nặng về mặt tài nguyên tính toán. EfficientNet cho độ chính xác cao trên tập ImageNet với số lượng tham số ít hơn đáng kể so với các mạng truyền thống.

- **Compound Scaling:** Kiến trúc này sử dụng cơ chế mở rộng đồng thời cả chiều sâu, chiều rộng và độ phân giải của ảnh theo một hệ số cố định, giúp tối ưu hóa hiệu năng.
- **Transfer Learning:** Nhóm tận dụng các mô hình đã được huấn luyện sẵn (pretrained) để thực hiện học chuyển đổi, điều này cực kỳ hiệu quả đối với các bài toán ảnh y tế vốn có lượng dữ liệu nhãn ác tính khan hiếm.
- **Phù hợp tài nguyên tính toán:** B3 không quá phức tạp nhưng cũng ko quá đơn giản phù hợp với tài nguyên tính toán hiện có và cho ra kết quả tốt hơn.

2.4. Chiến lược huấn luyện và xử lý mất cân bằng

Do dữ liệu ISIC 2024 có độ lệch lớp cực lớn (~1:1000), nhóm đã triển khai các kỹ thuật sau để tránh hiện tượng mô hình chỉ dự đoán lớp lành tính:

2.4.1. Tăng cường dữ liệu (Data Augmentation)

Sử dụng thư viện Albumentations để thực hiện tăng cường dữ liệu online trong quá trình huấn luyện:

- **Biến đổi hình học:** Bao gồm lật ngang (Horizontal Flip), lật dọc (Vertical Flip) và xoay kết hợp thay đổi tỉ lệ (ShiftScaleRotate).
- **Biến đổi màu sắc:** Điều chỉnh độ sáng, độ tương phản (RandomBrightnessContrast) và thay đổi hệ màu (HueSaturationValue) để mô hình thích nghi với các điều kiện ánh sáng khác nhau của ảnh chụp.

2.4.2. WeightedRandomSampler

Đây là kỹ thuật then chốt để xử lý dữ liệu thiểu số:

- Thay vì để mô hình học theo phân phối tự nhiên, WeightedRandomSampler sẽ ép mỗi batch huấn luyện (ví dụ 32 ảnh) phải duy trì tỉ lệ 50% Lành tính và 50% Ác tính.
- Việc này buộc mô hình phải quan sát và học các đặc trưng của lớp ác tính liên tục trong mọi bước cập nhật gradient, ngăn chặn việc mô hình bị "lười" và bỏ qua lớp bệnh.

2.5. Hàm mất mát và Tối ưu hóa ngưỡng

- **Hàm mất mát:** Nhóm đã thử nghiệm với CrossEntropyLoss và BinaryFocalLoss. Trong đó, Focal Loss được thiết kế để tập trung vào các mẫu khó phân loại bằng cách giảm trọng số của các mẫu dễ.
- **Threshold Tuning:** Nhóm không sử dụng ngưỡng mặc định là 0.5 mà thực hiện tìm ngưỡng tối ưu trên tập Validation (ngưỡng tối ưu tìm được là 0.43) để cân bằng giữa độ chính xác và khả năng bắt bệnh (Recall).
- **Test Time Augmentation (TTA):** Trong quá trình dự đoán (Inference), ảnh đầu vào được biến đổi nhiều lần (lật, xoay) và kết quả cuối cùng là trung bình cộng của các lần dự đoán này, giúp tăng độ ổn định và giảm nhiễu.

CHƯƠNG 3: KẾT QUẢ THỰC NGHIỆM VÀ ĐÁNH GIÁ CHI TIẾT

3.1. Thiết lập môi trường thực nghiệm

Để đảm bảo tính khách quan, tin cậy và khả năng tái lập kết quả (reproducibility), nhóm nghiên cứu đã thiết lập môi trường thực nghiệm với các thông số kỹ thuật được kiểm soát chặt chẽ:

- **Tài nguyên phần cứng:** Sử dụng GPU NVIDIA RTX 5090 32GBVRAM để tăng tốc quá trình huấn luyện mạng nơ-ron tích chập sâu.
- **Kiến trúc mô hình (Backbone):** EfficientNet-B3 được khởi tạo với bộ trọng số (weights) đã huấn luyện trước trên tập dữ liệu ImageNet (Pre-trained). Việc này giúp mô hình hội tụ nhanh hơn nhờ tận dụng các đặc trưng thị giác cơ bản đã học được từ trước.
- **Siêu tham số (Hyperparameters):**
 - Kích thước batch (Batch size): 32.
 - Hàm tối ưu (Optimizer): AdamW với tốc độ học (Learning rate) khởi tạo là $1e^{-3}$.
 - Lịch trình huấn luyện (Scheduler): CosineAnnealingLR để điều chỉnh learning rate giảm dần, giúp mô hình hội tụ về điểm cực trị địa phương tốt hơn.
 - Số lượng Epoch: 10 (kết hợp cơ chế Early Stopping để chống Overfitting).
- **Chiến lược phân chia dữ liệu (Stratified Split):** Tập dữ liệu được chia thành Train (70%) - Val (10%) - Test (20%). Quan trọng nhất, nhóm sử dụng phương pháp phân tầng để đảm bảo tỷ lệ 1:1000 giữa lớp ác tính và lành tính được giữ nguyên vẹn trong cả 3 tập, tránh hiện tượng lệch phân phối.

3.2. Tiêu chí đánh giá đặc thù

Trong bối cảnh dữ liệu mất cân bằng cực đoan (1:1000), các độ đo truyền thống như Accuracy trở nên vô nghĩa (mô hình đoán toàn bộ là lành tính vẫn đạt Accuracy 99.9%). Do đó, nhóm tập trung phân tích hai chỉ số sau:

- **pAUC (Partial Area Under the ROC Curve - FPR <= 0.01):**
 - Định nghĩa: Diện tích dưới đường cong ROC nhưng chỉ tính trong vùng Tỷ lệ dương tính giả (FPR) từ 0% đến 1%.
 - Ý nghĩa y học: Trong thực tế sàng lọc, chúng ta không thể chấp nhận việc báo động giả quá nhiều (gây quá tải cho bác sĩ). pAUC đo lường khả năng phát hiện bệnh của mô hình trong điều kiện "khắt khe" nhất: chỉ được phép báo sai dưới 1%. Đây là chỉ số quan trọng nhất của cuộc thi ISIC.
- **Recall (Sensitivity) trên lớp Malignant:**

- Định nghĩa: Tỷ lệ ca bệnh ác tính thực tế được mô hình phát hiện đúng.
- Ý nghĩa y học: Ung thư hắc tố nếu bỏ sót sẽ dẫn đến tử vong. Do đó, mục tiêu tối thượng là tối đa hóa Recall ("Thà báo nhầm còn hơn bỏ sót").

3.3. Phân tích chi tiết quá trình tối ưu hóa qua 5 phiên bản

Nhóm nghiên cứu đã thực hiện quy trình cải tiến theo mô hình xoắn ốc, đi từ mô hình cơ sở đến các kỹ thuật nâng cao. Dưới đây là phân tích sâu về từng phiên bản.

Bảng tổng hợp kết quả thực nghiệm:

	Chiến lược chính	Recall (Malignant)	pAUC (0.01)	Đánh giá tổng quan
V1	Baseline (CrossEntropy)	0.0000	0.6012	Thất bại do thiên kiến dữ liệu
V2	Weighted Random Sampler	0.2152	0.6325	Khả năng học tốt nhất
V3	Focal Loss + Bias Init	0.1266	0.5897	Suy giảm do xung đột kỹ thuật
V4	V3 + TTA (Inference)	0.1519	0.5961	Cải thiện độ ổn định
V5	V2 + TTA + Tuning	0.2658	0.6165	Giải pháp tối ưu (SOTA)

3.3.1. Phiên bản 1 (V1) - Sự thất bại của Baseline (CrossEntropy thuận túy)

- **Cơ chế:** Huấn luyện mô hình trực tiếp trên dữ liệu gốc với hàm mất mát CrossEntropyLoss.
- **Kết quả:** Recall = 0.00. Mô hình không phát hiện được bất kỳ ca ung thư nào.
- **Phân tích nguyên nhân:**
 - Đây là ví dụ điển hình của hiện tượng "Majority Class Bias" (Thiên kiến lớp đa số).
 - Hàm mất mát (Loss function) tính toán trung bình lỗi trên toàn bộ batch. Vì tỷ lệ là 1:1000, 99.9% tín hiệu Gradient (đạo hàm để cập nhật trọng số) đến từ các ảnh lành tính.
 - Các tín hiệu yếu ớt từ ảnh ác tính bị "nuốt chửng". Mô hình nhanh chóng tìm ra cách dễ nhất để giảm Loss: Dự đoán tất cả ảnh đầu vào đều là Lành tính. Tại điểm này, Loss rất thấp nhưng mô hình hoàn toàn vô dụng.

3.3.2. Phiên bản 2 (V2) - Bước ngoặt nhờ thay đổi phân phối (Sampler)

- **Cải tiến:** Áp dụng WeightedRandomSampler. Kỹ thuật này gán trọng số lấy mẫu nghịch đảo với tần suất xuất hiện của lớp (Lớp hiếm có trọng số cao).
- **Kết quả:** Hiệu năng tăng vọt. Recall đạt 0.2152 và pAUC đạt đỉnh 0.6325.
- **Biện luận khoa học:**
 - Sampler không thay đổi dữ liệu gốc, nhưng thay đổi dữ liệu mà mô hình nhìn thấy. Trong mỗi batch huấn luyện (32 ảnh), thuật toán ép buộc tỷ lệ lấy mẫu là 50% Lành - 50% Ác.
 - Điều này tạo ra một "thực tế giả lập" cân bằng cho mô hình. Gradient từ lớp ác tính trở nên mạnh mẽ và cân bằng với lớp lành tính.
 - Mô hình buộc phải học các đặc trưng (features) để phân biệt hai lớp thay vì đoán mò. Đây chứng minh là phương pháp hiệu quả nhất để xử lý dữ liệu lệch cực đoan.

3.3.3. Phiên bản 3 (V3) - Bài học từ sự phức tạp (Focal Loss & Bias Init)

- **Cải tiến:**
 - Thay CrossEntropy bằng Binary Focal Loss: Hàm loss chuyên dụng để giảm trọng số của các mẫu dễ (easy negatives), tập trung vào mẫu khó.
 - Bias Initialization: Khởi tạo lớp cuối cùng của mạng sao cho xác suất dự đoán ban đầu xấp xỉ 0.01 (tương ứng với tỷ lệ thực tế).
- **Kết quả:** Hiệu năng suy giảm nghiêm trọng. Recall tụt xuống 0.1266.
- **Phân tích nguyên nhân thất bại:**
 - Xung đột chiến lược (Conflict): Ở V2, ta đã dùng Sampler để cân bằng dữ liệu (biến bài toán thành 50/50). Khi áp dụng thêm Focal Loss (vốn thiết kế cho dữ liệu lệch), nó tiếp tục giảm trọng số của các mẫu mà nó cho là "dễ". Điều này dẫn đến việc triệt tiêu Gradient (Gradient Vanishing), khiến mô hình học rất chậm và kém hiệu quả.
 - Hậu quả của Bias Init: Việc ép xác suất khởi tạo về 1% khiến mô hình bắt đầu với trạng thái cực kỳ "thận trọng" (Conservative). Các giá trị dự đoán (logits) đều rất thấp. Khi dùng ngưỡng mặc định 0.5, hầu hết các ca bệnh (dù mô hình có hơi nghi ngờ) đều bị loại bỏ, dẫn đến Recall thấp.

3.3.4. Phiên bản 4 (V4) - Tăng độ ổn định với TTA

- **Cải tiến:** Giữ nguyên mô hình V3, nhưng áp dụng Test Time Augmentation (TTA) trong giai đoạn suy luận. Mỗi ảnh kiểm tra được tạo ra 5 biến thể (xoay, lật), dự đoán 5 lần và lấy trung bình.
- **Kết quả:** Recall tăng nhẹ từ 0.12 lên 0.15.
- **Biện luận:**
 - Mô hình Deep Learning thường có phương sai (variance) cao: một thay đổi nhỏ ở góc ảnh có thể làm thay đổi kết quả dự đoán.

- TTA hoạt động như một cơ chế "hội chẵn tập thể" (Ensembling), giúp làm mượt các nhiễu ngẫu nhiên và tăng độ tin cậy. Tuy nhiên, do nền tảng mô hình V3 (Focal Loss) vốn đã yếu, TTA không thể tạo ra đột phá lớn.

3.3.5. Phiên bản 5 (V5) - Sự kết hợp tối ưu (The Final Architecture)

- **Chiến lược:** Nhóm nhận định rằng V2 là mô hình có khả năng học tốt nhất, còn TTA là kỹ thuật suy luận tốt nhất. Do đó, V5 là sự kết hợp của:
 - **Base Model:** Quay lại sử dụng V2 (CrossEntropy + WeightedRandomSampler).
 - **Inference:** Áp dụng TTA.
 - **Post-processing (Mới):** Kỹ thuật Threshold Tuning (Dò tìm ngưỡng tối ưu).
- **Thực hiện Tuning:** Thay vì dùng ngưỡng cứng 0.5, nhóm thực hiện vét cạn (grid search) trên tập Validation để tìm ngưỡng cắt tối ưu. Kết quả tìm được ngưỡng 0.43.
- **Kết quả cuối cùng:** Recall đạt kỷ lục 0.2658 (tăng 23% so với V2 gốc). pAUC đạt 0.6165.
- **Kết luận:** V5 là phiên bản cân bằng tốt nhất, tận dụng sức mạnh học tập của Sampler và sự tinh tế trong hậu xử lý để "vớt" được tối đa số lượng bệnh nhân.

3.4. Tổng hợp ưu điểm và hạn chế của giải pháp đề xuất

3.4.1. Ưu điểm nổi bật

- **Giải quyết triệt để vấn đề mất cân bằng:** Thực nghiệm đã chứng minh việc can thiệp vào phân phối dữ liệu (Data Distribution) thông qua WeightedRandomSampler mang lại hiệu quả vượt trội so với việc thay đổi hàm Loss hay kiến trúc mạng.
- **Tính ổn định cao:** Sự kết hợp giữa Online Augmentation (trong lúc train) và TTA (trong lúc test) giúp mô hình có khả năng chống chịu tốt với các biến thể hình ảnh, giảm thiểu hiện tượng học vẹt (Overfitting).
- **Tư duy thực tế:** Việc tối ưu hóa ngưỡng (Threshold Tuning) cho thấy cách tiếp cận hướng đến ứng dụng thực tiễn: chấp nhận đánh đổi một lượng nhỏ độ chính xác để cứu sống bệnh nhân (tăng Recall).

3.4.2. Hạn chế tồn tại

- **Phụ thuộc vào ngưỡng:** Kết quả rất nhạy cảm với ngưỡng cắt. Ngưỡng 0.43 tối ưu trên tập Validation có thể cần được hiệu chỉnh lại khi triển khai trên các tập dữ liệu lâm sàng khác.
- **Giới hạn của CNN:** Mô hình vẫn dựa trên các đặc trưng cục bộ (texture, màu sắc). Các trường hợp Melanoma không sắc tố (amelanotic melanoma) hoặc có hình thái quá giống nốt ruồi lành tính vẫn là thách thức lớn đối với thuật toán hiện tại.

CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

4.1. Kết luận

Sau quá trình nghiên cứu, triển khai và đánh giá thực nghiệm nghiêm túc qua 5 phiên bản mô hình khác nhau trên bộ dữ liệu ISIC 2024, nhóm nghiên cứu rút ra các kết luận quan trọng sau:

- Hiệu quả của mô hình đề xuất:** Phiên bản V5 (Final Combo) là giải pháp tối ưu nhất với chỉ số Recall đạt 0.27 và pAUC đạt 0.62. Đây là kết quả của sự cộng hưởng giữa chiến lược lấy mẫu thông minh (Sampler), kỹ thuật suy luận đa chiều (TTA) và sự tinh chỉnh tham số hậu xử lý (Threshold Tuning).
- Vai trò của dữ liệu:** Trong các bài toán y tế với dữ liệu lệch cực đoan (1:1000), việc xử lý dữ liệu đầu vào (Tiền xử lý Blackhat, Sampling, Augmentation) đóng vai trò quyết định, quan trọng hơn việc lựa chọn các kiến trúc mạng quá phức tạp.
- Khả năng ứng dụng:** Với chỉ số pAUC ổn định ở mức cao, hệ thống có tiềm năng lớn để phát triển thành công cụ tham vấn (Second Opinion), giúp sàng lọc bước đầu và cảnh báo các ca nguy cơ cao cho bác sĩ da liễu.

4.2. Bài học kinh nghiệm

- Chiến lược Loss Function:** Không phải lúc nào các hàm Loss hiện đại (như Focal Loss) cũng tốt hơn. Khi dữ liệu quá ít (few-shot), CrossEntropy kết hợp với Sampler thường cho kết quả ổn định và hội tụ nhanh hơn.
- Tầm quan trọng của Threshold:** Ngưỡng 0.5 là một con số mặc định tùy ý. Việc tìm ngưỡng tối ưu dựa trên tập Validation là bước bắt buộc để khớp mô hình với yêu cầu thực tế (ví dụ: ưu tiên Recall).
- Augmentation là chìa khóa:** Online Augmentation và TTA là những kỹ thuật "chi phí thấp - hiệu quả cao" để cải thiện độ chính xác mà không cần thu thập thêm dữ liệu.

4.3. Hướng phát triển trong tương lai

Để nâng cao hơn nữa hiệu năng và tính ứng dụng của hệ thống, nhóm đề xuất các hướng phát triển sau:

- **Tích hợp dữ liệu đa phương thức (Multimodal Learning):** Kết hợp dữ liệu hình ảnh với thông tin lâm sàng (Metadata) như tuổi, giới tính, vị trí giải phẫu... để cung cấp thêm ngữ cảnh cho mô hình.
- **Sử dụng Generative AI:** Ứng dụng các mô hình sinh ảnh tiên tiến (như GANs hoặc Diffusion Models) để tạo ra các dữ liệu ảo thư nhân tạo chất lượng cao, giúp làm giàu tập dữ liệu huấn luyện.
- **Thử nghiệm kiến trúc mới:** Nghiên cứu áp dụng các kiến trúc Vision Transformer (ViT) hoặc Swin Transformer, vốn có khả năng nắm bắt các mối quan hệ toàn cục (global dependencies) tốt hơn so với CNN truyền thống.

TÀI LIỆU THAM KHẢO

1. International Skin Imaging Collaboration (ISIC), "ISIC 2024 Challenge Dataset", <https://challenge2024.isic-archive.com/>.
2. Mingxing Tan & Quoc V. Le (Google Research), "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", 2019.
3. Albumentations Team, "Fast and flexible image augmentations", <https://albumentations.ai/>.
4. Tài liệu môn học, "Xử lý ảnh và Thị giác máy tính", Trường Đại học Công nghệ Thông tin - ĐHQG TP.HCM.
5. Lin et al., "Focal Loss for Dense Object Detection" (Kiến trúc BinaryFocalLoss ứng dụng trong bài toán).
6. Top Hat and Black Hat Transform using Python-OpenCV, <https://www.geeksforgeeks.org/python/top-hat-and-black-hat-transform-using-python-opencv/>
7. Image Inpainting, https://docs.opencv.org/3.4/d9/d3d/tutorial_py_inpainting.html
8. EfficientNet Model <https://viblo.asia/p/efficientnet-cach-tiep-can-moi-ve-model-scaling-cho-convolutional-neural-networks-Qbc5QQzm5D8>
9. EfficientNet - một cách nghĩ mới trong scale model với accuracy cao <https://www.youtube.com/watch?v=ZLnY2oaMaBY>
10. ISIC 2024 metric Pauc <https://www.kaggle.com/code/yunsuxiaozi/isic-2024-metric-pauc>
11. Partial Area Under the ROC Curve (PAUC) <https://encyclopedia.pub/entry/30583>

