

Genetic Algorithm Optimized Stacking Approach to Skin Disease Detection

NGUYEN LE THANH MINH¹ NGUYEN BA THONG², AND LUONG TRUONG THINH.³

¹University of Information Technology - VNUHCM, Thu Duc, Ho Chi Minh City Vietnam (Student ID: 22520875)

²University of Information Technology - VNUHCM, Thu Duc, Ho Chi Minh City Vietnam (Student ID: 22521433)

³University of Information Technology - VNUHCM, Thu Duc, Ho Chi Minh City Vietnam (Student ID: 22521412)

ABSTRACT Do sự tồn tại của khoảng 3000 bệnh da khác nhau, việc chẩn đoán và điều trị các bệnh lý này gặp nhiều thách thức. Điều này đòi hỏi các phương pháp phát hiện bệnh da phải có độ chính xác cao để đảm bảo hiệu quả điều trị. Các công cụ hiện tại dựa trên học sâu thường hạn chế về phạm vi bao phủ bệnh, sử dụng tập dữ liệu huấn luyện nhỏ và hiếm khi đạt độ chính xác Top-5 vượt quá 70%. Một hệ thống lý tưởng cần có khả năng phát hiện và phân loại bệnh da toàn diện, đồng thời tính đến các yếu tố môi trường và bối cảnh. Nhằm giải quyết vấn đề này, chúng tôi đề xuất một hệ thống học sâu sử dụng phương pháp ensemble, kết hợp các kiến trúc hiện có và áp dụng kỹ thuật mã hóa để tối ưu khả năng thích ứng với nhiều dạng dữ liệu đầu vào. Hệ thống được huấn luyện trên tập dữ liệu DermNet và sử dụng thuật toán di truyền để tối ưu hóa quá trình ensemble, đạt độ chính xác Top-5 là 87%, cải thiện 13% so với các nghiên cứu được so sánh.

INDEX TERMS Genetic algorithm, deep learning, ensembling, skin disease classification, dermnet

I. INTRODUCTION

Mặc dù các bệnh về da rất phổ biến, nhiều người lại không tìm đến bác sĩ hay chuyên gia để kiểm tra tình trạng da của mình. Các bệnh về da đã gây ra hơn 98.000 ca tử vong trên toàn thế giới và khoảng 5 tỷ ca mắc bệnh trong năm 2019 [1]. Ngoài ra, bệnh về da có thể dẫn đến tàn tật, làm giảm chất lượng cuộc sống cũng như tuổi thọ. Một thước đo được sử dụng để đánh giá số năm bị mất do tàn tật hoặc sức khỏe kém được gọi là Disability-Adjusted Life Years (DALY), giúp định lượng gánh nặng bệnh tật. Theo thống kê, trong năm 2019, các bệnh về da chiếm 1,79% gánh nặng bệnh tật toàn cầu tính theo DALY và vẫn là nguyên nhân chính gây ra các khuyết tật kéo dài suốt đời [1]. Bảng 1 dưới đây tổng hợp dữ liệu từ Nghiên cứu Gánh nặng Bệnh tật Toàn cầu [1], cho thấy số ca mắc bệnh, DALY và số ca tử vong do bệnh da ngày càng tăng qua các năm.

Year	Incidents (in Billions)	DALYs (in Millions)	Deaths (in thousands)
1990	3.127	32.918	49.281
1995	3.399	32.073	59.896
2000	3.660	33.331	61.621
2005	3.953	34.981	76.265
2010	4.261	38.668	95.835
2015	4.620	42.035	82.105
2019	4.859	42.883	98.521

Bng 1. Number of incidents, DALYs and deaths due to skin diseases[1]

Nhiều phương pháp phát hiện bệnh da đã được triển khai trong những năm qua, nhưng vẫn còn tồn tại một

số hạn chế cần khắc phục để xây dựng một hệ thống hiệu quả hơn. Hầu hết các giải pháp hiện có được nghiên cứu trong quá trình phát triển hệ thống này chỉ sử dụng một tập dữ liệu nhỏ gồm các hình ảnh để huấn luyện mô hình học sâu. Tập dữ liệu nhỏ dễ dẫn đến độ chính xác thấp do hiện tượng underfitting, và có thể không cung cấp đủ thông tin cần thiết để xây dựng một hệ thống mạnh mẽ. Một số hệ thống khác chỉ được thiết kế để phát hiện một số lượng nhỏ các bệnh da, làm giảm tính linh hoạt của chúng. Ngoài ra, một số hệ thống không thể xử lý tốt các thay đổi về môi trường và kết cấu trong hình ảnh đầu vào – đây là yếu tố quan trọng vì điều kiện ánh sáng và môi trường không phải lúc nào cũng được kiểm soát. Bên cạnh đó, một số hệ thống sử dụng các mô hình không đủ mạnh để phát hiện và phân tích một danh sách đa dạng các bệnh da. Số lượng lớn các bệnh về da [2], cùng với những thách thức từ môi trường thay đổi và tập dữ liệu nhỏ, khiến việc phân loại bệnh da trở thành một vấn đề quan trọng.

Ứng dụng của học sâu đã lan rộng ra nhiều lĩnh vực, trong đó có y tế. Học sâu trong lĩnh vực y tế đang nổi lên nhờ vào khả năng nâng cao chất lượng chăm sóc sức khỏe thông qua các phương pháp tiếp cận dựa trên học sâu [3], [4], [5]. Một số ứng dụng phổ biến của học sâu bao gồm phân tích hình ảnh [6], [7], phân tích dữ liệu y tế, và phát triển thuốc. Các hệ thống học sâu được xây dựng để phát hiện bệnh da có thể hỗ trợ bệnh nhân và bác sĩ trong việc chẩn đoán sớm, từ đó có phương pháp

điều trị kịp thời nhằm ngăn ngừa tử vong, giảm nguy cơ biến chứng và tàn tật. Các bệnh về da có thể được nhận biết dựa trên các đặc điểm hình ảnh của chúng, điều này gợi ý rằng một mô hình học sâu có thể được sử dụng để phát hiện bệnh da từ các hình ảnh. Điều này sẽ hỗ trợ các chuyên gia y tế trong việc đưa ra quyết định tốt hơn. Hệ thống này cũng có thể được bệnh nhân sử dụng để tự phân tích tại nhà, từ đó giảm thiểu nguy cơ dẫn đến các hậu quả nghiêm trọng. Một hệ thống mạnh mẽ và dễ tiếp cận hơn sẽ giúp thu hẹp khoảng cách giữa cơ sở y tế và chuyên gia, đồng thời làm cho việc kiểm tra và chẩn đoán trở nên đáng tin cậy, chính xác và tiện lợi hơn.

Mục tiêu chính của việc phát triển hệ thống học sâu là sử dụng một tập dữ liệu toàn diện, bao quát đầy đủ các loại bệnh da khác nhau – yếu tố quan trọng với số lượng lớn các tình trạng da ảnh hưởng đến con người. Hệ thống này cũng phải có khả năng phát hiện nhiều bệnh da khác nhau, điều này sẽ nâng cao ứng dụng của nó trong chẩn đoán y tế. Ngoài ra, hệ thống cần đạt độ chính xác cao để có thể tin cậy trong các chẩn đoán thực tế.

Hệ thống đề xuất có thể được sử dụng trong môi trường y tế, hỗ trợ các chuyên gia trong việc xác định bệnh da và định hướng chẩn đoán. Độ chính xác Top-5 đáng kể của hệ thống chứng minh hiệu quả trong việc cung cấp danh sách các bệnh có khả năng cao, hỗ trợ các chuyên gia y tế đưa ra chẩn đoán chính xác.

II. LITERATURE REVIEW

Một phương pháp sử dụng mạng nơ-ron tích chập (CNN) [8] đã được áp dụng để trích xuất đặc trưng từ một tập dữ liệu gồm 100 hình ảnh và thực hiện việc phát hiện bệnh da thông qua máy vector hỗ trợ (SVM). Cách tiếp cận này cho thấy tiềm năng của việc sử dụng CNN để trích xuất đặc trưng, giúp cải thiện độ chính xác. Phân loại bệnh da bằng các kiến trúc như MobileNetV2 [9], [10] sử dụng hình ảnh làm đầu vào, sau đó đầu ra của mô hình MobileNetV2 được chuyển đến một đơn vị Long Short-Term Memory (LSTM) và cuối cùng là lớp dày đặc (dense layer) để tạo ra kết quả phân loại hình ảnh, đạt độ chính xác khoảng 85%. Một hệ thống kết hợp CNN như một công cụ trích xuất đặc trưng và mạng nơ-ron nhân tạo (ANN) [11] để thực hiện phân loại đã đạt độ chính xác cao hơn là 85,14

Một hệ thống dự đoán bệnh da sử dụng dữ liệu tổng hợp và lựa chọn đặc trưng [12] áp dụng hai phương pháp. Một phương pháp sử dụng mô hình tổ hợp với ba thuật toán kết hợp cùng phân tích phương sai tuyến tính (LDA), bộ phân loại bị động (PAC), bộ phân loại Radius Neighbors (RNC), bộ phân loại Bernoulli Naïve Bayesian (BNB), Naïve Bayes (NB) và Extra Tree Classifier (ETC) cùng với gradient boosting. Phương pháp còn lại sử dụng phương pháp tương tự nhưng thêm bộ trích xuất đặc trưng để giảm kích thước tập dữ liệu. Hệ thống này không làm việc với tập dữ liệu

hình ảnh và chỉ áp dụng cho số lượng lớp nhỏ hơn.

Phát hiện ung thư da bằng MobileNetV2 [13] thông qua phương pháp học chuyển giao (transfer learning) đã đạt độ chính xác 98,2%, giúp phân loại hình ảnh thành ung thư hoặc không ung thư. Một số phương pháp khác bao gồm việc chẩn đoán bệnh da với kiến trúc mô hình tùy chỉnh sử dụng bộ lọc Gabor 2D, hoạt động trên các hình ảnh với sắc độ da khác nhau [14]. Một kiến trúc CNN tùy chỉnh khác sử dụng phân tích thành phần chính (PCA) đạt độ chính xác 90% [15], tuy nhiên hệ thống này chỉ nhận diện tốt các tế bào bạch cầu mà gặp khó khăn với một số độ phân giải ảnh.

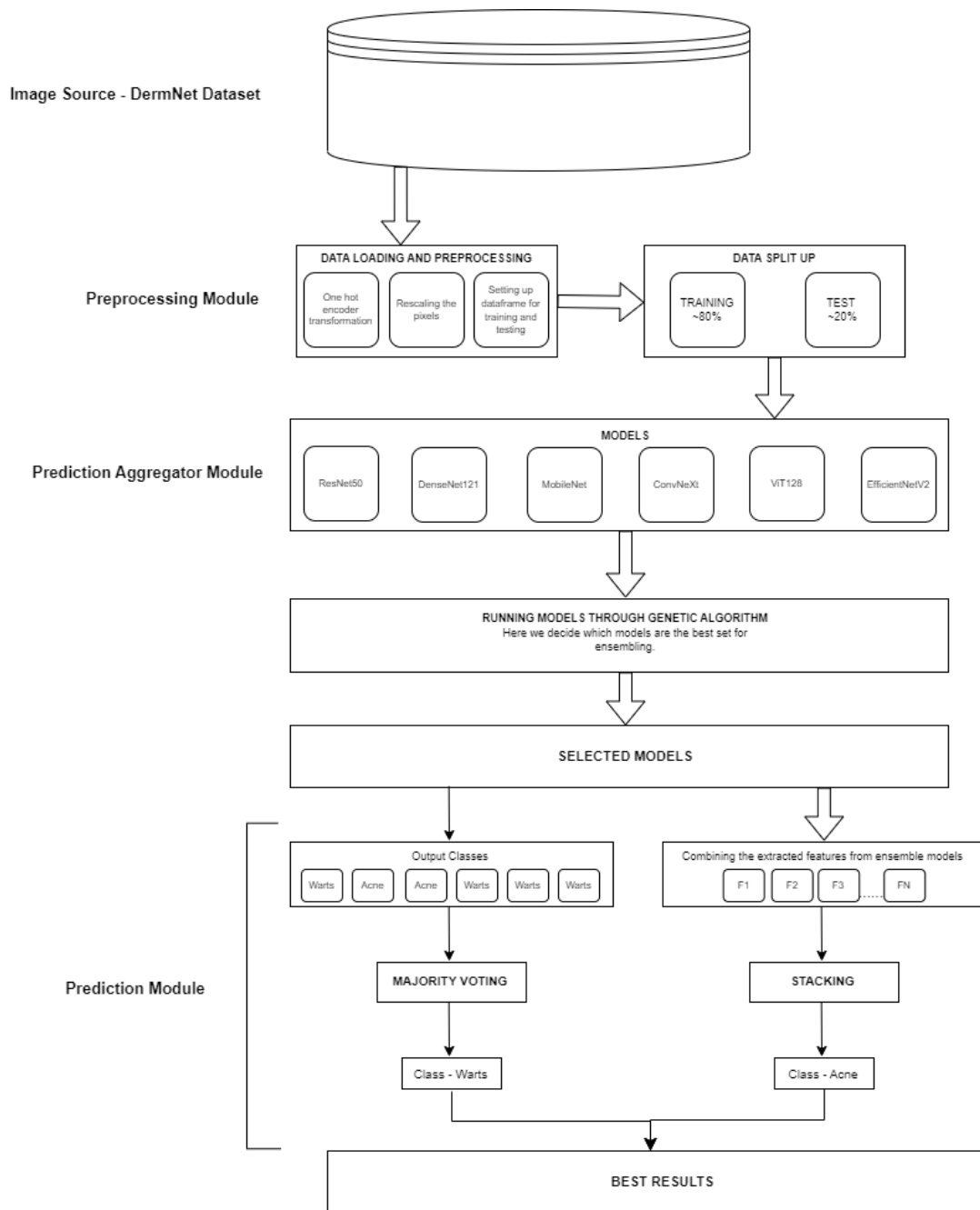
Một nghiên cứu sử dụng thuật toán máy học cho việc phát hiện bệnh da [16] đã đánh giá nhiều mô hình như SVM, hồi quy logistic, rừng ngẫu nhiên (Random Forest), Naïve Bayes và CNN, kết luận rằng mô hình CNN đạt độ chính xác cao nhất 96%. Trong khi đó, việc áp dụng mô hình ResNet [17] để phát hiện ung thư da [18] thông qua học chuyển giao lại có độ chính xác thấp hơn (57%) so với các mô hình khác. Một phương pháp học chuyển giao mới [19] được sử dụng với các tập dữ liệu lớn hơn và đạt được độ chính xác cao hơn thông qua CNN sâu (DCNN), đạt điểm F1 lên đến 98% khi kết hợp với học chuyển giao kép, mặc dù thời gian huấn luyện dài hơn.

Một phương pháp sử dụng CNN để phân loại bệnh da [20] dựa trên các giai đoạn cụ thể của melanoma, bao gồm tổn thương ác tính, melanoma lan tỏa bề mặt và melanoma nốt. Để giải quyết vấn đề thiếu tập dữ liệu lớn, DermGAN [21] được áp dụng để tổng hợp hình ảnh với các tình trạng da bằng mạng đối kháng sinh (GAN). Một nghiên cứu so sánh giữa các mô hình học máy và học sâu [22] chỉ ra rằng mô hình cây tổ hợp Bagged Tree Ensemble đạt độ chính xác 92,99%, tuy nhiên hiệu năng của các mô hình học máy có thể không duy trì tốt trên các tập dữ liệu lớn hơn.

Cuối cùng, việc sử dụng phương pháp tổ hợp với CNN để phân loại bệnh da yêu cầu tập dữ liệu lớn hơn, bằng cách tăng số lượng dữ liệu thông qua sao chép [23], đã đạt được độ chính xác trung bình khoảng 92%. Các đặc trưng về màu sắc và kết cấu của hình ảnh [24] cũng được khai thác để phân loại bệnh da, tuy nhiên phương pháp này có thể bị ảnh hưởng bởi các yếu tố môi trường. Việc sử dụng các hệ thống tổ hợp với các phương pháp tối ưu hóa như thuật toán di truyền [29] đã cho thấy khả năng cải thiện độ chính xác vượt trội khi so sánh với các hệ thống tổ hợp khác.

III. PROPOSED SYSTEM

Hệ thống được đề xuất sẽ triển khai ba mô-đun chính để hoạt động hiệu quả. Ba mô-đun bao gồm: mô-đun tiền xử lý dữ liệu, mô-đun tổng hợp dự đoán dành cho các phương pháp tiếp cận ensemble, và mô-đun dự đoán. Các hình ảnh trong nghiên cứu được lấy từ tập dữ liệu DermNet. Mô-đun tiền xử lý đảm nhiệm việc tiền xử lý phù hợp các hình ảnh trong tập dữ liệu, bao gồm

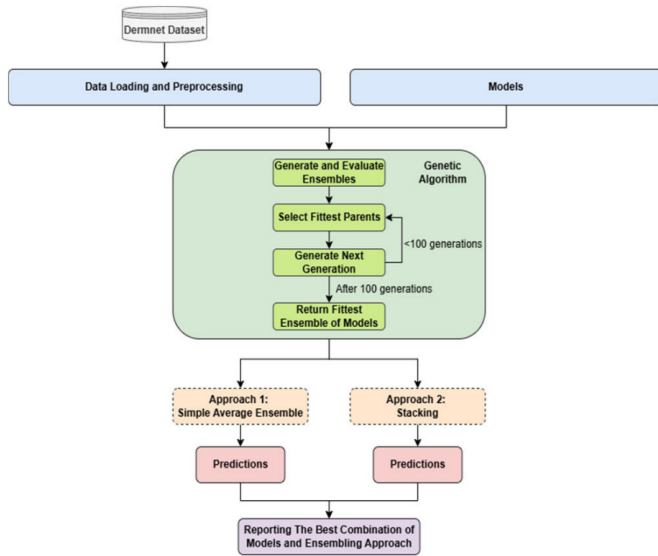


Hình 1. Block diagram of the proposed system

thay đổi kích thước, chuẩn hóa và hiệu chỉnh hình ảnh. Đối với các phương pháp sử dụng ensemble, mô-đun tổng hợp dự đoán sẽ thu thập kết quả từ tất cả các mô hình trong tập hợp để đưa ra dự đoán cuối cùng. Mô-đun dự đoán cung cấp kết quả đầu ra từ hệ thống đề xuất cho một hình ảnh đầu vào bất kỳ. Hệ thống học máy trong nghiên cứu sử dụng năm mô hình học sâu, bao gồm ResNet50[25], DenseNet121[26], MobileNet[10], ConvNeXt[27] và ViT128[28]. Để xác định các tổ hợp mô hình tối ưu, thuật toán di truyền (GA)[29][30] được áp dụng nhằm xây dựng các tổ hợp mô hình tích hợp

với độ chính xác vượt trội so với từng mô hình riêng lẻ. Phương pháp này giúp khắc phục các hạn chế khi sử dụng các mô hình độc lập.

Do yêu cầu hệ thống cần đạt hiệu suất cao, nhiều mô hình đã được lựa chọn và đánh giá trên tập dữ liệu mục tiêu nhằm xác định mô hình phù hợp. Các bước xử lý dữ liệu và đánh giá mô hình được mô tả trong Hình 1. Đầu tiên, tập dữ liệu DermNet được tải lên, sau đó thực hiện tiền xử lý hình ảnh, bao gồm mã hóa, biến đổi và chuẩn hóa giá trị điểm ảnh để đảm bảo tính đồng nhất trước khi đưa vào mô hình. Dữ liệu sau



Hình 2. GA System

khi được tiền xử lý được chia thành tập huấn luyện và tập kiểm thử. Quá trình huấn luyện sử dụng tập huấn luyện, trong khi tập kiểm thử được dùng để đánh giá hiệu suất của mô hình sau khi huấn luyện. Trong quá trình huấn luyện, giá trị mất mát (loss) và độ chính xác (accuracy) trong mỗi vòng lặp (epoch) được lưu lại để phân tích. Sau khi các mô hình được huấn luyện xong, hiệu suất của chúng được đánh giá dựa trên độ chính xác trên tập kiểm thử. Giá trị hiệu suất này được sử dụng làm tiêu chí để thuật toán di truyền tìm kiếm các tổ hợp mô hình tối ưu. Thuật toán GA sau đó tạo các tổ hợp dựa trên soft stacking, hard stacking, soft voting, và blending.

Các bước triển khai thuật toán di truyền được trình bày trong Hình 2. Giá trị độ phù hợp (fitness value) của từng tổ hợp được tính toán dựa trên độ chính xác kiểm thử. Dựa vào giá trị fitness, thuật toán tiến hành tạo các phép lai ghép (crossover) giữa hai tổ hợp khác nhau, ưu tiên các tổ hợp có giá trị fitness cao. Đồng thời, một số biến số như tỷ lệ lai ghép ngẫu nhiên (random crossover rate) và tỷ lệ đột biến (mutation rate) được điều chỉnh để đảm bảo tính đa dạng của các tổ hợp. Điều kiện dừng của thuật toán là hoàn thành 100 vòng lặp. Sau đó, tổ hợp mô hình tốt nhất sẽ được chọn để triển khai.

Kiến trúc cuối cùng của hệ thống sử dụng các phương pháp tích hợp GA đã lựa chọn để phát triển một hệ thống tối ưu trong việc phát hiện bệnh da liễu.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Thiết lập thử nghiệm của công trình được đề xuất bao gồm việc sử dụng thư viện Pytorch để giúp xây dựng một hệ thống máy học và OpenCV để thực hiện tăng cường và xử lý hình ảnh. Mô hình được huấn luyện trên tập dữ liệu DermNet lấy từ Dermatology Resource, một

thư viện chứa các bộ dữ liệu về bệnh da liễu. Tập dữ liệu bao gồm hình ảnh bệnh da ở định dạng JPG và các hình ảnh được chia thành 23 lớp khác nhau dựa trên các loại bệnh về da. Bộ dữ liệu có hơn 19.000 hình ảnh, được chia thành các 2 tập huấn luyện và kiểm tra để huấn luyện và đánh giá mô hình. Các lớp có trong bộ dữ liệu như sau:

- Acne and Rosacea : Mụn trứng cá và chứng đỏ mặt
- Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions : Chứng dày sừng ánh sáng, ung thư tế bào đáy và các tổn thương ác tính khác
- Atopic Dermatitis : Viêm da cơ địa
- Bullous Disease : Bệnh bóng nước
- Cellulitis Impetigo and other Bacterial Infections : Viêm mô tế bào, chốc lở và các bệnh nhiễm trùng do vi khuẩn khác
- Eczema : Bệnh chàm
- Exanthems and Drug Eruptions : Ban xuất huyết và phát ban do thuốc
- Hair Loss Alopecia and other Hair Diseases : Rụng tóc, hói đầu và các bệnh về tóc khác
- Herpes HPV and other STDs : Herpes, HPV và các bệnh lây truyền qua đường tình dục khác
- Light Diseases and Pigmentation Disorders : Các bệnh về ánh sáng và rối loạn sắc tố
- Lupus and Connective Tissue Diseases : Lupus và các bệnh mô liên kết
- Melanoma Skin Cancer Nevi and Moles : U hắc tố da, nốt ruồi và nốt ruồi son
- Nail Fungus and other Nail Diseases : Nấm móng và các bệnh về móng khác
- Poison Ivy and other contact dermatitis : Cây thường xuân độc và các bệnh viêm da tiếp xúc khác
- Psoriasis Lichen Planus : Bệnh vẩy nến, lichen phẳng
- Scabies Lyme and other infestations and Bites : Bệnh ghẻ, Lyme và các bệnh nhiễm trùng và vết cắn khác
- Seborrheic Keratoses and other Benign tumors : Chứng dày sừng tiết bã và các khối u lành tính khác
- Systemic diseases : Các bệnh hệ thống
- Tinea Ringworm Candidiasis and other Fungal Infections : Nấm da, nấm Candida và các bệnh nhiễm trùng do nấm khác
- Urticaria Hives : Mày đay
- Vascular Tumors : Các khối u mạch máu
- Vasculitis : Viêm mạch máu
- Warts Molluscum and Viral Infections : Mụn cóc, u mềm lây và nhiễm trùng do virus

Bảng 2 trình bày các tham số được sử dụng trong quá trình huấn luyện và kiểm thử các mô hình. Việc lựa chọn kích thước batch, bộ tối ưu hóa, và hàm mất mát được thực hiện dựa trên các thí nghiệm trước đó,

Hyperparameter	Value
Batch Size	128
Optimizer	Adam
Loss function	Categorical Cross Entropy
Train Test split ratio	80% for train, 20% for test
Epoch	40

Bng 2. Initialization and hyperparameters.

nơi các giá trị được chọn thể hiện hiệu suất tốt hơn với cùng số lượng epoch. Trong các thí nghiệm này, nhiều giá trị khác nhau đã được kiểm tra, bao gồm kích thước batch là 16, 32, 64, 128 và 256. Kích thước batch là 8 ban đầu đã được xem xét nhưng sau đó bị loại bỏ do làm tăng thời gian huấn luyện, dẫn đến việc áp dụng thuật toán di truyền trở nên không khả thi.

Các lựa chọn cho bộ tối ưu hóa bao gồm Adam Optimizer, Stochastic Gradient Descent Optimizer (SGD), và RMSProp Optimizer. Đối với hàm mất mát, cả categorical cross-entropy và sparse categorical cross-entropy đều đã được thử nghiệm. Tỷ lệ chia dữ liệu train-test được xác định theo thực tiễn chuẩn, với việc phân bổ ngẫu nhiên 80% dữ liệu để huấn luyện và 20% để kiểm thử.

Để đánh giá hiệu suất của từng mô hình và hệ thống tổ hợp được huấn luyện, các chỉ số đánh giá như độ chính xác (accuracy), độ chính xác (precision), khả năng thu hồi (recall), điểm F1 (F1 score), và Top-5 accuracy đã được sử dụng. Việc sử dụng nhiều chỉ số hiệu suất mang lại sự hiểu biết toàn diện về hiệu quả của mỗi mô hình và cho phép đưa ra quyết định có cơ sở trong việc lựa chọn các mô hình được đưa vào. Tất cả mọi mô hình đều được huấn luyện qua 40 epoch để đảm bảo tính công bằng.

Một số nghiên cứu được xem xét đã ưu tiên sử dụng thống kê Top-5 để hiểu rõ hơn và xếp loại hiệu suất của các mô hình được huấn luyện và triển khai. Độ chính xác Top-5 (Top-5 accuracy) sử dụng 5 dự đoán hàng đầu của mô hình dựa trên điểm số độ tin cậy và so sánh chúng với giá trị thực để tính toán thống kê này. Để có sự so sánh tương thích hơn, các mô hình được huấn luyện trong công trình này đã được so sánh với các mô hình khác dựa trên độ chính xác Top-5 cùng với chỉ số độ chính xác cổ điển.

Các mô hình được liệt kê dưới đây đã được kiểm thử, và độ chính xác của mô hình cũng như mất mát của mô hình đã được ghi nhận.

A. ALL LAYERS FROZEN

Các mô hình sau đây đã được đóng băng tất cả các trọng số, chỉ một vài lớp ở phía đầu vào và đầu ra được để mở để cho phép thay đổi trọng số của các lớp đó. Bảng 3 là kết quả của các mô hình khi đóng băng trọng số.

Khi đóng băng trọng số của các mô hình, nhìn chung hiệu suất của các mô hình không đạt mức cao. Trong số

đó, mô hình ConvNeXt nổi bật với hiệu quả tốt nhất, đạt độ chính xác (Accuracy) cao nhất là 0.4 và điểm F1 (F1 Score) cũng đạt 0.4, vượt trội hơn so với các mô hình còn lại.

B. ALL LAYERS UNFROZEN

Ở phần này, mọi mô hình đều được mở trọng số, bên cạnh đó vẫn cần phải tinh chỉnh lớp đầu vào và đầu ra để đảm bảo phù hợp với bộ dữ liệu. Bảng 4 là kết quả của các mô hình khi đóng băng trọng số.

Trường hợp fine-tune trên toàn bộ các lớp của các mô hình trên bộ dữ liệu này, mô hình EfficientNet đạt được độ chính xác cao nhất là 0.54 và điểm F1 là 0.54. Bên cạnh đó, độ đo top-5 accuracy cũng được cải thiện, với hiệu suất đạt được là 0.83.

C. ENSEMBLE MODELS

Trong nghiên cứu này, chúng tôi áp dụng thuật toán di truyền (Genetic Algorithm) để tối ưu hóa việc lựa chọn tổ hợp các mô hình thành phần, nhằm đạt được độ chính xác cao nhất khi áp dụng các kỹ thuật Ensemble như Stacking và Voting. Quá trình thực nghiệm bao gồm việc so sánh hai phương pháp Blending (huấn luyện meta-model trên một phần dữ liệu) và Stacking (huấn luyện meta-model trên toàn bộ dữ liệu). Bên cạnh đó, chúng tôi cũng đánh giá hiệu suất của cả hai biến thể Hard và Soft của Stacking và Voting. Cụ thể, Hard Stacking/Voting chỉ sử dụng nhãn dự đoán đầu ra từ các mô hình cơ sở (base models), trong khi Soft Stacking/Voting tận dụng vector xác suất đầu ra của chúng. Kết quả chi tiết của các mô hình Ensemble được đề xuất, cùng với kết quả từ bài báo tham khảo, được trình bày trong Bảng 5.

Phân tích kết quả thực nghiệm cho thấy tổ hợp các mô hình cơ sở do chúng tôi đề xuất đã mang lại sự cải thiện đáng kể về hiệu suất. Cụ thể, độ chính xác (Accuracy) đạt 0.61, vượt trội so với 0.45 được báo cáo trong bài báo tham khảo và 0.54 của mô hình Efficientnet. Kết quả cũng chỉ ra rằng việc sử dụng vector xác suất đầu ra (Soft Stacking/Voting) hiệu quả hơn đáng kể so với việc chỉ sử dụng nhãn dự đoán (Hard Stacking/Voting). Thử nghiệm với phương pháp Blending, trong đó meta-model chỉ được huấn luyện trên một phần dữ liệu, cho kết quả thậm chí còn kém hơn so với việc sử dụng một mô hình riêng lẻ, với độ chính xác chỉ đạt 0.44. Đáng chú ý, hiệu suất của Voting khá tương đồng với Stacking, cho thấy Voting là một lựa chọn hiệu quả với chi phí huấn luyện thấp hơn đáng kể so với Stacking mà vẫn đạt được hiệu suất gần tương đương.

V. CONCLUSION AND FUTURE WORK

Hệ thống phát hiện bệnh da liễu với độ chính xác cao là điều cần thiết do tầm quan trọng của các quyết định được đưa ra trong lĩnh vực này. Trong nỗ lực tìm kiếm một phương pháp hiệu quả để phát hiện bệnh da liễu, nhiều mô hình khác nhau đã được nghiên cứu, cùng

Model	Accuracy	F1 Score	Precision	Recall	Top-5 Acc
Densenet frozen	0.3365	0.3348	0.3754	0.3365	0.6816
Resnet50 frozen	0.2623	0.2564	0.3132	0.2623	0.6246
Mobilenet frozen	0.358	0.3622	0.376	0.358	0.7071
Vit frozen	0.3823	0.3786	0.3987	0.3823	0.7418
Efficientnet frozen	0.2696	0.2753	0.3051	0.2696	0.6014
Convnext frozen	0.3993	0.4008	0.4259	0.3993	0.7701

Bng 3. Performance metrics for various frozen models.

Model	Accuracy	F1 Score	Precision	Recall	Top-5 Acc
Mobilenet unfrozen	0.4725	0.4673	0.4764	0.4725	0.7928
Densenet unfrozen	0.5174	0.5097	0.5221	0.5174	0.82
Resnet50 unfrozen	0.2123	0.2017	0.2461	0.2123	0.5694
Convnext unfrozen	0.5194	0.51	0.5264	0.5194	0.8143
Efficientnet unfrozen	0.5419	0.54	0.5447	0.5419	0.8293

Bng 4. Performance metrics for various unfrozen models.

Model	Accuracy	F1 Score	Precision	Recall	Top-5 Acc
Paper	0.4545	0.362	0.8125	0.232	0.7338
GA Hard Stacking	0.2296				
GA Blending	0.44				
Stacking Model All	0.5517	0.5432	0.5697	0.5517	0.8318
GA Soft Voting (4 models)	0.5995	0.5986	0.5965	0.5988	0.8758
EfficientNet unfrozen	0.5419	0.54	0.5447	0.5419	0.8293
GA Soft Stacking (ConvNeXt frozen, ConvNeXt unfrozen, DenseNet frozen, DenseNet unfrozen, MobileNet frozen, EfficientNet unfrozen, ViT frozen)	0.6059	0.5996	0.6064	0.6059	0.8696

Bng 5. Performance metrics for various models.

với việc áp dụng các kỹ thuật tích hợp (ensembling) và thuật toán di truyền để xác định các tổ hợp mô hình tối ưu. Việc nâng cao hiệu quả của mô hình trong các môi

trường khác nhau là điều thiết yếu, và mô hình xử lý hình ảnh đã góp phần cải thiện khả năng thích nghi của hệ thống. Một phương pháp tích hợp dựa trên stacking

đã được triển khai, kết hợp nhiều mô hình nhằm nâng cao hiệu suất tổng thể của hệ thống.

Mô hình phát hiện bệnh da liễu tích hợp đã chứng minh một hướng tiếp cận tiềm năng trong việc xây dựng hệ thống phát hiện bệnh da liễu với độ chính xác cao. Cụ thể, mô hình đạt được độ chính xác Top-5 là 87% trên tập dữ liệu DermNet, cao hơn 14% so với nghiên cứu đối chiếu[31].

Để cải thiện hơn nữa các kết quả đạt được, nghiên cứu trong tương lai có thể hướng tới việc triển khai các kỹ thuật tích hợp dựa trên trung bình trọng số (weighted-average-based ensemble). Đồng thời, phương pháp tích hợp có thể được tinh chỉnh thông qua việc áp dụng các kiến trúc mạng nơ-ron nhân tạo khác nhau. Ngoài ra, việc thay đổi các tham số riêng lẻ trong các lớp của mô hình trong các thí nghiệm tiếp theo có thể mang lại kết quả tốt hơn, do các tham số này chưa được điều chỉnh trong quá trình nghiên cứu hiện tại.

Tài liệu

- [1] (IHME) IfHMaE. “Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results”. inSeattle, United States: (2023).
- [2] Roderick J Hay andothers. “The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions”. inJournal of investigative dermatology: 134.6 (2014), pages 1527–1534.
- [3] Qasem Abu Al-Haija and Adeola Adebajo. “Breast Cancer Diagnosis in Histopathological Images Using ResNet-50 Convolutional Neural Network”. in2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS): 2020, pages 1–7. DOI: 10.1109/IEMTRONICS51293.2020.9216455.
- [4] Tsamara Hanifa Arfan, Mardhiya Hayaty and Arifiyanto Hadinegoro. “Classification of Brain Tumours Types Based On MRI Images Using Mobilenet”. in2021 2nd International Conference on Innovative and Creative Information Technology (ICITech): 2021, pages 69–73. DOI: 10.1109/ICITech50181.2021.9590183.
- [5] Babak Ehteshami Bejnordi andothers. “Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images”. inJournal of Medical Imaging: 4.4 (2017), pages 044504–044504.
- [6] Tao Zhou andothers. “Dense convolutional network and its application in medical image analysis”. inBioMed Research International: 2022.1 (2022), page 2384830.
- [7] S Suganyadevi, V Seethalakshmi and Krishnasamy Balasamy. “A review on deep learning in medical image analysis”. inInternational Journal of Multimedia Information Retrieval: 11.1 (2022), pages 19–38.
- [8] Nawal Soliman ALKolifi ALEnezi. “A method of skin disease detection using image processing and machine learning”. inProcedia Computer Science: 163 (2019), pages 85–92.
- [9] Parvathaneni Naga Srinivasu andothers. “Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM”. inSensors: 21.8 (2021), page 2852.
- [10] Andrew G Howard andothers. “MobileNets: efficient convolutional neural networks for mobile vision applications (2017)”. inarXiv preprint arXiv:1704.04861: 126 (2017).
- [11] Jihan Alam. “An Efficient Approach for Skin Disease Detection using Deep Learning”. in2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE): 2021, pages 1–8. DOI: 10.1109/CSDE53843.2021.9718427.
- [12] Anurag Kumar Verma, Saurabh Pal and Surjeet Kumar. “Prediction of skin disease using ensemble data mining techniques and feature selection method—a comparative study”. inApplied biochemistry and biotechnology: 190.2 (2020), pages 341–359.
- [13] Javed Rashid andothers. “Skin cancer disease detection using transfer learning technique”. inApplied Sciences: 12.11 (2022), page 5714.
- [14] T Shanthi, RS Sabeenian and Raju Anand. “Automatic diagnosis of skin diseases using convolution neural network”. inMicroprocessors and Microsystems: 76 (2020), page 103074.
- [15] Jian Huang andothers. “[Retracted] Detection of Diseases Using Machine Learning Image Recognition Technology in Artificial Intelligence”. inComputational Intelligence and Neuroscience: 2022.1 (2022), page 5658641.
- [16] Shuchi Bhadula andothers. “Machine learning algorithms based skin disease detection”. inInternational Journal of Innovative Technology and Exploring Engineering (IJITEE): 9.2 (2019), pages 4044–4049.
- [17] Kaiming He andothers. “Deep residual learning for image recognition”. inProceedings of the IEEE conference on computer vision and pattern recognition: 2016, pages 770–778.
- [18] Kinnor Das andothers. “Machine learning and its application in skin cancer”. inInternational Journal of Environmental Research and Public Health: 18.24 (2021), page 13409.
- [19] Laith Alzubaidi andothers. “Novel transfer learning approach for medical imaging with limited labeled data”. inCancers: 13.7 (2021), page 1590.

- [20] Viswanatha Reddy Allugunti. “A machine learning model for skin disease classification using convolution neural network”. in *International Journal of Computing, Programming and Database Management*: 3.1 (2022), pages 141–147.
- [21] Amirata Ghorbani and others. “Dermgan: Synthetic generation of clinical skin images with pathology”. in *Machine learning for health workshop*: PMLR. 2020, pages 155–170.
- [22] Samir Bandyopadhyay, Amiya Bhaumik and Sandeep Poddar. “Skin disease detection: machine learning vs deep learning”. in (2021).
- [23] Medishetty Maniraju, Rudrangi Adithya and Gandu Srilekha. “Recognition of Type of Skin Disease Using CNN”. in *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR): 2022*, pages 1–4. DOI: 10.1109/ICAITPR51569.2022.9844199.
- [24] K. Veera Swamy and B. Divya. “Skin Disease Classification using Machine Learning Algorithms”. in *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4): 2021*, pages 1–5. DOI: 10.1109/C2I454156.2021.9689338.
- [25] Kaiming He and others. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV]. URL: <https://arxiv.org/abs/1512.03385>.
- [26] Gao Huang and others. *Densely Connected Convolutional Networks*. 2018. arXiv: 1608.06993 [cs.CV]. URL: <https://arxiv.org/abs/1608.06993>.
- [27] Zhuang Liu and others. *A ConvNet for the 2020s*. 2022. arXiv: 2201.03545 [cs.CV]. URL: <https://arxiv.org/abs/2201.03545>.
- [28] Alexey Dosovitskiy and others. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. arXiv: 2010.11929 [cs.CV]. URL: <https://arxiv.org/abs/2010.11929>.
- [29] John H Holland. “Genetic algorithms”. in *Scientific american*: 267.1 (1992), pages 66–73.
- [30] Seyedali Mirjalili. “Evolutionary algorithms and neural networks”. in *Studies in computational intelligence*: 780 (2019), pages 43–53.
- [31] Ananthakrishnan Balasundaram and others. “Genetic Algorithm Optimized Stacking Approach to Skin Disease Detection”. in *IEEE Access*: 12 (2024), pages 88950–88962. DOI: 10.1109/ACCESS.2024.3412791.