

IMAGE RECOGNITION WITH IMBALANCED DATA BASED ON DEEP LEARNING

Tran Van Thanh¹, Nguyen Van Dai², Ha Manh Toan³, Duong Thi Nhung^{4*}

¹Lac Hong University, ²University of Science - VNU

³Institute of Information Technology - VAST, ⁴Thai Nguyen University

ARTICLE INFO		ABSTRACT
Received:	19/3/2025	Skin cancer has been a serious health problem to human society in recent times, and patients will easily face dangerous situations if their diseases are not detected early. To address this issue, this research has been conducted towards an automatic classification of skin lesion images that can be captured by using a normal camera. Experiments have been conducted on the HAM10000 set, which had 7 different lesion types and a significant imbalance between classes. Accordingly, this research focuses on handling data imbalance, which helps to increase the efficiency in identifying minority classes but still needs to ensure the performance in identifying majority classes. Comprehensive and comparative experiments are also conducted with popular deep learning architectures including ConvNeXtTiny, DenseNet 201, Inception-ResNet-v2, and MobileNet-v3 Small to discuss and clarify the hypothesis. The study confirmed the superiority of the proposed method with the highest balanced accuracy value of 0.7584 and the overall accuracy value of 0.8408 for the ConvNeXtTiny model.
Revised:	09/5/2025	
Published:	10/5/2025	

KEYWORDS

Skin cancer detection
Class imbalance
HAM10000
Convolution neural network
Balanced accuracy

NHẬN DẠNG HÌNH ẢNH VỚI DỮ LIỆU MẮT CÂN BẰNG DỰA TRÊN HỌC SÂU

Trần Văn Thành¹, Nguyễn Văn Đại², Hà Mạnh Toàn³, Dương Thị Nhung^{4*}

¹Trường Đại học Lạc Hồng, ²Trường Đại học Khoa học Tự nhiên - ĐHQG Hà Nội

³Viện Công nghệ thông tin - Viện Hàn lâm Khoa học và Công nghệ Việt Nam, ⁴Đại học Thái Nguyên

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận bài:	19/3/2025	Ung thư da là vấn đề sức khỏe nghiêm trọng đối với xã hội và người bệnh sẽ dễ dàng phải đối mặt với những tình huống nguy hiểm nếu không được phát hiện sớm. Để góp phần giải quyết, nghiên cứu này được thực hiện hướng đến việc tự động phân loại ảnh tổn thương da. Các thử nghiệm được tiến hành trên bộ HAM10000 với 7 loại tổn thương khác nhau và có sự mất cân bằng đáng kể giữa các lớp. Theo đó, nghiên cứu của chúng tôi tập trung vào việc xử lý mất cân bằng dữ liệu, giúp tăng sự hiệu quả trong việc nhận dạng các lớp thiểu số nhưng vẫn cần đảm bảo hiệu năng trên các lớp đa số. Chúng tôi cũng tiến hành thử nghiệm có tính toàn diện và so sánh với ConvNeXtTiny, DenseNet 201, Inception-ResNet-v2, và MobileNet-v3 Small để thảo luận làm rõ giả thuyết. Nghiên cứu đã khẳng định ưu thế vượt trội của phương pháp đề xuất với độ chính xác cân bằng cao nhất là 0,7584 và độ chính xác trên toàn tập 0,8408 cho mô hình ConvNeXtTiny.
Ngày hoàn thiện:	09/5/2025	
Ngày đăng:	10/5/2025	

TỪ KHÓA

Phát hiện tổn thương da
Dữ liệu mất cân bằng
HAM10000
Mạng nơ ron tích chập
Độ chính xác cân bằng

DOI: <https://doi.org/10.34238/tnu-jst.12337>

* Corresponding author. Email: dt Nhungtn@tnu.edu.vn

1. Giới thiệu

Ung thư da đang trở thành một trong những mối đe dọa nghiêm trọng đối với sức khỏe cộng đồng toàn cầu. Theo báo cáo của WHO năm 2022, mỗi năm có khoảng 1,2 triệu ca mắc mới ung thư da được ghi nhận trên toàn cầu [1]. Tại Hoa Kỳ, số ca mắc mới đã tăng đến 225% trong giai đoạn 1990-2018, với hơn 99.780 ca ung thư hắc tố được chẩn đoán trong năm 2022 và khoảng 7.650 ca tử vong [2]. Tại châu Âu, hơn 144.000 ca mắc mới được báo cáo hằng năm, với khoảng 27.000 ca tử vong [3]. Đặc biệt tại châu Á, mặc dù tỷ lệ mắc thấp hơn, xu hướng gia tăng vẫn được ghi nhận rõ rệt, với tốc độ tăng khoảng 3-8% mỗi năm tại các nước phát triển như Nhật Bản và Hàn Quốc [4]. Chi phí điều trị ung thư da, đặc biệt là ung thư hắc tố, đã vượt quá 3,3 tỷ đô la Mỹ mỗi năm tại Hoa Kỳ [5]. Điều đáng lo ngại là bệnh có khả năng di căn nhanh và tỷ lệ sống sót giảm đáng kể nếu không được phát hiện sớm, từ 98% ở giai đoạn sớm xuống còn 25% khi đã di căn [6]. Trong những năm gần đây, sự phát triển của công nghệ học sâu và hệ thống CAD/CADx đã cho thấy tiềm năng to lớn trong việc hỗ trợ chẩn đoán sớm ung thư da, với độ chính xác có thể so sánh với các chuyên gia da liễu [7]. Tuy nhiên, thách thức lớn nhất trong việc phát triển các hệ thống này là sự mất cân bằng nghiêm trọng của dữ liệu y tế, khi số lượng mẫu bệnh lý ác tính thường ít hơn nhiều so với các trường hợp lành tính [8].

Trong những năm gần đây, học sâu đã ghi dấu ấn mạnh mẽ trong lĩnh vực phân loại ảnh y tế, đặc biệt là với các bài toán liên quan đến tổn thương da. Esteva và cộng sự [9] đã tiên phong khi ứng dụng mạng nơ-ron tích chập (CNN) dựa trên kiến trúc Inception-v3 để phân loại tổn thương da, đạt hiệu suất ngang tầm với các bác sĩ da liễu chuyên môn. Nghiên cứu này được thực hiện trên một tập dữ liệu khổng lồ gồm 129.450 hình ảnh lâm sàng, mở ra tiềm năng to lớn của học sâu trong chẩn đoán y khoa. Trong khi đó, Syed và cộng sự [10] đã khai thác kỹ thuật tinh chỉnh (fine-tuning) trên mạng DenseNet với tập dữ liệu HAM10000, nâng cao đáng kể khả năng nhận diện các loại tổn thương da khác nhau. Zhang và cộng sự [11] lại tập trung vào việc cải thiện hiệu suất trên dữ liệu mất cân bằng bằng cách kết hợp EfficientNet-B4 với hàm mất mát focal loss, đạt được độ chính xác cân bằng lên đến 85,7%. Ngoài ra, Chaturvedi và cộng sự [12] đã đề xuất một phương pháp phân loại ung thư da cho 7 lớp sử dụng MobileNet, tận dụng tính nhẹ và hiệu quả của mô hình này để triển khai trên các thiết bị hạn chế tài nguyên. Tương tự, Lucius và cộng sự [13] đã phát triển các khung mạng nơ-ron sâu nhằm hỗ trợ bác sĩ tổng quát trong việc phân loại tổn thương da có sắc tố, với kết quả cải thiện độ chính xác đáng kể, đạt diện tích dưới đường cong ấn tượng trên tập dữ liệu thực tế.

Tuy nhiên, các nghiên cứu trên vẫn đối mặt với những hạn chế nhất định, đặc biệt là vấn đề mất cân bằng dữ liệu – một thách thức phổ biến trong các tập dữ liệu y tế như ISIC 2018, nơi tỷ lệ mẫu giữa các lớp có thể chênh lệch lên đến 60 lần [14]. Sự mất cân bằng này khiến mô hình dễ bị thiên lệch về các lớp chiếm ưu thế, làm giảm khả năng nhận diện các lớp thiểu số – vốn thường liên quan đến các tình trạng bệnh hiếm gặp nhưng quan trọng. Để khắc phục, nhiều giải pháp đã được đề xuất. Chẳng hạn, việc lấy mẫu lại giúp cân bằng số lượng mẫu giữa các lớp, nhưng có thể dẫn đến mất thông tin quan trọng khi giảm mẫu hoặc tạo ra dữ liệu nhân tạo không thực tế khi tăng mẫu [15]. Việc học nhạy chi phí điều chỉnh trọng số của các lớp trong hàm mất mát, nhưng hiệu quả phụ thuộc lớn vào cách lựa chọn trọng số, vốn không phải lúc nào cũng tối ưu [16]. Việc học theo nhóm kết hợp nhiều mô hình để tăng độ bền, song lại đòi hỏi tài nguyên tính toán lớn và phức tạp trong triển khai [17]. Lin và cộng sự [18] giới thiệu focal loss để tập trung vào các mẫu khó, nhưng việc lựa chọn các siêu tham số không phải lúc nào cũng dễ dàng và cách tiếp cận này đôi khi bỏ qua sự đóng góp của các mẫu dễ phân loại, dẫn đến hiệu suất không đồng đều. Wang và cộng sự [19] tích hợp tăng cường dữ liệu với DenseNet-121, đạt độ chính xác 87,3%, nhưng phương pháp này thiếu cơ chế kiểm soát tỷ lệ mẫu một cách linh hoạt, dễ gây quá khớp trên các lớp thiểu số. Trong khi đó, nghiên cứu của Chaturvedi và cộng sự [12] tuy hiệu quả với MobileNet, lại chưa giải quyết triệt để vấn đề mất cân bằng dữ liệu, còn Lucius và cộng sự [13] tập trung chủ yếu vào hỗ trợ lâm sàng mà không tối ưu hóa cho các tập dữ liệu đa dạng.

Trong bài báo này, chúng tôi tập trung vào việc phát hiện tổn thương da với trọng tâm là xử lý việc mất cân bằng dữ liệu vốn dẫn đến sự thiên lệch trong các kết quả phân loại. Dữ liệu được sử dụng là bộ HAM10000. Những đóng góp chính trong nghiên cứu này bao gồm những nội dung sau. Thứ nhất, chúng tôi áp dụng xử lý mất cân bằng dựa trên việc kết hợp các kỹ thuật về chuẩn hóa, tăng cường dữ liệu cũng như hiệu chỉnh hàm mất mát. Thứ hai, chúng tôi tiến hành thử nghiệm có tính so sánh với nhiều trường hợp khác nhau của các mô hình học sâu như ConvNeXtTiny, Inception ResNet v2, MobileNet V3 Small và Densenet 201 cho các trường hợp áp dụng và không áp dụng xử lý mất cân bằng để nội dung thử nghiệm được toàn diện. Cuối cùng, chúng tôi làm rõ hiệu quả của phương pháp xử lý mất cân bằng trên cơ sở phân tích chi tiết các trường hợp thử nghiệm với việc các độ đo tổng quát trên tập dữ liệu cũng như riêng rẽ trên từng lớp tổn thương. Bên cạnh đó, chúng tôi cũng tiến hành so sánh những kết quả đạt được với nghiên cứu khác cũng thử nghiệm trên tập dữ liệu HAM10000 và thảo luận.

Các nội dung trình bày tiếp theo như sau: mục 2 sẽ trình bày chi tiết các nội dung của nghiên cứu để làm rõ các bước thực hiện và đặc điểm của phương pháp nghiên cứu, mục 3 sẽ trình bày các kết quả cụ thể và bàn luận về các khía cạnh khác nhau. Cuối cùng là phần kết luận.

2. Phương pháp nghiên cứu

2.1. Một số kiến trúc học sâu

ConvNeXtTiny là một kiến trúc mạng được Liu và đồng nghiệp phát triển vào năm 2022 [20]. Đây là phiên bản nhỏ nhất trong họ ConvNeXt. Nghiên cứu này sử dụng cấu hình:

$$C = (96, 192, 384, 768), B = (3, 3, 9, 3) \quad (1)$$

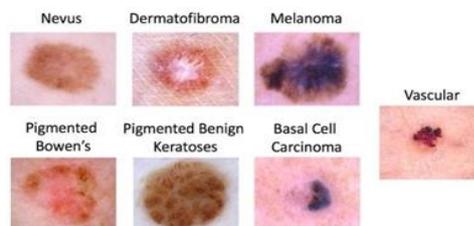
Trong đó, mạng gồm 4 giai đoạn, C là số kênh và B là số khối tại mỗi giai đoạn.

DenseNet là mô hình mạng được giới thiệu bởi Gao Huang và cộng sự [21] vào năm 2016. Mô hình này sử dụng kiến trúc đặc biệt có tên là “Dense Blocks”, trong đó các lớp trong mỗi block kết nối chặt chẽ với nhau giúp khắc phục vấn đề biến mất tín hiệu gradient và cải tiến hiệu suất huấn luyện. Nghiên cứu này sử dụng mô hình được đề xuất trong bài báo [21] là DenseNet 201 với số lớp tương ứng là 201.

InceptionResNetv2 là kiến trúc được phát triển bởi Szegedy và cộng sự [22] vào năm 2016. Kiến trúc này bao gồm các khối Inception-ResNet được cải tiến, trong đó tích hợp các kết nối tắt để cải thiện hiệu năng và tăng tốc độ hội tụ. Nghiên cứu này sử dụng mô hình được đề xuất trong [22] với cấu trúc gồm 164 lớp tích chập.

MobileNet V3 Small là một kiến trúc mạng được đề xuất vào năm 2019 bởi Howard và cộng sự [23]. So với nhiều mô hình khác, MobileNet V3Small giảm đáng kể số lượng tham số và độ phức tạp tính toán trong khi vẫn duy trì hiệu suất cao. Nghiên cứu này sử dụng mô hình được đề xuất trong [23] với cấu trúc gồm 11 khối tích chập.

2.2. Chuẩn bị dữ liệu



Hình 1. 7 lớp của tập dữ liệu

Nghiên cứu này sử dụng bộ dữ liệu HAM10000, bao gồm 10.015 ảnh soi chiếu da Dermoscopy với độ phân giải 600 x 450 pixel, được phân thành 7 loại tổn thương da khác nhau như minh họa trong Hình 1. Các loại tổn thương bao gồm: u hắc tố ác tính (melanoma - MEL), nốt ruồi (melanocytic nevus - NV), ung thư tế bào đáy (basal cell carcinoma - BCC), dày sừng

quang hóa (actinic keratosis - AKIEC), dày sừng lành tính (benign keratosis - BKL), u sợi da (dermatofibroma - DF) và tổn thương mạch máu (vascular lesions - VASC). Đặc điểm nổi bật của bộ dữ liệu này là sự phân bố không đồng đều giữa các lớp, phản ánh thực tế lâm sàng khi các trường hợp lành tính chiếm đa số trong khi các tổn thương ác tính xuất hiện với tần suất thấp hơn nhiều. Chi tiết về phân bố số lượng mẫu của từng lớp được trình bày trong Bảng 1, cho thấy rõ thách thức về tính mất cân bằng cần được giải quyết trong nghiên cứu này.

Bảng 1. Phân bố các lớp của bộ dữ liệu

Nhãn	MEL	NV	BCC	AKIEC	BKL	DF	VASC
Số lượng	1113	6705	514	327	1099	115	142

2.3. Chuẩn hóa và tăng cường dữ liệu

Để đối mặt với tình trạng mất cân bằng dữ liệu, quá trình tiền xử lý dữ liệu đóng vai trò quan trọng. Các hình ảnh được thay đổi kích thước thành 299 x 299 cho kiến trúc Inception ResNet v2[22] và 224 x 224 cho các kiến trúc khác. Các giá trị điểm ảnh cũng được chuẩn hóa về các miền cụ thể để phù hợp với thiết lập từng mô hình với sự hỗ trợ của thư viện. Điều này giúp dữ liệu đảm bảo tính nhất quán và tương thích với các trọng số đã được huấn luyện trước.

Quá trình tăng cường dữ liệu được thực hiện thông qua việc kết hợp nhiều kỹ thuật biến đổi hình ảnh khác nhau. Các biến đổi hình học cơ bản bao gồm việc thực hiện ngẫu nhiên xoay ảnh hai phía, dịch chuyển, cắt ảnh, co giãn, và lật ảnh. Ngoài ra, các biến đổi về cường độ như điều chỉnh độ sáng và thêm nhiễu Gaussian cũng được áp dụng. Đặc biệt, để tránh tạo ra các mẫu không thực tế, các thông số biến đổi được chọn lọc cẩn thận dựa trên đặc điểm của tổn thương da.

2.4. Hiệu chỉnh hàm mất mát

Phương pháp này thường được áp dụng trong các bài toán có sự mất cân bằng dữ liệu giữa các lớp, nhằm tăng cường tầm quan trọng của các lớp hiếm. Cụ thể, nó tạo ra một cơ chế phạt lớn hơn đối với các lỗi dự đoán trên các lớp có số lượng mẫu nhỏ, trong khi giảm tác động của các lớp chiếm ưu thế. Điều này đảm bảo rằng các lớp hiếm được chú ý nhiều hơn trong quá trình tối ưu hóa, đồng thời giảm thiểu nguy cơ mô hình hội tụ về một trạng thái thiên vị quá mức. Biểu thức thường được sử dụng để mô tả phương pháp này là:

$$L = \frac{1}{N} \sum_{i=1}^N w_i * l(y_i, \hat{y}_i) \quad (2)$$

Trong đó:

- w_i là trọng số của từng lớp, thường được xác định dựa trên tần suất xuất hiện của các lớp trong dữ liệu huấn luyện.

- N : số lớp.

- l : Hàm mất mát giữa nhãn thực tế và dự đoán.

- L : Tổng hàm mất mát có trọng số, trung bình trên tất cả các mẫu trong tập dữ liệu

Phương pháp này đã được chứng minh là hiệu quả trong việc giảm thiểu ảnh hưởng của sự mất cân bằng dữ liệu, đặc biệt trong các lĩnh vực như phát hiện bệnh hiếm hoặc phân loại sự kiện ít gặp [24].

2.5. Thiết lập thử nghiệm

Các thí nghiệm được thực hiện trên nền tảng Kaggle với GPU NVIDIA Tesla P100 có 16GB VRAM, được thiết kế tối ưu cho các tác vụ học sâu. Mô hình được triển khai sử dụng framework TensorFlow 2.x và Keras làm front-end API, tận dụng khả năng tính toán song song và tối ưu hóa tự động của TensorFlow. Môi trường thực thi bao gồm Python 3.7 với các thư viện hỗ trợ như NumPy cho xử lý mảng đa chiều, Pandas cho quản lý dữ liệu, và Scikit-learn cho các tác vụ tiền xử lý. Cấu hình này cho phép huấn luyện mô hình với hiệu suất cao và thời gian thực thi được tối ưu hóa, đặc biệt phù hợp cho các tác vụ học sâu quy mô lớn.

Trong bài toán phân loại, việc lựa chọn độ đo đánh giá phù hợp đóng vai trò then chốt để đánh giá chính xác hiệu suất của mô hình. Độ chính xác thông thường thường thiên vị về các lớp chiếm đa số trong tập dữ liệu và không phản ánh đúng hiệu suất của mô hình trên các tập dữ liệu mất cân bằng như HAM10000. Để khắc phục hạn chế này, chúng tôi áp dụng độ chính xác cân bằng, viết tắt là BACC. Độ đo này giúp tính toán độ chính xác cân bằng, giúp tránh việc đánh giá hiệu suất bị phóng đại trên các tập dữ liệu không cân bằng.

$$BACC = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (3)$$

Trong đó, TP là số mẫu thuộc lớp dương tính và được dự đoán đúng là dương tính, FN là số mẫu thuộc lớp dương tính nhưng bị dự đoán sai thành âm tính, TN là số mẫu thuộc lớp âm tính và được dự đoán đúng là âm tính và FP là số mẫu thuộc lớp âm tính nhưng bị dự đoán sai thành dương tính.

Để thống nhất và tiện theo dõi, trong bài báo này, ký hiệu BACC được dùng cho các kết quả ghi nhận được. Nghiên cứu còn sử dụng thêm các độ khác để đánh giá hiệu suất phân lớp trên dữ liệu mất cân bằng. Cụ thể là độ chính xác và F1-score:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$F1 - score = 2 \frac{TP}{2TP+FP+FN} \quad (5)$$

Quá trình huấn luyện mô hình được thực hiện với các tham số được tinh chỉnh cẩn thận để đạt hiệu suất tối ưu. Trong giai đoạn đầu, mô hình được huấn luyện với tốc độ học 0,01 trong 20 bước với kích thước lô 16. Sau đó, trong giai đoạn tiếp theo, tốc độ học được giảm xuống 0,001 và mô hình được huấn luyện thêm 40 bước. Chúng tôi sử dụng thuật toán tối ưu Adam [25] với epsilon là 0,1 cho tất cả 3 mô hình ở 3 thử nghiệm. Để ngăn ngừa quá khớp và tối ưu hóa quá trình huấn luyện, một số kỹ thuật đã được áp dụng: Kỹ thuật dừng sớm với đánh giá 10 để dừng huấn luyện khi mô hình không cải thiện, ModelCheckpoint để lưu trữ mô hình tốt nhất dựa trên sai số kiểm chứng, và điều chỉnh giảm tốc độ học khi sai số không cải thiện để tự động điều chỉnh tốc độ học khi mô hình đạt đến bão hòa.

3. Kết quả và bàn luận

Để đánh giá hiệu quả của chiến lược kết hợp đa kỹ thuật xử lý mất cân bằng dữ liệu, chúng tôi thực nghiệm trên bốn kiến trúc ConvNeXtTiny, DenseNet 201, Inception-ResNet-v2, và MobileNet-v3 Small. Các phương pháp được áp dụng tuần tự, và kết quả được phân tích chi tiết nhằm đánh giá phương pháp của nghiên cứu.

Trong trường hợp cơ sở, khi chưa áp dụng bất kỳ kỹ thuật cân bằng nào, các mô hình được kỳ vọng sẽ cho hiệu suất thấp trên các chỉ số cân bằng như BACC và F1-score (F1-score) của lớp thiểu số. Việc áp dụng thiết lập trọng số lớp dự kiến sẽ cải thiện đáng kể BACC mà không ảnh hưởng nhiều đến độ chính xác tổng thể. F1-score là trung bình điều hoà giữa tỉ lệ chính xác trong chẩn đoán dương tính và tỉ lệ chính xác trong các trường hợp thực sự mắc bệnh. Do đó, F1-score giúp đánh giá một cách cân bằng về hiệu năng của phương pháp.

3.1. Trường hợp cơ sở

Bảng 2 trình bày hiệu suất của các mô hình trên tập dữ liệu gốc (không áp dụng bất kỳ chiến lược cân bằng nào). Ta nhận thấy mô hình ConvNeXtTiny và DenseNet 201 có các F1-score và độ chính xác được cũng xấp xỉ nhau, và kém hơn một chút lần lượt là Inception ResNet v2 và Mobilenet v3 Small. Điều này có thể thấy mô hình nhỏ như Mobilenet v3 small không khá quát hóa tốt trong trường hợp cơ sở này. Xét trên chỉ số BACC, ta thấy các chỉ số thấp hơn cho các mô hình và đây là điều đã được dự đoán. Tuy nhiên, Inception ResNet v2 và Mobilenet v3 small có BACC khá thấp với 0,4492 và 0,3111 đã phản ánh rõ ràng sự mất cân bằng đã ảnh hưởng đáng kể đến hai mô hình này.

Bảng 2. Kết quả đánh giá các mô hình trong trường hợp cơ sở

Mô hình	F1-score	Độ chính xác	BACC
DenseNet 201	0,8450	0,8343	0,6923
Inception ResNet v2	0,734	0,7347	0,4492
Mobilenet v3 small	0,6917	0,7161	0,3111
ConvNeXtTiny	0,8418	0,8415	0,6878

Bảng 3. F1 - score mỗi lớp trong trường hợp cơ sở

Mô hình	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Mean
DenseNet 201	0,51	0,72	0,67	0,69	0,68	0,93	0,82	0,85
Inception ResNet v2	0,40	0,49	0,49	0,43	0,40	0,88	0,36	0,73
Mobilenet v3 small	0,69	0,43	0,42	0,00	0,31	0,87	0,00	0,69
ConvNeXtTiny	0,60	0,73	0,71	0,58	0,62	0,93	0,79	0,84

Bảng 3 cho thấy rõ hơn các ảnh hưởng của việc mất cân bằng. Cụ thể, lớp NV có số lượng mẫu cao nhất đã được các mô hình học thiên lệch khá rõ ràng với các chỉ số từ 0,87 đến 0,93. Trong khi đó, có những lớp có chỉ số thấp bằng 0 như DF và VASC với mô hình Mobilenet v3 small. DenseNet 201 ghi nhận BACC đạt 0,6923, với F1-score trung bình 0,845, nhưng F1-score trên lớp thiểu số như AKIEC với 0,51 và DF với 0,69 cho thấy mô hình chưa khắc phục được mất cân bằng. Inception ResNet v2 thì có đa số các lớp có F1-score dưới 0,5. Như vậy từ Bảng 3, việc so sánh các số liệu giữa các lớp cho các mô hình đã phản ánh hạn chế trong việc học đặc trưng từ các lớp thiểu số. Trong đó, Mobilenet v3 small và Inception ResNet v2 cho hiệu suất kém hơn.

Phân tích này xác nhận rằng, trong điều kiện không xử lý mất cân bằng, tất cả mô hình đều bị thiên lệch về lớp đa số là NV, dẫn đến hiệu suất kém trên các lớp thiểu số quan trọng như MEL và VASC. Điều này khẳng định nhu cầu áp dụng các kỹ thuật cân bằng dữ liệu.

3.2. Trường hợp xử lý mất cân bằng

Bảng 4. Kết quả sau khi áp dụng phương pháp xử lý mất cân bằng

Mô hình	F1-score	Độ chính xác	BACC
DenseNet 201	0,8330	0,8175	0,7237
Inception ResNet v2	0,7287	0,6991	0,6109
Mobilenet v3 small	0,6246	0,5837	0,3188
ConvNeXtTiny	0,8490	0,8408	0,7584

Bảng 4 thể hiện hiệu quả khi áp dụng phương pháp xử lý mất cân bằng và cho thấy sự cải thiện rõ rệt. ConvNeXtTiny tăng BACC từ 0,6878 lên 0,7584, tức là xấp xỉ 10,2%, với F1-score trung bình đạt 0,8490. DenseNet 201 cải thiện BACC từ 0,6923 lên 0,7237, tức là xấp xỉ 4,5%. Trong khi đó, MobileNet v3 Small chỉ tăng nhẹ BACC lên 0,3188, không phản ánh rõ ràng được sự cải thiện. Như vậy từ số liệu trong Bảng 4, việc so sánh kết quả của phương pháp đề xuất đã chỉ ra kết quả cho các mô hình đều có sự cải thiện BACC so với trường hợp cơ sở. Trong đó, mô hình ConvNeXtTiny đạt cao nhất là 0,7584 với mức tăng khoảng 10,2%.

Bảng 5. F1-score mỗi lớp khi áp dụng phương pháp xử lý mất cân bằng

Mô hình	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Mean
DenseNet 201	0,59	0,69	0,69	0,69	0,65	0,92	0,77	0,83
Inception ResNet v2	0,35	0,46	0,49	0,59	0,46	0,86	0,70	0,73
Mobilenet v3 small	0,22	0,07	0,24	0,09	0,35	0,82	0,00	0,62
ConvNeXtTiny	0,59	0,80	0,73	0,63	0,65	0,92	0,90	0,85

Bảng 5 trình bày chi tiết hơn với các chỉ số F1-score cho mỗi lớp và sự cải thiện trên các lớp ít mẫu đã thể hiện rõ ràng. ConvNeXtTiny cải thiện trên lớp VASC từ 0,79 lên 0,90. DenseNet 201

cải thiện F1-score lớp AKIEC tăng từ 0,51 lên 0,59. Bên cạnh đó, trường hợp MobileNet v3 Small không thể hiện được sự cải thiện với việc vẫn có giá trị bằng 0. Do đó, việc sử dụng mô hình cấu trúc nhỏ như MobileNet v3 Small không thực sự phù hợp với điều kiện nghiên cứu này. Kết quả của từng lớp cho các mô hình đã phản ánh rõ ràng sự cải thiện giữa các lớp thiếu số với mức độ khác nhau dựa trên các kiến trúc. ConvNeXtTiny đã cho kết quả cao nhất với đa số các độ đo đều trên 0,7.

3.3. So sánh với nghiên cứu khác

Bảng 6. So sánh với nghiên cứu [12]

Nghiên cứu	Model	Trung bình F1-score	Độ chính xác
[12]	MobileNet	0,83	0,8315
Nghiên cứu này	ConvNeXtTiny	0,8490	0,8408

Chúng tôi cũng tiến hành so sánh kết quả đạt được với nghiên cứu khác trong Bảng 6. Cụ thể, nghiên cứu [12] đã áp dụng mô hình MobileNet cho việc phân loại 7 trường hợp tổn thương da và cũng thử nghiệm trên bộ dữ liệu HAM10000. Họ đạt được kết quả với trung bình F1-score cho các lớp là 0,83 và giá trị độ chính xác là 0,8315. So với kết quả tốt nhất trong nghiên cứu này với ConvNeXtTiny, kết quả của chúng tôi cao hơn một chút. Cụ thể là 0,8408 so với 0,8315 trên độ chính xác và 0,8490 so với 0,83 trên trung bình F1-score. So sánh kết quả của các mô hình khác có số liệu thấp hơn được chúng tôi thử nghiệm, Densenet 121 có trung bình F1-score cao hơn một chút là 0,8330 nhưng có độ chính xác kém hơn với 0,8175.

4. Kết luận

Trong bài báo này, chúng tôi đã tiến hành thử nghiệm trên tập dữ liệu HAM10000 với những đặc điểm rõ ràng về sự mất cân bằng giữa các lớp tổn thương da. Nghiên cứu đã khẳng định ưu thế vượt trội của phương pháp đề xuất với độ chính xác cân bằng cao nhất là 0,7584 cho mô hình ConvNeXtTiny và thứ hai là 0,7237 cho DenseNet 201. Bên cạnh đó, độ chính xác trên toàn tập vẫn đạt được 0,8408 cho mô hình ConvNeXtTiny và 0,8175 cho DenseNet 201. Các thí nghiệm đã chỉ ra sự cải thiện rõ rệt trên các lớp thiếu số của tập dữ liệu khi áp dụng phương pháp xử lý mất cân bằng trong khi chất lượng chẩn đoán đối với các lớp chiếm đa số bị ảnh hưởng ở phạm vi tối thiểu. Điều này đã chứng minh sự khái quát hóa vượt trội của các mô hình khi áp dụng phương pháp nghiên cứu trong bài báo. Đây là cơ sở quan trọng cho những nghiên cứu tương lai với sự áp dụng các mô hình vào thực tế nơi việc thu thập dữ liệu sẽ gặp nhiều khó khăn khi triển khai tại các cơ sở với những hạn chế nhất định về con người cũng như phần cứng.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] World Health Organization, "Cancer today: Data visualization tools for exploring the global cancer burden in 2022," International Agency for Research on Cancer, 2022. [Online]. Available: <https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf>. [Accessed March. 13, 2025].
- [2] American Cancer Society, "Cancer Facts & Figures 2022," American Cancer Society, Atlanta, GA, 2022. [Online]. Available: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2022/2022-cancer-facts-and-figures.pdf>. [Accessed March. 13, 2025].
- [3] European Cancer Information System, *Estimates of cancer incidence and mortality in 2022*, European Commission, Joint Research Centre, 2022.
- [4] H. Koh *et al.*, "Changing epidemiology of skin cancer in Asia," *Journal of Dermatological Science*, vol. 94, no. 1, pp. 2-9, 2019.
- [5] G. P. Guy *et al.*, "Annual total cost of skin cancer treatment in the U.S.," *American Journal of Preventive Medicine*, vol. 48, no. 2, pp. 183-187, 2015.
- [6] A. C. Society, "Survival Rates for Melanoma Skin Cancer," *American Cancer Society Medical Content and News*, 2022.

- [7] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 168-172.
- [8] M. E. Celebi *et al.*, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362-373, 2007.
- [9] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115-118, 2017.
- [10] S. R. Hassan, S. Afroge, and M. Mizan, "Skin lesion classification using densely connected convolutional network," in *2020 IEEE Region 10 Symposium (TENSYP)*, 2020, pp. 750-753.
- [11] X. Zhang *et al.*, "Multi-class skin lesion classification using deep learning with attention mechanism," *Computers in Biology and Medicine*, vol. 129, 2021, Art. no. 104380.
- [12] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobilenet," in *International Conference on Advanced Machine Learning Technologies and Applications*. Springer, 2020, pp. 165-176.
- [13] M. Lucius, J. De All, J. A. De All, M. Belvisi, L. Radizza, M. Lanfranconi, V. Lorenzatti, and C. M. Galmarini, "Deep Neural Frameworks Improve the Accuracy of General Practitioners in the Classification of Pigmented Skin Lesions," *Diagnostics*, vol. 10, no. 11, Nov. 2020, Art. no. 969, doi: 10.3390/diagnostics10110969.
- [14] P. Tschandl *et al.*, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, 2018, Art. no. 180161.
- [15] N. V. Chawla *et al.*, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [16] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, 2009.
- [17] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221-232, 2016.
- [18] T. Y. Lin *et al.*, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 2020.
- [19] J. Wang *et al.*, "Automated skin lesion classification using deep learning for skin cancer detection," *Medical & Biological Engineering & Computing*, vol. 58, pp. 1665-1679, 2020.
- [20] Z. Liu *et al.*, "A ConvNet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11976-11986.
- [21] G. Huang *et al.*, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.
- [22] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017, pp. 4278 - 4284.
- [23] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1314-1324.
- [24] Y. Cui *et al.*, "Class-Balanced Loss Based on Effective Number of Samples," *arXiv preprint arXiv:1901.05555*, 2019, pp. 9268-9277.
- [25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014, pp. 1-13.