

Detecção da Psoríase Utilizando Visão Computacional: Uma Abordagem Comparativa Entre CNNs e Vision Transformers

Natanael Lucena¹, Fabio S. da Silva¹, Ricardo Rios¹

¹Escola Superior de Tecnologia — Universidade do Estado do Amazonas
Manaus, AM, Brasil

{nldm.eng20, fssilva, rrios}@uea.edu.br

Abstract. *This paper presents a comparison of the performance of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) in the task of multiclassifying images containing lesions of psoriasis and diseases similar to it. Models pre-trained on ImageNet were adapted to a specific data set. Both achieved high predictive metrics, but the ViTs stood out for their superior performance with smaller models. Dual Attention Vision Transformer-Base (DaViT-B) obtained the best results, with an f1-score of 96.4%, and is recommended as the most efficient architecture for automated psoriasis detection. This article reinforces the potential of ViTs for medical image classification tasks.*

Resumo. *Esse artigo apresenta uma comparação de desempenho de Redes Neurais Convolucionais (CNNs) e Vision Transformers (ViTs) na tarefa de multiclassificação de imagens contendo lesões de psoríase e de enfermidades similares a essa doença. Modelos pré-treinados no ImageNet foram adaptados a um conjunto de dados específico. Ambos alcançaram métricas preditivas elevadas, mas os ViTs se destacaram por apresentarem desempenho superior com modelos menores. O Dual Attention Vision Transformer-Base (DaViT-B) obteve os melhores resultados, com um f1-score de 96,4%, e é recomendado como a arquitetura mais eficiente para detecção automatizada de psoríase. Esse artigo reforça o potencial dos ViTs para tarefas de classificação de imagens médicas.*

1. Introdução

A psoríase é uma doença inflamatória da pele, crônica e sem cura, que afeta aproximadamente 3% da população mundial, ou seja, cerca de 125 milhões de pessoas [de Dermatologia 2023]. No Brasil, estima-se que 90% da população desconhece a doença, e apenas 6% reconhece suas lesões [Brasil 2020]. O diagnóstico atual é realizado por dermatologistas com base em achados clínicos e pistas visuais [Dash et al. 2020], sendo um processo subjetivo e demorado [Bin Ji 2022], que pode exigir biópsias para confirmação [Rodrigues and Teixeira 2009]. A complexidade do diagnóstico aumenta devido à semelhança visual da psoríase com outras doenças de pele, como dermatite, líquen plano e pitiríase rosada [Roslan et al. 2020].

A aplicação de técnicas de visão computacional e inteligência artificial (IA) tem se mostrado promissora para auxiliar no diagnóstico de doenças de pele, incluindo a psoríase [Milani et al. 2023, Olescki 2021, Silva et al. 2022]. Métodos baseados em aprendizado profundo (*Deep Learning*), como Redes Neurais Convolucionais (CNNs) e *Vision Transformers* (ViTs), têm alcançado desempenho próximo ou superior ao humano em tarefas de classificação de imagens [Wei et al. 2023, Mohan et al. 2024, Roslan et al. 2020,

Maduranga and Nandasena 2022]. Essas técnicas podem explorar características da pele que não são facilmente visíveis ao olho humano, oferecendo uma solução objetiva e potencialmente automatizada para o diagnóstico.

Esse artigo apresenta um estudo comparativo entre modelos de CNNs e ViTs para a tarefa de multi-classificação de imagens de pele com psoríase e doenças visualmente similares. O objetivo é identificar o modelo mais adequado para atuar como ferramenta de apoio ao diagnóstico dermatológico. Para isso, foram selecionados modelos de CNNs (EfficientNet-V2-L [Tan and Le 2021], ConvNeXt-L [Liu et al. 2022] e Inception-v3 [Szegedy et al. 2016]) e ViTs (ViT-L/16 [Alexey 2020], MaxViT-T [Tu et al. 2022] e DaViT-B [Ding et al. 2022]), pré-treinados no ImageNet [Deng et al. 2009] e adaptados a um conjunto de dados específico de imagens de psoríase e doenças similares.

Este trabalho expande a abordagem adotada em [Milani et al. 2023] de classificação binária para multi classificação, incluindo imagens de doenças similares, e a utilização de ViTs na comparação.

Esse artigo está organizado como segue. Na Seção 2, são apresentados os trabalhos relacionados. A Seção 3 apresenta os materiais e métodos utilizados neste trabalho. Na Seção 4, discutem-se os resultados obtidos a partir do treinamento dos modelos. Por fim, na Seção 5 são apresentadas as conclusões do trabalho.

2. Trabalhos Relacionados

[Milani et al. 2023] focou na aplicação de CNNs para a detecção de psoríase. Foi comparado o desempenho de diferentes arquiteturas de CNN. Os autores utilizaram técnicas como validação cruzada K-Fold [Kohavi 1995] e aumento de dados. O modelo Inception-v3 destacou-se com um desempenho superior, alcançando uma acurácia de 97,5%. Foi ressaltada a importância do aumento de dados para a generalização do modelo. Entretanto, o estudo foca exclusivamente na classificação binária das imagens de psoríase e pele saudável, o que simplifica o problema e favorece resultados otimistas. Além disso, compara três modelos clássicos de CNN — ResNet50 [He et al. 2016], Inception-v3 [Szegedy et al. 2016] e VGG19 [Simonyan 2014] — que, embora amplamente utilizados, são relativamente antigos.

[Zhao et al. 2020] utilizou CNNs para a identificação de psoríase a partir de um conjunto de dados estruturado de imagens clínicas, incluindo doenças de pele visualmente similares, como líquen plano e eczema. O modelo Inception-v3 obteve o melhor desempenho, com uma AUC de $0,981 \pm 0,015$, superando outros modelos como Xception [Chollet 2017] e DenseNet121 [Huang et al. 2017]. O sistema de IA demonstrou maior precisão e menor taxa de erros em comparação com dermatologistas humanos. No entanto, devido aos dados não estarem disponíveis publicamente, a comunidade científica pode enfrentar limitações para reproduzir o trabalho. Além disso, foi mencionado o uso de técnicas de aumento de dados, mas os autores não esclareceram se essas técnicas foram devidamente restritas ao conjunto de treinamento.

[Mohan et al. 2024] explorou o uso de modelos baseados em *Transformers*, incluindo ViTs e *Swin Transformers* [Liu et al. 2021], para a classificação de doenças de pele, incluindo psoríase. Foi utilizado um conjunto de dados ampliado de 31 classes e alcançou uma precisão de teste de 96,48% com o modelo DinoV2 [Oquab et al. 2023],

superando *benchmarks* anteriores. A pesquisa concluiu que os *Transformers* têm potencial para revolucionar o diagnóstico de doenças de pele. Os autores não explicam se o conjunto de dados foi ampliado ou refinado. Apesar de o artigo apontar que os modelos ViT do experimento obtiveram um maior desempenho em relação à CNN EfficientNetB2, não fica claro se o conjunto de dados utilizado contém imagens de doenças visualmente similares à psoríase. Neste trabalho, o conjunto de dados contempla imagens de psoríase e de doenças similares, o que aumenta o desafio da classificação.

3. Materiais e Métodos

Esse trabalho utiliza modelos baseados na técnica de Aprendizado Supervisionado para detectar a incidência de psoríase. Para o treinamento dos modelos, fez-se o uso de um servidor composto de duas GPUs NVIDIA GeForce GTX 1080 Ti, cada uma com 11 GB de VRAM, um processador Intel(R) Core(TM) i7-8700 CPU @ 3,20 GHz, 56 GB de memória principal e 2,4 TB de memória secundária.

3.1. Conjunto de Dados

A obtenção de imagens médicas representou um desafio na construção da base de dados devido a questões éticas e à necessidade de preservar a privacidade dos pacientes. O acesso a imagens dermatológicas requer autorizações explícitas, o que dificulta a coleta e o compartilhamento. Por isso, o conjunto de imagens, de diferentes áreas do corpo, foi coletado a partir de repositórios de domínio público [DermNetNZ 2024, DermIS 2022, Dermatoweb 2002, Atlas Dermatologico 2024, DanderM 2017, Hellenic Dermatological Atlas 2011], cujos dados de imagem podem ser utilizados para fins educativos e acadêmicos. Essas imagens possuem etiquetas de diagnóstico, validadas por dermatologistas certificados.

A coleta de imagens abrangeu lesões rotuladas como psoríase, dermatite, líquen plano e pitiríase rosada, resultando inicialmente em um total de 3.248 imagens. Imagens de áreas específicas, como unhas e couro cabeludo, foram desconsideradas por não apresentarem características visuais relevantes. Além disso, imagens desfocadas ou sem foco foram removidas, reduzindo o conjunto final para 2.781 imagens.

O conjunto de imagens de pele saudável, também de diversas áreas do corpo humano, foi coletado de dados NTU [Huynh et al. 2014] do *Biometrics and Forensics Lab*. O total de imagens coletadas dessa fonte foi uma quantidade sucinta de 20 imagens. O restante das imagens de pele saudável foi obtido do Google Images®, resultando em um total de 1.176 imagens de pele saudável.

Devido à coleta de dados de múltiplas fontes, um processo de refinamento foi necessário para eliminar imagens duplicadas. Um *script* em Python, utilizando o algoritmo *dHash*, foi implementado para detectar e remover duplicatas. O procedimento identificou e excluiu redundâncias, garantindo um conjunto de dados único e mais representativo.

Após todas as etapas de refinamento, o conjunto final de dados totalizou 3.843 imagens distribuídas entre as classes de psoríase, dermatite, líquen plano, pitiríase rosada e pele saudável, conforme apresenta o gráfico da Figura 1.

3.2. Particionamento do Conjunto de Dados

O particionamento foi realizado em duas etapas. Na primeira, o conjunto de dados foi dividido de modo que 80% dos dados foram destinados a participar da validação cruzada,

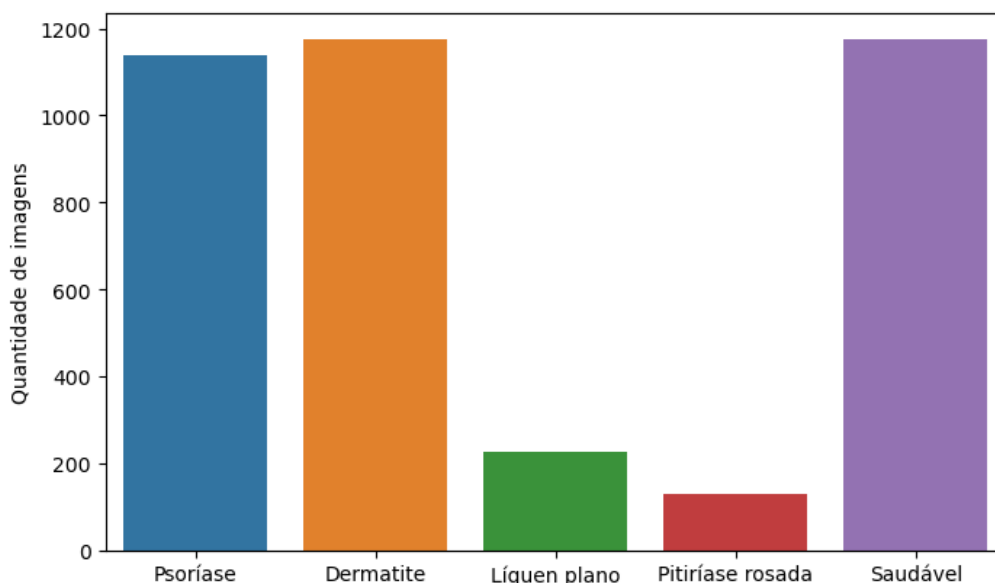


Figura 1. Distribuição do conjunto de dados por classe.

representando os dados de treino e validação. A outra parte, 20%, foi separada exclusivamente para compor os dados de teste, que não participam da validação cruzada e nem da etapa de treino. Nessa etapa foi empregada a técnica de Amostragem Estratificada [Meng 2013], que garante que os subgrupos estejam adequadamente representados no conjunto de treinamento.

Na segunda etapa, foi utilizada a validação cruzada *K-Fold* [Kohavi 1995] no subconjunto de treino e validação. Optou-se por utilizar cinco (05) *folds*, sendo quatro para treinamento e um exclusivo para validação. Durante cada iteração do processo de validação, um dos *folds* de treinamento é selecionado para validação, permitindo que o modelo seja avaliado em diferentes subconjuntos de dados. Esse método contribui para minimizar o risco de *overfitting* em apenas um conjunto específico e aumenta a confiabilidade e a precisão dos resultados obtidos [Géron 2019].

3.3. Pré-Processamento das Imagens

Todas as imagens tiveram suas intensidades de pixel normalizadas, de modo a colocá-las em uma escala entre 0 (zero) e 1 (um). Além disso, o conjunto de dados destinado ao treinamento passou por um processo de aumento artificial de dados. Essa técnica incluiu rotações incrementais das imagens em ângulos aleatórios entre 0° e 20°, além da aplicação de giros tanto horizontais quanto verticais. Por fim, as imagens foram redimensionadas para o tamanho especificado por cada modelo pré-treinado no ImageNet.

3.4. Modelos de CNN e ViT

Optou-se por utilizar a versão pré-treinada (a partir da base do ImageNet) de cada modelo, disponível nas bibliotecas TorchVision [Marcel and Rodriguez 2010] e Timm [Wightman 2019].

A tabela 1 apresenta o tamanho e a quantidade de parâmetros de cada modelo, além de suas respectivas acurácias *top-1* e *top-5* na base de dados do ImageNet.

Modelo	Categoria	Tamanho (MB)	Parâmetros	Acurácia <i>top-1</i> (%)	Acurácia <i>top-5</i> (%)
Inception-v3	CNN	104	27,2M	77,3%	93,5%
EfficientNetV2-L	CNN	455	118,5M	85,8%	97,8%
ConvNeXt-L	CNN	755	197,8M	84,4%	97,0%
ViT-L/16	ViT	1.161	304,3M	79,7%	94,6%
MaxViT-T	ViT	118	30,9M	83,7%	96,7%
DaViT-B	ViT	352	88,0M	84,6%	96,9%

Tabela 1. Modelos selecionados e suas respectivas acurácias no conjunto de dados ImageNet [Marcel and Rodriguez 2010, Wightman 2019].

Os modelos foram selecionados com base em suas características específicas, buscando avaliar os resultados obtidos em diferentes configurações e levando em conta seu desempenho na competição ImageNet. Um fator comum entre eles é a ampla disponibilidade em bibliotecas de aprendizado profundo, como TorchVision e Timm, o que facilita sua implementação.

3.5. Transferência de Aprendizado

Treinar uma rede neural profunda completa para uma tarefa específica demanda elevados recursos computacionais e um grande volume de dados. Por isso, é comum recorrer a modelos pré-treinados, que podem ser utilizados como ponto de partida ou como extratores de características para o problema em questão.

A estratégia definida para transferência de aprendizado foi a de utilizar um modelo pré-treinado como extrator de características, devido às suas vantagens significativas de eficiência, simplicidade e redução do risco de *overfitting*.

Essa abordagem consiste em empregar um modelo pré-treinado, geralmente desenvolvido com base em um conjunto de dados extenso e genérico, como, por exemplo, o ImageNet, para obter as características das novas imagens. Posteriormente, essas características extraídas são utilizadas para treinar um modelo adaptado a um conjunto de dados menor e mais específico. Nesse processo, a última camada do modelo é substituída por uma camada densa com o número de unidades correspondente ao total de classes do novo conjunto de dados [Milani et al. 2023].

3.6. Configurações do Treinamento

Durante o treinamento dos modelos, foi definido um critério para acompanhar as métricas de perda (*loss*) e acurácia (*accuracy*). Esse critério de otimização foi definido para o caso de o modelo apresentar melhorias na taxa de aprendizado (*learning rate*). A seguir, serão apresentadas as configurações utilizadas em todos os treinamentos realizados.

- Otimizador: utilizou-se o otimizador *AdaMax*, que é uma versão do Adam [Kingma and Ba 2014] mais robusta para grandes gradientes e valores de parâmetros elevados.
- *ReduceLROnPlateau*: essa função disponível na biblioteca PyTorch implementa a técnica de redução da taxa de aprendizado mediante platô. A métrica utilizada para identificar o platô da taxa de aprendizagem é a perda (*loss*) do conjunto de

validação. A redução aplicada tem um fator de $1e^{-3}$, e a paciência (*patience*) foi definida como 3, o que significa que o modelo espera por três épocas sem melhorias na métrica antes de diminuir a taxa de aprendizagem usando o fator.

- *Early Stopping*: a acurácia do conjunto de validação é a métrica utilizada para avaliar o desempenho do modelo, e a paciência foi definida como 7 épocas.
- *Checkpoint*: utilizou-se as funções *save* e *load_state_dict* para salvar e carregar, respectivamente, os parâmetros do modelo.
- *Batch size*: utilizou-se um *batch* de tamanho 32 levando em consideração os recursos computacionais disponíveis e o tempo de treinamento.
- Épocas: Para esse experimento, definiu-se 50 épocas.

3.7. Métricas de Avaliação

A matriz de confusão é representada por uma matriz quadrada, utilizada para comparar as previsões do modelo com os valores reais (verdadeiros). Os valores localizados na diagonal principal correspondem às previsões corretas, enquanto as demais posições representam os erros do modelo. Ela apresenta quatro categorias de rótulos:

- Verdadeiro Negativo (TN): quantidade de amostras corretamente classificadas como pertencentes à classe negativa.
- Verdadeiro Positivo (TP): quantidade de amostras corretamente classificadas como pertencentes à classe positiva.
- Falso Negativo (FN): número de amostras da classe positiva incorretamente classificadas como negativas.
- Falso Positivo (FP): número de amostras da classe negativa incorretamente classificadas como positivas.

Com isso, a partir dos parâmetros que compõem a matriz de confusão, é possível obter cada métrica de avaliação. Neste trabalho foram utilizadas métricas ponderadas.

As métricas ponderadas calculam as métricas individuais para cada classe, mas ponderam essas métricas pelo número de amostras de cada classe antes de calcular a média final. Desse modo, as métricas ponderadas conseguem equilibrar a análise de classes minoritárias e majoritárias, sendo úteis para conjuntos de dados desbalanceados onde é desejável refletir o impacto real das classes no desempenho geral [Luque et al. 2019].

Nas fórmulas de cada métrica a seguir, a variável N representa o número total de classes:

- Acurácia: mede a proporção de previsões corretas em relação ao total de amostras avaliadas, conforme a equação 1.

$$\text{Acurácia} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (1)$$

- Precisão: é a métrica utilizada para representar quantos casos de verdadeiros positivos são identificados corretamente pelo modelo, conforme a equação 2.

$$\text{Precisão} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i} \quad (2)$$

- Revocação (*recall*): também chamada de *sensibilidade*, é utilizada para representar as amostras positivas que foram corretamente rotuladas, conforme a equação 3.

$$\text{Revocação} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i} \quad (3)$$

- *F1-score*: é calculado a partir da média harmônica da precisão e sensibilidade, conforme a equação 4. O *f1-score* é especialmente útil em contextos onde há um desbalanceamento entre as classes, pois penaliza modelos que apresentam alta precisão, mas baixa revocação, ou vice-versa.

$$F1\text{-score} = \frac{1}{N} \sum_{i=1}^N \frac{2 \times \text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (4)$$

4. Resultados e Discussão

A Tabela 2 apresenta as métricas de cada um dos seis modelos selecionados. Cada métrica foi calculada a partir da média ponderada considerando os resultados da validação cruzada K-fold.

Tabela 2. Métricas gerais em termos de média ponderada.

Modelo	Parâmetros	Acurácia	Precisão	Revocação	F1-score
Inception-v3	27,2M	94,5%	94,6%	94,5%	94,5%
EfficientNetV2-L	118,5M	95,4%	95,5%	95,4%	95,4%
ConvNeXt-L	197,8M	95,6%	95,6%	95,6%	95,5%
ViT-L/16	304,3M	88,5%	88,0%	88,5%	87,8%
MaxViT-T	30,9M	96,0%	96,1%	96,0%	96,0%
DaViT-B	88,0M	96,4%	96,4%	96,4%	96,4%

Dentre os modelos de CNN, ConvNeXt-L e EfficientNetV2-L alcançaram um *f1-score* de 95,4% e 95,5%, respectivamente. No entanto, esse desempenho foi obtido a custo de uma quantidade de parâmetros significativamente elevada de ambos os modelos (118,5 milhões e 197,8 milhões, respectivamente). Uma vez que a quantidade de parâmetros de uma rede neural pode influenciar diretamente no seu custo computacional [LeCun et al. 1989], essa quantidade elevada de parâmetros pode indicar uma maior limitação quanto à sua implementação prática em ambientes com recursos limitados. Por outro lado, o modelo Inception-v3 se apresenta como uma alternativa mais econômica em termos de recursos computacionais, dado que conta com apenas 27,2 milhões de parâmetros, ao custo de um desempenho de predição ligeiramente inferior, mas ainda notável, com um *f1-score* de 94,5%.

Dentre os modelos de ViT, MaxViT-T e DaViT-B alcançaram as métricas mais altas (acurácia, precisão, revocação e *f1-score*) ao custo de quantidades de parâmetros relativamente baixas (30,9 milhões e 88,0 milhões, respectivamente). O modelo ViT-L/16,

apesar de sua grande contagem de parâmetros (304,3 milhões), obteve o pior desempenho em comparação com os outros modelos, alcançando um *f1-score* inferior, de 87,8%. Esse desempenho pode ser justificado pelo fato de o modelo não ter sido bem ajustado para o conjunto de dados do ImageNet, conforme aponta [Tan and Le 2021].

É possível observar o desempenho superior dos modelos DaViT-B e MaxViT-T para a tarefa proposta, principalmente por ambos terem alcançado métricas de predição excepcionais, mesmo sendo modelos relativamente leves. Entretanto, apesar de ser ligeiramente mais pesado que o MaxViT-T, o DaViT-B acertou 97% das classificações de psoríase, conforme apontado em sua matriz de confusão. Esse fator, somado a sua conquista de maior *f1-score* (96,4%) dentre os modelos, consolida o DaViT-B como o modelo com maior potencial para atuar como ferramenta de auxílio no diagnóstico da psoríase.

Os dados da matriz de confusão do Davit-B (Figura 2) indicam que a principal dificuldade de classificação para o modelo residiu na semelhança visual dessas condições com a psoríase e, especialmente, com a dermatite. O modelo classificou corretamente 84% das imagens de líquen plano, sendo os principais erros a confusão com dermatite (11%) e psoríase (4,9%); outras doenças, como a pitiríase rosada (6,9% dos seus casos), também foram erroneamente classificadas como líquen plano. Para a pitiríase rosada, a taxa de acerto foi de 80%, com confusões primárias com dermatite (11%) e líquen plano (6,9%).

Finalmente, evidencia-se que a dermatite constitui o principal fator de confusão para ambas as patologias em questão. Ademais, uma dificuldade notável foi observada na diferenciação mútua entre líquen plano e pitiríase rosada. Esses erros podem ser justificados pelo nível notável de similaridade visual das lesões cutâneas, que impõe limitações ao desempenho diagnóstico do DaViT-B.

5. Conclusões

O experimento aponta que ambas as categorias de modelos (CNN e ViT) apresentam um grande potencial para a resolução da tarefa proposta, com todos os modelos, com exceção do ViT-L/16, alcançando um *f1-score* superior a 94%. Em particular, foi observado que DaViT-B melhor eficiência na detecção da psoríase, sendo o modelo selecionado como o mais indicado para a tarefa.

A implementação de modelos como o DaViT-B em ambientes clínicos para auxiliar dermatologistas na detecção da psoríase é promissora, uma vez que o desempenho robusto do modelo em diferenciar psoríase de outras condições de pele visualmente similares demonstra seu potencial para reduzir diagnósticos equivocados. Isso favorece a detecção mais refinada da doença, para que tratamentos mais precisos e eficazes sejam prescritos.

A alta similaridade visual das lesões cutâneas pode justificar a principal confusão na classificação de líquen plano e pitiríase rosada pelo modelo DaViT-B, que ocorre com a dermatite e entre si. Entretanto, é possível que haja mais fatores que corroborem com essa confusão e, por isso, reconhece-se a necessidade de investigações futuras mais aprofundadas para elucidar os fatores que contribuem para a essas confusões e para o desenvolvimento de estratégias mitigatórias, como o enriquecimento do conjunto de dados.

Os trabalhos futuros identificados durante a execução do trabalho são:

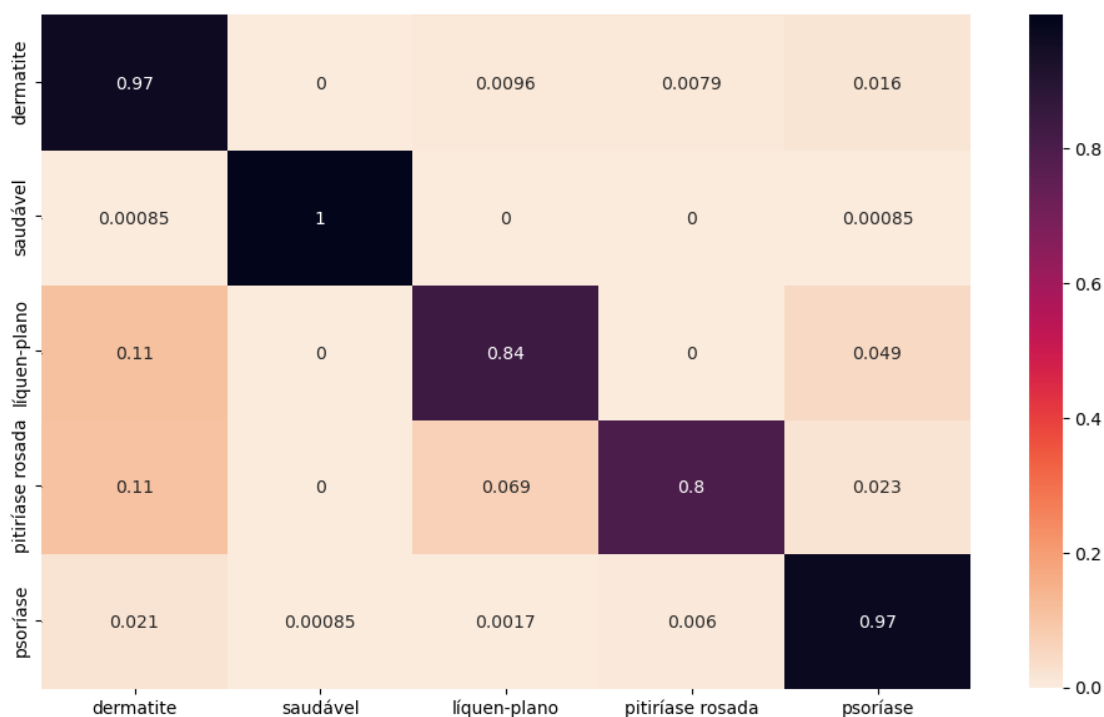


Figura 2. Matriz de Confusão do modelo DaViT-B.

- Expandir o conjunto de dados de imagens, incluindo imagens de um leque maior de doenças de pele que se assemelhem à psoríase, como, por exemplo, o melanoma e outros tipos de câncer de pele;
- Realizar uma análise investigativa mais aprofundada sobre os erros de predição apresentados por todos os modelos para as classes Líquen Plano, Pitíriase Rosada e Dermatite do conjunto de dados;
- Explorar mais a estratégia de transferência de aprendizado do ajuste fino do modelo pré-treinado, citado na Seção 3.5;
- Investigar a utilização do otimizador *Nesterov-accelerated Adaptive Moment Estimation* (Nadam) [Dozat 2016];
- Definir e coletar métricas de avaliação sobre o custo computacional de cada modelo, por exemplo a métrica de predições por segundo ou do tempo de uma predição;
- Verificar o desempenho, com ajuste fino, de modelos multimodais populares e modernos, como, por exemplo, Llama 3.2 [Dubey et al. 2024], Gemini 1.5 [Team et al. 2024] e GPT-4o [Hurst et al. 2024];
- Desenvolver um sistema que utilize o DaViT-B para o auxílio ao diagnóstico da psoríase e coletar métricas de teste desse sistema em ambientes de uso prático.

Referências

- [Alexey 2020] Alexey, D. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv: 2010.11929*.
- [Atlas Dermatologico 2024] Atlas Dermatologico (2024). Dermatology atlas. <https://atlasdermatologico.com.br/>. Acesso em: 19/04/2024.

- [Bin Ji 2022] Bin Ji, Yiyi Wang, D. Z. (2022). Automatic detection and evaluation of nail psoriasis based on deep learning: a preliminary application and exploration. *SPIE International Conference on Computer Application and Information Security*.
- [Brasil 2020] Brasil, A. (2020). Estudo mostra que mais de 90% da população desconhecem a psoríase. <https://agenciabrasil.ebc.com.br/saude/noticia/2020-11/estudo-mostra-que-mais-de-90-da-populacao-desconhecem-psoriedade>. Acesso em: 29/04/2024.
- [Chollet 2017] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.
- [DanderM 2017] DanderM (2017). Atlas of dermatology. <https://www.danderM.dk/atlas/>. Acesso em: 17/04/2024.
- [Dash et al. 2020] Dash, M. et al. (2020). A cascaded deep convolution neural network based cadx system for psoriasis lesion segmentation and severity assessment. *Applied Soft Computing*.
- [de Dermatologia 2023] de Dermatologia, S. B. (2023). O que é a psoríase? <https://www.sbd.org.br/psoriedade/>. Acesso em: 29/04/2024.
- [Deng et al. 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- [Dermatoweb 2002] Dermatoweb (2002). Web docente de dermatologia. <http://dermatoweb.udl.es/>. Acesso em: 22/04/2024.
- [DermIS 2022] DermIS (2022). Dermatology information system. <https://www.dermis.net/>. Acesso em: 26/04/2024.
- [DermNetNZ 2024] DermNetNZ (2024). The worlds leading free dermatology website. <https://dermnetnz.org/>. Acesso em: 29/04/2024.
- [Ding et al. 2022] Ding, M., Xiao, B., Codella, N., Luo, P., Wang, J., and Yuan, L. (2022). Davit: Dual attention vision transformers. In *European conference on computer vision*, pages 74–92. Springer.
- [Dozat 2016] Dozat, T. (2016). Incorporating nesterov momentum into adam. In *Proceedings of the 4th International Conference on Learning Representations, Workshop Track*, pages 1–4.
- [Dubey et al. 2024] Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- [Géron 2019] Géron, A. (2019). *Mãos À Obra: Aprendizado de Máquina com Scikit-Learn E TensorFlow*. Alta Books.
- [He et al. 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

- [Hellenic Dermatological Atlas 2011] Hellenic Dermatological Atlas (2011). For health professionals and public. <http://www.hellenicdermatlas.com/en/>. Acesso em: 16/04/2024.
- [Huang et al. 2017] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- [Hurst et al. 2024] Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., et al. (2024). Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.
- [Huynh et al. 2014] Huynh, N. Q., Xu, X., Kong, A. W. K., and Subbiah, S. (2014). A preliminary report on a full-body imaging system for effectively collecting and processing biometric traits of prisoners. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*.
- [Kingma and Ba 2014] Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.
- [Kohavi 1995] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. Morgan Kaufmann Publishers Inc.
- [LeCun et al. 1989] LeCun, Y., Denker, J., and Solla, S. (1989). Optimal brain damage. *Advances in neural information processing systems*, 2.
- [Liu et al. 2021] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022.
- [Liu et al. 2022] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986.
- [Luque et al. 2019] Luque, A., Carrasco, A., Martín, A., and de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91:216–231.
- [Maduranga and Nandasena 2022] Maduranga, P. and Nandasena, D. (2022). Mobile-based skin disease diagnosis system using convolutional neural networks (cnn). *International Journal of Image, Graphics and Signal Processing*, 14:47–57.
- [Marcel and Rodriguez 2010] Marcel, S. and Rodriguez, Y. (2010). Torchvision the machine-vision package of torch. <https://pytorch.org/vision/>.
- [Meng 2013] Meng, X. (2013). Scalable simple random sampling and stratified sampling. In Dasgupta, S. and McAllester, D., editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 531–539, Atlanta, Georgia, USA. PMLR.
- [Milani et al. 2023] Milani, A. et al. (2023). A deep learning application for psoriasis detection. *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 315–329.

- [Mohan et al. 2024] Mohan, J., Sivasubramanian, A., Sowmya, V., and Vinayakumar, R. (2024). Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable ai. *arXiv preprint arXiv:2407.14757*.
- [Olescki 2021] Olescki, G. (2021). Detecção de tromboembolia pulmonar utilizando redes neurais convolucionais e extração de características. *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 381–391.
- [Oquab et al. 2023] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. (2023). Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*.
- [Rodrigues and Teixeira 2009] Rodrigues, A. P. and Teixeira, R. M. (2009). Desvendando a psoríase. *RBAC*.
- [Roslan et al. 2020] Roslan, R., Razly, I., Sabri, B., and Ibrahim, Z. (2020). Evaluation of psoriasis skin disease classification using convolutional neural network. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 9:349.
- [Silva et al. 2022] Silva, G. et al. (2022). Cardiac arrhythmia detection in ecg signals using graph convolutional network. *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 25–35.
- [Simonyan 2014] Simonyan, K. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [Szegedy et al. 2016] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.
- [Tan and Le 2021] Tan, M. and Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR.
- [Team et al. 2024] Team, G., Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., Tanzer, G., Vincent, D., Pan, Z., Wang, S., et al. (2024). Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.
- [Tu et al. 2022] Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., and Li, Y. (2022). Maxvit: Multi-axis vision transformer. In *European conference on computer vision*, pages 459–479. Springer.
- [Wei et al. 2023] Wei, M., Wu, Q., Ji, H., Wang, J., Lyu, T., Liu, J., and Zhao, L. (2023). A skin disease classification model based on densenet and convnext fusion. *Electronics*, 12(2):438.
- [Wightman 2019] Wightman, R. (2019). Pytorch image models. <https://github.com/rwightman/pytorch-image-models>.
- [Zhao et al. 2020] Zhao, S., Xie, B., Li, Y., Zhao, X.-y., Kuang, Y., Su, J., He, X.-y., Wu, X., Fan, W., Huang, K., et al. (2020). Smart identification of psoriasis by images using convolutional neural networks: a case study in china. *Journal of the European Academy of Dermatology and Venereology*, 34(3):518–524.