

Phân loại tổn thương da qua hình ảnh soi da

Trần Đại Hiến^{1,2,3}, Hồ Ngọc Mai^{1,2,3}, Lê Ngọc Thiên Phúc^{1,2,3},

Nguyễn Gia Tuấn Anh^{1,2,4}, Trần Quốc Khánh^{1,2,4}

¹ Trường Đại học Công nghệ Thông tin, Thành phố Hồ Chí Minh, Việt Nam

² Đại học Quốc gia, Thành phố Hồ Chí Minh, Việt Nam

³ {22520426, 22520839, 22521117}@gm.uit.edu.vn

⁴ {anhngt, khanhtq}@uit.edu.vn

Abstract

Phân loại tổn thương da là một nhiệm vụ quan trọng trong chẩn đoán ung thư da. Việc chẩn đoán chính xác và kịp thời các tổn thương da làm tăng đáng kể cơ hội điều trị thành công. Tuy nhiên, việc phân biệt giữa các tổn thương chỉ dựa trên kiểm tra trực quan là một thách thức đối với các bác sĩ da liễu. Nghiên cứu này đề xuất một hệ thống dựa trên mô hình học máy như Support Vector Machine, Random Forest, và LightGBM để phân loại tổn thương da qua hình ảnh soi da. Bộ dữ liệu toàn diện do nhóm xây dựng gồm 8707 hình ảnh soi da từ hai nguồn DermNet NZ và kho lưu trữ ISIC, cùng kỹ thuật tăng cường dữ liệu để giải quyết vấn đề mất cân bằng và cải thiện khả năng khái quát hóa. Mô hình được huấn luyện để phân biệt các loại tổn thương khác nhau như u da ung thư tế bào đáy, ung thư biểu mô tế bào vảy, v.v. Hiệu suất của mô hình được đánh giá trên tập dữ liệu kiểm thử độc lập. Kết quả đánh giá chứng minh rằng mô hình của nhóm đạt được độ chính xác phân loại là 56% chính xác trên tập dữ liệu kiểm thử. Nghiên cứu này góp phần phát triển công cụ chẩn đoán với khả năng hỗ trợ bác sĩ da liễu trong việc chẩn đoán ung thư da một cách chính xác hơn.

1 Giới thiệu

Phân loại tổn thương da là một lĩnh vực quan trọng trong chẩn đoán da liễu, giúp bác sĩ xác định loại tổn thương da một cách chính xác và hiệu quả. Phân loại thủ công dựa trên kinh nghiệm của bác sĩ có nhiều hạn chế như sai sót do con người, tốn thời gian, và phụ thuộc vào chuyên môn của bác sĩ. Do đó, việc phát triển hệ thống phân loại tự động các tổn thương da dựa trên hình ảnh soi da có tiềm năng to lớn trong việc hỗ trợ chẩn đoán da liễu.

Nghiên cứu này nhằm mục đích phát triển một hệ thống phân loại tự động các tổn thương da phổ

biến dựa trên hình ảnh thu thập từ máy soi da. Hệ thống sử dụng các mô hình học máy như Random Forest, Support Vector Machine, và LightGBM để phân loại hình ảnh.

Nghiên cứu này đóng góp những phần sau:

- **Thu thập và xử lý dữ liệu:** Thu thập dữ liệu từ các nguồn uy tín như DermNet NZ và kho lưu trữ ISIC, sau đó làm sạch và chuẩn bị dữ liệu để nâng cao chất lượng của bộ dữ liệu.
- **Phát triển hệ thống phân loại tự động:** Sử dụng mô hình học máy như LightGBM, Random Forest, và Support Vector Machine để phân loại tổn thương da qua hình ảnh soi da, giúp giảm sai sót, tiết kiệm thời gian, và chẩn đoán chính xác hơn.

Nghiên cứu này góp phần phát triển một hệ thống phân loại tự động các tổn thương da phổ biến dựa trên hình ảnh soi da, nâng cao hiệu quả và chất lượng dịch vụ khám chữa bệnh da liễu, hỗ trợ bác sĩ trong việc chẩn đoán và theo dõi các bệnh lý da liễu.

2 Nền tảng

2.1 Tổng quan đề tài

Nghiên cứu này tập trung vào việc phát triển hệ thống phân loại tự động các tổn thương da phổ biến dựa trên hình ảnh soi da. Bài toán nhóm cần xử lý trong nghiên cứu lần này là phân loại tổn thương da đa lớp với nhiệm vụ chính trong nghiên cứu là nhận hình ảnh soi da được thu thập từ nguồn tin đáng tin cậy làm đầu vào, sau đó phân loại và trả về nhãn phân loại tổn thương da và điểm tin cậy tương ứng. Hình ảnh được thu thập từ máy soi da, có kích thước $W \times H \times 3$ pixel, với W và H là số nguyên dương, số 3 thể hiện cho ảnh ở kênh màu RGB. Hệ thống có khả năng phân loại 7 loại tổn thương da phổ biến:

bkl, akiec, bcc, mel, df, vasc, nv. Hệ thống cũng cung cấp mức độ tin cậy trong việc phân loại để hỗ trợ bác sĩ da liễu đánh giá mức độ chính xác của kết quả.

2.2 Công trình liên quan

Phân loại tổn thương da là lĩnh vực nghiên cứu quan trọng, với tiềm năng cách mạng hóa việc chẩn đoán và điều trị tổn thương da khác nhau, bao gồm cả ung thư da. Trong nghiên cứu, nhóm khám phá các công việc liên quan đến việc phát hiện và phân loại tổn thương da, khám phá các phương pháp, cách tiếp cận và tiến bộ mà các nhà nghiên cứu đã sử dụng trong lĩnh vực này.

- **Phân loại tổn thương da bằng mô hình học sâu:** Esteva và cộng sự (Esteva et al., 2017) đã sử dụng kiến trúc Inception v3 CNN và đạt độ chính xác ấn tượng với 86.6%, chứng minh tiềm năng của công nghệ học sâu trong phân loại tổn thương da là vô cùng to lớn.
- **Kỹ thuật tăng cường dữ liệu:** Fabio Perez và cộng sự (Perez et al., 2018) đề xuất kỹ thuật tăng cường dữ liệu sử dụng các phép biến đổi hình học ngẫu nhiên để tăng kích thước và tính đa dạng dữ liệu. Điều này giúp cải thiện hiệu suất mô hình một cách đáng kể so với việc sử dụng tập dữ liệu gốc, nêu bật tầm quan trọng của việc tăng cường dữ liệu đối với việc khái quát hóa mô hình một cách đáng kể so với việc sử dụng tập dữ liệu gốc.
- **So sánh tính chính xác của các mô hình học máy truyền thống:** Grignaffini và cộng sự (Grignaffini et al., 2022) đã đánh giá các kỹ thuật học máy khác nhau để phát hiện và phân loại ung thư da từ nhiều bộ dữ liệu đồng thời cung cấp bản tóm tắt về các kỹ thuật phân loại ung thư da cùng với các chỉ số hiệu suất.
- **Phân tích tập trung các loại tổn thương cụ thể:** Zafar và cộng sự (Zafar et al., 2023) đã phân loại các tổn thương và cung cấp cái nhìn tổng quan về các kỹ thuật phân tích và chẩn đoán ung thư da từ nhiều bộ dữ liệu khác nhau.

3 Bộ dữ liệu

3.1 Nguồn dữ liệu

Để xây dựng một bộ dữ liệu toàn diện và đa dạng phục vụ cho việc huấn luyện và đánh giá mô hình học máy phân loại tổn thương da, nhóm đã thu thập ảnh soi da từ hai nguồn chính là

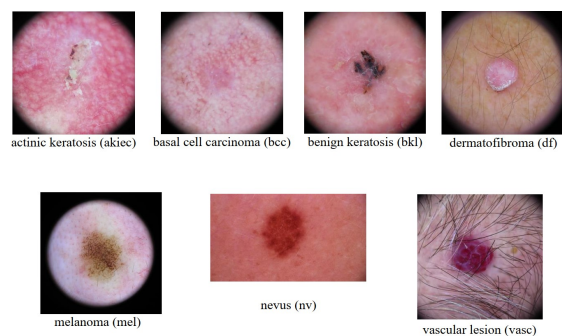
DermNet NZ(dermnetnz.org) và ISIC(www.isic-archive.com). DermNet NZ là một trang web uy tín do các bác sĩ da liễu tạo ra, duy trì quản lý và cung cấp miễn phí những thông tin và hình ảnh về các bệnh lý da khác nhau. Trang web cung cấp những thông tin y khoa được quản lý và kiểm duyệt bởi các bác sĩ da liễu chuyên khoa, đảm bảo tính chính xác và chất lượng cao. Nhóm đã sử dụng thư viện Beautiful Soup của Python để trích xuất dữ liệu và thu thập ảnh từ DermNet NZ. Nhóm thu thập được tổng cộng 107 ảnh từ DermNet NZ.

Tuy nhiên, do số lượng ảnh có sẵn trên DermNet NZ bị hạn chế nên nhóm đã quyết định thu thập thêm ảnh từ kho lưu trữ ISIC, một nguồn tài nguyên công cộng hàng đầu và uy tín trong lĩnh vực da liễu. ISIC có quy trình quản lý và kiểm soát chất lượng dữ liệu chặt chẽ, đảm bảo tính chính xác của chẩn đoán hình ảnh. Nhóm đã sử dụng công cụ isic-cli để tải xuống ảnh từ kho lưu trữ ISIC. Nhóm đã thu thập được tổng cộng 8600 ảnh từ kho lưu trữ ISIC, cung cấp bộ dữ liệu toàn diện và đa dạng cho nghiên cứu của nhóm.

Bằng cách kết hợp dữ liệu từ DermNet NZ và ISIC, nhóm đã thu thập được bộ dữ liệu hoàn chỉnh chứa tổng cộng 8707 hình ảnh soi da, bao gồm nhiều loại tổn thương da khác nhau. Bộ dữ liệu này giúp nhóm có được một bộ dữ liệu đa dạng và toàn diện hơn, phục vụ cho việc nghiên cứu và phát triển mô hình học máy.

3.2 Mô tả dữ liệu

Bộ dữ liệu có tổng cộng **8707** hình ảnh. Bộ dữ liệu bao gồm bảy loại tổn thương da, đó là **bkl**, **bcc**, **akiec**, **vasc**, **df**, **mel** và **nv**. Cụ thể như sau:



Hình 1: Hình ảnh mẫu cho 7 loại tổn thương da.

- **Benign keratosis/ Seborrheic keratosis (bkl):** là dày sừng tiết bã da dầu, là ví dụ về “chứng dày sừng lành tính” (bao gồm **1045** hình ảnh trong tập dữ liệu).

- **Actinic keratosis/ Bowen’s disease (akiec):** là các loại ung thư biểu mô tế bào vảy ở da và có thể được điều trị tại chỗ mà không cần phẫu thuật(bao gồm **1320** hình ảnh trong tập dữ liệu).
- **Basal Cell Carcinoma (bcc):** là ung thư tế bào đáy, một loại ung thư da biểu mô hiếm khi lây lan nhưng nếu không được điều trị có thể gây tử vong(bao gồm **1542** hình ảnh trong tập dữ liệu).
- **Melanoma (mel):** là ung thư hắc tố, một khối u ung thư ác tính phát triển từ tế bào hắc tố và có thể có nhiều dạng khác nhau. Nếu được phát hiện đủ sớm, nó có thể được điều trị bằng quy trình phẫu thuật cơ bản (bao gồm **1200** hình ảnh trong tập dữ liệu).
- **Dermatofibroma (df):** là u da lành tính, các tổn thương da tăng trưởng lành tính hoặc phản ứng viêm đối với chấn thương nhẹ(bao gồm **300** hình ảnh trong tập dữ liệu).
- **Vascular lesion (vasc):** là tổn thương mạch máu. U mạch anh đào, u mạch sừng và u hạt sinh mủ là những ví dụ về u mạch lành tính hoặc ác tính. (bao gồm **300** hình ảnh trong tập dữ liệu).
- **Melanocytic Nevus (nv):** là nốt ruồi hắc tố, là khối u lành tính của tế bào hắc tố và xuất hiện với nhiều hình dạng và kích cỡ khác nhau(bao gồm **3000** hình ảnh trong tập dữ liệu).

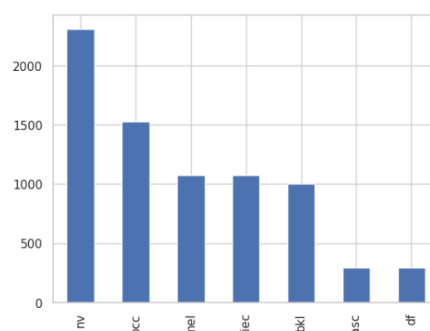
Metadata của nhóm bao gồm **8707 dòng** và **7 cột thuộc tính** như sau:

Thuộc tính	Ý nghĩa
image_id	Mã định danh duy nhất cho mỗi hình ảnh
attribution	Nguồn cung cấp hình ảnh
age_approx	Tuổi xấp xỉ của bệnh nhân
anatom_site_general	Vị trí giải phẫu của tổn thương da
dx	Chẩn đoán y tế cho tổn thương da
sex	Giới tính của bệnh nhân
lesion_id	Mã định danh duy nhất cho mỗi tổn thương da trong hình ảnh

Hình 2: Danh sách các thuộc tính và ý nghĩa của chúng

3.3 Khai phá dữ liệu

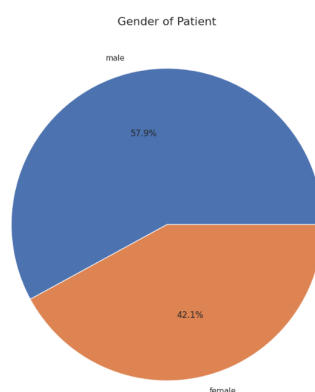
Loại tổn thương da thường gặp nhất là nv (nevus) gấp 10 lần so với hai loại tổn thương da có tỉ lệ



Hình 3: Sự phân bố của các loại tổn thương da

nhỏ nhất, bao gồm df (dermatofibroma) và vasc (Vascular lesion).

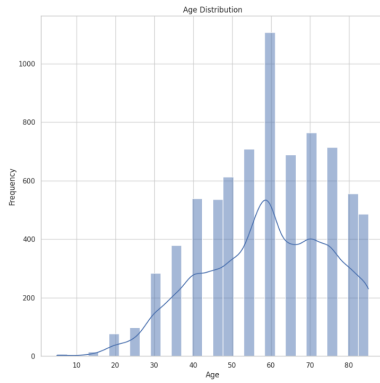
Chính vì lý do đó, từ hình trên, nhóm đưa ra nhận định rằng tập dữ liệu bị mất cân bằng, với một số lớp có nhiều hình ảnh hơn đáng kể so với các lớp khác. Trong tập dữ liệu, chúng tôi quan sát thấy lớp tổn thương da dermatofibroma (df) và vascular lesion (vasc) có kích thước nhỏ nhất và chênh lệch số lượng nhiều so với các lớp khác. Đây là một thách thức phổ biến trong các nhiệm vụ học máy và có thể ảnh hưởng đến hiệu suất của các mô hình phân loại. Chính vì vậy, nhóm đã đề xuất việc tăng cường dữ liệu để khắc phục số lượng hình ảnh hạn chế và giảm tình trạng quá khớp.



Hình 4: Sự phân bố giới tính của bệnh nhân trong nghiên cứu của nhóm.

Theo hình 4, phần lớn bệnh nhân trong nghiên cứu là nam, được biểu thị bằng miếng bánh lớn hơn, chiếm khoảng 57.9%. Thiểu số bệnh nhân trong nghiên cứu là nữ giới, được biểu thị bằng miếng bánh nhỏ hơn, chiếm khoảng 42,1%. Nhìn chung, dữ liệu cho thấy tổn thương da tập trung nhiều ở nam giới hơn.

Từ hình 5, nhóm đưa ra nhận định rằng tổn thương da phân bố ở mọi lứa tuổi. Tuy nhiên, số lượng người sau 50 tuổi tăng đáng kể. Nhóm tuổi



Hình 5: Sự phân bố độ tuổi của bệnh nhân trong tập dữ liệu phân loại tổn thương da của nhóm

có tần suất mắc bệnh cao nhất là 50-60 tuổi. Nhìn chung, dữ liệu cho thấy tổn thương da tập trung nhiều ở người trung niên hơn.

3.4 So sánh với các tập dữ liệu khác

Trong phần này, nhóm sẽ so sánh tập dữ liệu đã xây dựng với các bộ dữ liệu có kích thước từ vừa đến nhỏ (từ 10000 ảnh trở xuống). Xem xét ưu và nhược của từng bộ. Các bộ dữ liệu được so sánh: HAM10000(Tschandl et al., 2018), ISIC 2017(Cassidy et al., 2021), PH2(Cassidy et al., 2021).

	Akiec	Bcc	Bkl	Df	Mel	Nv	Vasc	Tổng
HAM10000	327	514	1099	115	1113	6705	142	10015
ISIC2017			254		374	1372		2000
PH2					40	160		200
Dataset nhóm	1320	1542	1045	300	1200	3000	300	8707

Hình 6: So sánh dataset

Đối với tập dữ liệu ISIC 2017 và PH2, dữ liệu các thuộc tính đều thiếu so với dataset của nhóm đã xây dựng. Đối với dataset HAM10000, tuy có nhiều dữ liệu ảnh hơn tập dữ liệu của nhóm, nhưng các dữ liệu ở loại da chiếm đa số như nevus quá nhiều, trong khi dữ liệu ở các loại da chiếm thiểu số như dermatofibroma(df) và vascular lesion (vasc) chỉ chiếm 115 và 142. Bộ dữ liệu của nhóm sẽ cân bằng lại sự mất cân bằng dữ liệu trầm trọng này trong khi đảm bảo độ đa dạng của các loại da. Vì vậy, tập dữ liệu của nhóm xây dựng sẽ phù hợp hơn các dataset so sánh với kích thước từ vừa đến nhỏ.

4 Tiền xử lý dữ liệu

4.1 Loại bỏ tóc

Trong một số hình ảnh về tổn thương da, đôi khi những yếu tố không mong muốn sẽ xuất hiện như

thước đo milimet hay tóc trong hình ảnh. Điều này có thể dẫn đến việc khó khăn trong phân loại và có thể sai lệch. Vì vậy nhóm đề xuất kỹ thuật xóa tóc bằng kỹ thuật thị giác máy tính sử dụng phương pháp DullRazor (Lee et al., 1997). Các bước bao gồm:

Chuyển đổi ảnh gốc sang ảnh xám (grayscale): Việc chuyển đổi ảnh màu sang ảnh xám giúp đơn giản hóa quá trình xử lý ảnh. Thay vì xử lý ba kênh màu (RGB), bây giờ ta chỉ cần xử lý một kênh duy nhất.

Phương pháp biến đổi blackhat: Phương pháp biến đổi blackhat là một kỹ thuật hình thái học giúp làm nổi bật các đặc điểm tối trên nền sáng. Bằng cách tính toán sự khác biệt giữa ảnh gốc và ảnh đã được xử lý bằng phương pháp morphology closing, các vùng tóc (thường là các đặc điểm tối nhỏ) sẽ được làm nổi bật so với phần còn lại của ảnh (H Khan et al., 2021).

Tạo mặt nạ bằng binary threshold: Ngưỡng nhị phân giúp tạo ra một ảnh nhị phân (binary image) từ ảnh đã qua xử lý, trong đó các pixel thuộc vùng tóc được gán giá trị 1 (trắng) và các pixel khác được gán giá trị 0 (đen).

Loại bỏ tóc: Sử dụng phương pháp inpaint, các pixel thuộc vùng tóc (xác định bởi mặt nạ) trong ảnh gốc được thay thế bằng các pixel xung quanh. Điều này giúp việc xóa tóc khỏi ảnh trông tự nhiên và việc khôi phục lại các vùng da bị che khuất bởi tóc mà không làm mất thông tin quan trọng (Bertalmio et al., 2000).



Hình 7: Quá trình loại bỏ tóc.

4.2 Data Augmentation

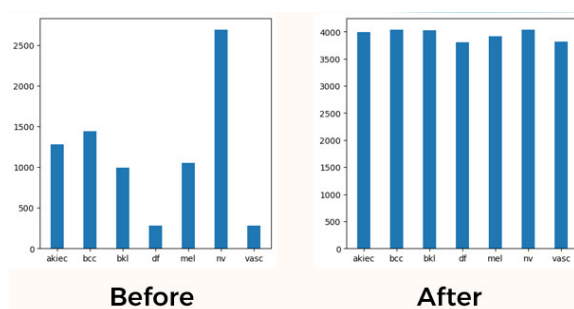
Data Augmentation là một kỹ thuật quan trọng trong học máy và thị giác máy tính, giúp tăng cường kích thước và tính đa dạng của tập dữ liệu huấn luyện, từ đó giúp mô hình học máy tổng quát hóa tốt hơn và tránh overfitting. Điều này đặc biệt quan trọng cho các bài toán như phân loại ảnh tổn thương da, nơi mà dữ liệu có thể bị hạn chế và thiếu đa dạng (Perez et al., 2018).

Trong nghiên cứu này, nhóm đã áp dụng ngẫu nhiên lần lượt các kỹ thuật Data Augmentation sau:

- Xoay ảnh ngẫu nhiên trong khoảng 360°.

- Dịch chuyển ảnh theo chiều ngang và chiều dọc ngẫu nhiên với tỷ lệ tối đa 10%.
- Phóng to hoặc thu nhỏ ảnh ngẫu nhiên trong khoảng 10%.
- Ngẫu nhiên lật ngang, lật dọc ảnh.
- Điền các pixel bị mất bằng các pixel gần nhất.

Việc áp dụng các kỹ thuật này đã giúp tạo ra khoảng 4000 ảnh cho mỗi loại tổn thương da, từ đó tạo ra một tập dữ liệu huấn luyện phong phú và đa dạng, góp phần cải thiện hiệu suất của mô hình.



Hình 8: Phân bố ảnh trước và sau khi data augmentation.

4.3 Feature Extraction

Trong bài toán phân loại ảnh tổn thương da, một bước quan trọng là trích xuất đặc trưng (feature extraction) từ ảnh để cung cấp cho mô hình học máy các thông tin cần thiết cho việc học và phân loại. Để làm điều này, nhóm đã sử dụng mô hình ResNet-50 đã được huấn luyện trước (pre-trained) từ thư viện img2vec. Dưới đây là mô tả chi tiết về quy trình và ý nghĩa của việc sử dụng ResNet-50 để trích xuất đặc trưng.

Resnet-50: ResNet-50 là một mô hình mạng nơ-ron tích chập (CNN) với 50 lớp, sử dụng các khối residual để giải quyết vấn đề vanishing gradient trong mạng nơ-ron sâu. Mỗi khối residual bao gồm hai lớp tích chập được nối với nhau bằng một kết nối cộng. Kiến trúc này giúp ResNet-50 học sâu hiệu quả và trích xuất các đặc trưng phức tạp từ ảnh (He et al., 2015).

Thư viện img2vec: Thư viện img2vec cung cấp một giao diện đơn giản để sử dụng ResNet-50 cho việc trích xuất đặc trưng từ ảnh. Các đặc trưng được trích xuất từ ResNet-50 sau đó được chuyển đổi thành vector 512 chiều và được sử dụng làm đầu vào cho mô hình học máy phân loại tổn thương da (Pan et al., 2023).

5 Mô hình máy học

5.1 Support Vector Machine

Support Vector Machine – SVM (Máy học vectơ hỗ trợ) là một thuật toán học máy có giám sát được sử dụng để phân loại dữ liệu bằng cách tìm một đường thẳng hoặc siêu phẳng tối ưu giúp khoảng cách giữa các lớp trong không gian N chiều được tối đa hóa. SVM được phát triển vào những năm 1990 bởi Vladimir N. Vapnik và các đồng nghiệp của ông. (Jakkula, 2006)

5.2 Random Forest

Random Forest (Rừng ngẫu nhiên) là một thuật toán học máy được sử dụng phổ biến, được Leo Breiman và Adele Cutler đăng ký bản quyền. Thuật toán này kết hợp đầu ra của nhiều cây quyết định để đưa ra một kết quả duy nhất. Tính dễ sử dụng và linh hoạt của nó chính là lý do nó được ưa chuộng, vì nó xử lý được cả các bài toán phân loại và hồi quy. (Biau and Scornet, 2016)

5.3 Light Gradient-Boost Machine

Light Gradient Boosting Machine (Máy học tăng cường độ dốc nhanh) là một cấu trúc gradient boosting sử dụng các thuật toán dựa trên cây. Nó được thiết kế để phân tán và mang lại hiệu quả với những ưu điểm như: tốc độ huấn luyện nhanh hơn và hiệu quả cao hơn; sử dụng bộ nhớ thấp hơn; độ chính xác tốt hơn; hỗ trợ học song song, phân tán và trên GPU; khả năng xử lý dữ liệu lớn. (Ke et al., 2017)

6 Thực nghiệm và kết quả

6.1 Kết quả

Trong phần này, kết quả của các mô hình sẽ được huấn luyện với cấu hình: AMD Ryzen 7 5800H, Radeon Graphics, 16.0 GB RAM, 64-bit.

Thông thường accuracy score là độ đo quen thuộc thường được sử dụng, tuy nhiên với bài toán phân loại đa lớp accuracy score không phải là một độ đo phù hợp bởi vì test set bị mất cân bằng (Valverde-Albacete and Peláez-Moreno, 2014) với nhãn chủ yếu là nevus(nv). Vì vậy, ba độ đo được nhóm sử dụng để đánh giá hiệu suất của mô hình phân loại là precision (độ chính xác), recall (độ phủ), F1-score (điểm số F1). Recall thể hiện khả năng phát hiện tất cả các positive, tỷ lệ này càng cao thì cho thấy khả năng bỏ sót các điểm positive là thấp. Precision thể hiện sự chuẩn xác của việc phát hiện các điểm positive. Số này càng cao thì

model nhận các điểm positive càng chuẩn. F1-score là số dung hòa Recall và Precision giúp ta có căn cứ để lựa chọn model. Các độ đo đều dùng phương pháp macro để tính toán.

Sau khi áp dụng các mô hình máy học ở Phần 5 với những đặc trưng đã lựa chọn được ở Phần 4, ta thu được kết quả sau:

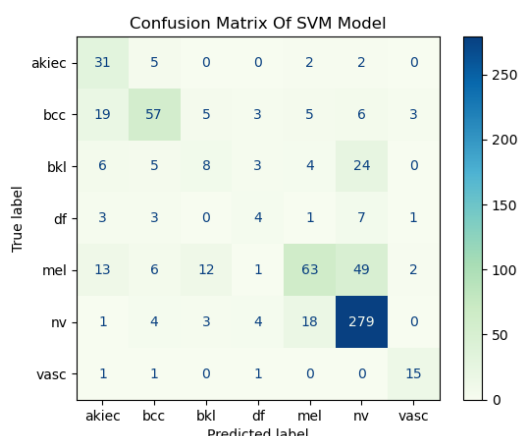
	Recall	Precision	F1-score
SVM	56%	54%	53%
Random Forest	46%	51%	46%
Ligth-GBM	54%	55%	53%

Hình 9: Kết quả thang đo hiệu suất các mô hình học máy

Có thể thấy rằng, kết quả hiệu suất của các mô hình đều không được cao, nhưng đa số đều trên 50%. Tuy không vượt trội hơn hẳn nhưng hai mô hình SVM và Light-GBM có phần thể hiện tốt hơn Random Forest.

6.2 Thực nghiệm và phân tích lỗi

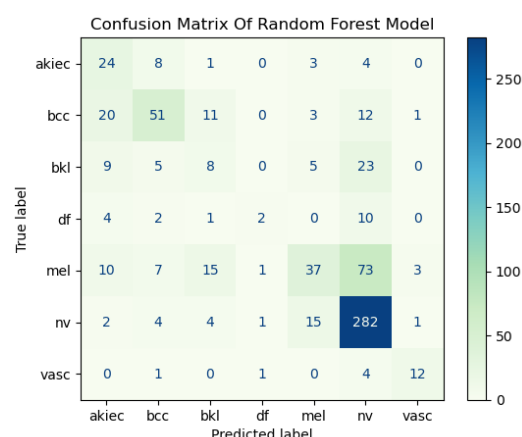
Để minh họa cho hiệu suất của mô hình phân lớp, nhóm đã sử dụng confusion matrix (ma trận nhầm lẫn) để trực quan hóa sự mơ hồ giữa nhãn thực tế và nhãn dự đoán.



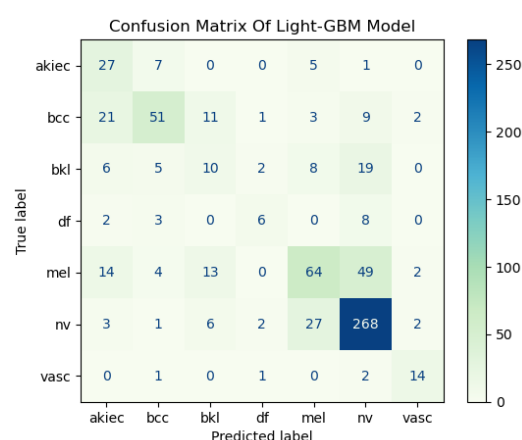
Hình 10: Ma trận nhầm lẫn của mô hình SVM.

Thông qua ma trận nhầm lẫn, có thể thấy:

- **Lớp akiec:** Độ chính xác thấp, độ nhạy cao nhưng độ đặc hiệu thấp. Mô hình có xu hướng



Hình 11: Ma trận nhầm lẫn của mô hình Random Forest.



Hình 12: Ma trận nhầm lẫn của mô hình Light-GBM.

dự đoán chính xác các trường hợp thuộc lớp akiec nhưng cũng hay nhầm lẫn với các lớp khác.

- **Lớp bcc:** Độ chính xác trung bình, độ nhạy thấp, độ đặc hiệu cao. Mô hình bỏ sót nhiều trường hợp thuộc lớp bcc nhưng ít dự đoán sai với các lớp khác.
- **Lớp bkl:** Độ chính xác thấp, độ nhạy thấp, độ đặc hiệu thấp. Mô hình gặp khó khăn trong việc phân biệt lớp bkl với các lớp khác.
- **Lớp df:** Độ chính xác trung bình, độ nhạy thấp, độ đặc hiệu cao. Mô hình bỏ sót nhiều trường hợp thuộc lớp df nhưng ít dự đoán sai với các lớp khác.
- **Lớp mel:** Độ chính xác trung bình, độ nhạy thấp, độ đặc hiệu cao. Mô hình bỏ sót nhiều trường hợp thuộc lớp mel nhưng ít dự đoán sai với các lớp khác.

- **Lớp nv:** Độ chính xác cao, độ nhạy cao, độ đặc hiệu cao. Đây là lớp có hiệu quả tốt nhất trong mô hình.
- **Lớp vasc:** Độ chính xác cao, độ nhạy cao, độ đặc hiệu cao. Đây là lớp có hiệu quả tốt thứ hai trong mô hình.

Nguyên nhân có thể là do bộ dữ liệu chưa đủ đa dạng, một số lớp có số lượng dữ liệu ít hơn, dẫn đến mô hình học tập không đầy đủ các đặc trưng của các lớp đó. Ngoài ra trong phân loại ảnh da liễu, có thể do sự kết hợp của nhiều yếu tố như sự phức tạp của ảnh y khoa, khả năng phân biệt đặc điểm hình ảnh, kiến thức y khoa. Thêm vào đó nhóm chưa thể cải thiện hiệu quả của công cụ xóa tóc trên các ảnh. Để nâng cao hiệu quả tổng thể của mô hình, nhóm cần tiếp tục nghiên cứu thu thập thêm dữ liệu cho các lớp có số lượng ít, đặc biệt là các lớp có hiệu quả thấp và cải thiện mô hình, dữ liệu và quá trình huấn luyện.

7 Kết luận và hướng phát triển

Nghiên cứu đã cho thấy tiềm năng của machine learning trong việc phân loại các tổn thương da. Tuy nhiên đồ án còn tồn tại nhiều hạn chế.

Việc chỉ sử dụng các kỹ thuật Data Augmentation cơ bản như xoay, dịch chuyển, phóng to/thu nhỏ ảnh có thể dẫn đến việc tạo ra các ảnh mới không hoàn toàn đại diện cho thực tế và thiếu đa dạng. Trong bài báo A Survey on Deep Learning for Skin Lesion Segmentation, phương pháp GAN được đề xuất cho việc augmentation. Phương thức này học các mẫu và biến thể có mặt trong dữ liệu thực tế. Vì vậy, nó có thể tạo ra các mẫu dữ liệu mới hoàn toàn và thực tế, mở rộng đáng kể tập dữ liệu đào tạo và tạo nên sự đa dạng mà phương pháp augmentation truyền thống không có (Mirikharaji et al., 2023).

Việc thu thập dữ liệu từ một khu vực nhất định như Châu Mỹ và Châu Âu có thể khiến mô hình không học được các đặc điểm của tổn thương da phổ biến ở các khu vực khác trên thế giới, đặc biệt là Châu Á. Điều này có thể dẫn đến việc mô hình phân loại sai các ảnh da của người dân ở khu vực này. Trong tương lai cần thu thập dữ liệu từ nhiều khu vực địa lý và nguồn khác nhau để đảm bảo tính đa dạng và đại diện cho tập dữ liệu huấn luyện.

Phương pháp DullRazor, mặc dù phổ biến, vẫn tồn tại một số hạn chế khi áp dụng cho việc xóa tóc trong ảnh da: Hiệu suất không ổn định: hiệu quả có thể thay đổi tùy thuộc vào ảnh da, đặc biệt ảnh

có tóc sáng hoặc nền phức tạp (Abbas et al., 2011) hơn nữa, phương pháp này còn có thể tạo ra hiện vật: Có thể xuất hiện các vết sáng hoặc đường viền không đồng đều trên da sau khi xử lý (Lameski et al., 2019), ngoài ra còn gây khó khăn trong việc xử lý trường hợp phức tạp: phương pháp gặp khó khăn khi xử lý tóc xoăn, dày hoặc che khuất chi tiết quan trọng (Perez et al., 2018). Do những hạn chế này, việc áp dụng DullRazor cho ảnh da có thể dẫn đến kết quả không mong muốn hoặc không hiệu quả. Cần cân nhắc sử dụng các phương pháp mới như dựa trên CNN (He et al., 2017) hoặc GAN (Zhu et al., 2017) để khắc phục những hạn chế này và mang lại hiệu quả tốt hơn.

Việc sử dụng mô hình học máy để phân loại tự động các tổn thương da có tiềm năng to lớn trong việc hỗ trợ chẩn đoán sớm, nâng cao hiệu quả điều trị và giảm chi phí y tế. Sau đây là một số hướng phát triển mà nhóm cho là tiềm năng cho mô hình phân loại tổn thương da qua ảnh soi da: nâng cao hiệu quả mô hình bằng cách áp dụng các mô hình học sâu, mở rộng nguồn dữ liệu; mở rộng khả năng phân loại nhiều loại tổn thương da, phân biệt các giai đoạn phát triển của tổn thương, phát hiện dấu hiệu của ung thư; phát triển ứng dụng di động hỗ trợ chuẩn đoán,...

References

- Qaisar Abbas, M Emre Celebi, and Irene Fondón García. 2011. Hair removal methods: A comparative study for dermoscopy images. *Biomedical Signal Processing and Control*, 6(4):395–404.
- Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. 2000. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424.
- Gérard Biau and Erwan Scornet. 2016. A random forest guided tour. *Test*, 25:197–227.
- Bill Cassidy, Connah Kendrick, Andrzej Brodzicki, Joanna Jaworek-Korjakowska, and Moi Hoon Yap. 2021. *Analysis of the isic image datasets: Usage, benchmarks and recommendations*. *Medical Image Analysis*.
- Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118.
- Flavia Grignaffini, Francesco Barbuto, Lorenzo Piazza, Maurizio Troiano, Patrizio Simeoni, Fabio Mangini, Giovanni Pellacani, Carmen Cantisani, and Fabrizio

- Frezza. 2022. Machine learning approaches for skin cancer classification from dermoscopic images: a systematic review. *Algorithms*, 15(11):438.
- Adil H Khan, DNF Iskandar, Jawad F Al-Asad, and Samir El-Nakla. 2021. Classification of skin lesion with hair and artifacts removal using black-hat morphology and total variation. *International Journal of Computing and Digital Systems*, 10:597–604.
- Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. [Deep residual learning for image recognition](#).
- Vikramaditya Jakkula. 2006. Tutorial on support vector machine (svm). *School of EECS, Washington State University*, 37(2.5):3.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Jane Lameski, Andrej Jovanov, Eftim Zdravevski, Petre Lameski, and Sonja Gievska. 2019. Skin lesion segmentation with deep learning. In *IEEE EUROCON 2019-18th International Conference on Smart Technologies*, pages 1–5. IEEE.
- Tim Lee, Vincent Ng, Richard Gallagher, Andrew Coldman, and David McLean. 1997. Dullrazor®: A software approach to hair removal from images. *Computers in biology and medicine*, 27(6):533–543.
- Zahra Mirikharaji, Kumar Abhishek, Alceu Bissoto, Catarina Barata, Sandra Avila, Eduardo Valle, M Emre Celebi, and Ghassan Hamarneh. 2023. A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, page 102863.
- Heng Pan, Chenyang Liu, Wenxiao Wang, Li Yuan, Hongfa Wang, Zhifeng Li, and Wei Liu. 2023. [Img2vec: A teacher of high token-diversity helps masked autoencoders](#).
- Fábio Perez, Cristina Vasconcelos, Sandra Avila, and Eduardo Valle. 2018. Data augmentation for skin lesion analysis. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5*, pages 303–311. Springer.
- Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. 2018. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9.
- Francisco J Valverde-Albacete and Carmen Peláez-Moreno. 2014. 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PloS one*, 9(1):e84217.
- Mehwish Zafar, Muhammad Imran Sharif, Muhammad Irfan Sharif, Seifedine Kadry, Syed Ahmad Chan Bukhari, and Hafiz Tayyab Rauf. 2023. Skin lesion analysis and cancer detection based on machine/deep learning techniques: A comprehensive survey. *Life*, 13(1):146.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.