

ANÁLISE COMPARATIVA DE MODELOS BASEADOS EM DEEP LEARNING NA CLASSIFICAÇÃO DE ACNE E ROSÁCEA

Guilherme Yugo Ishii¹
Paloma Passerani Moreira¹
João Ricardo Favan²
Eloíza Martins Primo Capeloci²

RESUMO

As dermatoses faciais acne e rosácea estão entre as condições mais frequentes na dermatologia, sendo associadas a impactos negativos na qualidade de vida. Por apresentarem sinais visuais semelhantes, sua diferenciação pode ser imprecisa, o que favorece a autoavaliação inadequada. Este trabalho teve como objetivo comparar duas abordagens de aprendizado profundo, utilizando as arquiteturas ResNet50 e InceptionV3 para automatizar a classificação entre acne, rosácea e pele normal, avaliando seu desempenho atuando de forma integral e como extratoras de características somadas aos classificadores SVM, MLP e regressão logística. O conjunto de dados foi composto por 1.002 imagens distribuídas entre as três classes, pré-processadas e divididas em treino, validação e teste. Os oito modelos foram treinados com aprendizado por transferência e avaliados em 10 execuções com sementes aleatórias, utilizando as métricas macro de *F1-score*, precisão e *recall*, bem como o teste de Friedman com *post-hoc* de Nemenyi e análise dos tempos computacionais. Todos os modelos alcançaram *F1-score* superior a 0,9680, com a ResNet50 apresentando o maior valor de 0,9883. As abordagens híbridas com SVM e regressão logística foram estatisticamente equivalentes às CNNs completas, porém com tempo de treinamento reduzido. Conclui-se que o uso de arquiteturas de aprendizado profundo apresenta alto potencial para o diagnóstico diferencial de dermatoses faciais, podendo atuar como ferramenta de apoio na área dermatológica e contribuindo para avaliações mais eficientes.

Palavras-chave: Aprendizado profundo; Redes neurais convolucionais; Algoritmos classificadores; Dermatoses faciais.

1 INTRODUÇÃO

No cenário dermatológico, a acne e a rosácea se destacam pela elevada frequência de casos diagnosticados. Segundo um estudo epidemiológico global realizado com mais de 50 mil indivíduos em 20 países, a prevalência da acne atinge

¹Discente em Big Data no Agronegócio na FATEC Pompeia, Pompeia-SP,

²Docentes do curso Big Data no Agronegócio, FATEC Pompeia, Pompeia-SP.

cerca de 20,5% da população, com a América Latina apresentando uma das maiores taxas do mundo. De forma paralela, a rosácea afeta aproximadamente 5,1% (Saurat *et al.*, 2024). No contexto brasileiro, dados do Inquérito Epidemiológico da Sociedade Brasileira de Dermatologia (SBD, 2025) ressaltam a acne como a razão mais frequente em atendimentos dermatológicos na rede privada, representando 9,5% do total de casos registrados. Em relação à rosácea, embora não existam estudos populacionais amplos em todo o território nacional, uma pesquisa realizada em 13 ambulatórios de dermatologia no Brasil avaliou uma prevalência de 13% entre pacientes adultos (Bonamigo *et al.*, 2025).

Segundo Rocha *et al.* (2024), a acne é classificada como uma doença imunoinflamatória crônica da unidade pilossebácea, conforme a Classificação Internacional de Doenças (CID-11) e manifesta-se pela presença de comedões, pápulas e pústulas, enquanto a rosácea é definida como uma dermatose, ou seja, uma doença de pele, de caráter inflamatório crônico que acomete principalmente as regiões centrais da face, caracterizando-se por manchas vermelhas persistentes no rosto, denominadas eritema facial (Bonamigo *et al.*, 2025).

Ambas as dermatoses exercem um impacto negativo na qualidade de vida, sendo associadas à ansiedade, redução da autoestima, prejuízos nas relações sociais e depressão (Rocha *et al.*, 2024; Bonamigo *et al.*, 2025). Um grande desafio reside no fato de que elas compartilham sinais visuais semelhantes, levando muitos indivíduos ao autodiagnóstico e ao uso inadequado de produtos cosméticos (Gerges; Shih; Azar, 2022). Diante disso, torna-se necessário e benéfico o desenvolvimento de métodos automatizados para realizar essa classificação com precisão.

A inteligência artificial é a área abrangente voltada ao desenvolvimento de sistemas que simulam funções da inteligência humana, sendo o aprendizado de máquina (*machine learning*) uma de suas subáreas, que permite que os sistemas aprendam sozinhos a partir de dados (Paiva, 2025). Dentro do aprendizado de máquina, o aprendizado profundo (*deep learning*) utiliza redes neurais artificiais para identificar padrões complexos em diferentes níveis de complexidade (Carrera; Almeida; Oliveira, 2024).

Na dermatologia, conforme abordado por Cachinski *et al.* (2025), o uso de inteligência artificial tem se estabelecido como uma tecnologia transformadora por meio de algoritmos de aprendizado profundo, proporcionando análises rápidas e grande assertividade na identificação de padrões visuais, com precisão equivalente ou superior à de especialistas no diagnóstico a partir de imagens clínicas (Silva *et al.*, 2025).

Nesse cenário, o grande destaque são as *Convolutional Neural Networks* (CNNs) ou redes neurais convolucionais, que se inspiram no comportamento dos neurônios e operam aplicando filtros a partir da camada convolucional, isto é, uma matriz inicializada aleatoriamente para extrair as características da imagem. Na sequência, a camada de *pooling* (camada de agrupamento) reduz o tamanho dessas informações, diminuindo a quantidade de pesos. Por fim, as camadas densas, também conhecidas como camadas totalmente conectadas, são responsáveis por identificar os padrões relevantes (Nóbrega, 2023).

Entre as arquiteturas de CNNs mais utilizadas, a ResNet50 e InceptionV3 são consideradas referência em classificação de imagens. A ResNet50 é composta por 50 camadas e baseada em blocos residuais, possuindo como principal diferencial o uso de conexões de atalho entre as camadas, o que contorna o problema do desvanecimento do gradiente ao permitir que ele flua de forma direta, tornando o treinamento de redes profundas mais eficaz e estável (Maeda-Gutiérrez *et al.*, 2025). Por sua vez, a InceptionV3, com 48 camadas, é organizada em blocos que realizam várias operações em paralelo, sendo conhecida, conforme Pan *et al.* (2023), pela capacidade de lidar com imagens de diferentes tamanhos e resoluções. O objetivo dessas operações é justamente capturar características em várias escalas de forma simultânea e, ao fim dos blocos, combinar seus resultados.

Essas redes viabilizam o aprendizado por transferência por meio do *ImageNet*, pois como apontado por Ikechukwu *et al.* (2021), treinar uma rede neural do zero exigiria muito tempo computacional e uma grande quantidade de dados de treinamento rotulados. Além disso, Jeong, Park e Henao (2023) ampliam que modelos que fazem uso de aprendizado por transferência podem alcançar e até ultrapassar o desempenho de especialistas no diagnóstico de condições de pele.

A partir do pré-treino com o *ImageNet*, as CNNs podem ser adaptadas ao objetivo por meio do *fine-tuning*, processo que ajusta os pesos da rede para refinar o aprendizado na nova tarefa. Nessa aplicação, as arquiteturas atuam de forma integral, extraindo características visuais e classificando as imagens. Para isso, as camadas convolucionais da CNN pré-treinada são mantidas e suas camadas finais substituídas por camadas voltadas ao problema, incluindo camadas densas com função de ativação *ReLU* (*Rectified Linear Unit*), que introduz não-linearidade para o aprendizado de padrões complexos, e uma camada de saída *softmax*, que converte os resultados em probabilidades para cada classe. Comumente, para contornar o *overfitting* (sobreajuste), utiliza-se o *Dropout*, que desativa parte dos neurônios aleatoriamente e favorece uma generalização mais robusta. No treinamento, pode-se utilizar o otimizador Adam, que combina técnicas de momento e taxa de aprendizado adaptativa para atualizar os pesos da rede e minimizar erros.

Por outro lado, estudos recentes apontam vantagens de se utilizar abordagens híbridas, nas quais as CNNs atuam apenas como extratoras de características, enquanto algoritmos de *machine learning*, como o *Multilayer Perceptron* (MLP), a regressão logística e o *Support Vector Machine* (SVM) assumem o papel de classificadores finais (Aziz *et al.*, 2023; Mehan, 2025; Pereira, 2025). O MLP é uma rede neural artificial composta por múltiplas camadas de neurônios interconectados, capaz de aprender representações complexas e não lineares. A regressão logística, por sua vez, é um algoritmo linear que calcula a probabilidade de uma amostra pertencer a uma determinada classe, transformando os resultados numéricos em probabilidades por meio da função sigmoide. Somando-se a eles, o SVM destaca-se por traçar hiperplanos que maximizam a distância entre as classes, mesmo em espaços de alta dimensionalidade.

Nessa abordagem, como o foco é a extração de características provenientes das camadas convolucionais das CNNs pré-treinadas, as demais camadas são descartadas. Dessa forma, essas redes não passam por reajuste de pesos, somente extraem as características e as atribuem aos algoritmos de aprendizado de máquina, que se encarregam então da etapa de classificação, distinguindo as classes.

Diante desse panorama, o presente trabalho tem como objetivo desenvolver e comparar duas abordagens baseadas em *deep learning* para o diagnóstico

diferencial entre acne, rosácea e pele normal, visando o potencial de aplicação em sistemas computacionais de apoio à dermatologia. A finalidade é avaliar o desempenho das arquiteturas ResNet50 e InceptionV3 operando de forma completa como classificadoras, em contraste com o desempenho dessas redes atuando apenas como extratoras de características, integradas aos algoritmos de classificação SVM, MLP e regressão logística. Busca-se, assim, identificar qual modelo oferece a solução mais assertiva entre desempenho e eficiência para a classificação dessas dermatoses.

2 MATERIAL E MÉTODOS

Este estudo utilizou imagens faciais obtidas em conjuntos de dados, de acne e rosácea, complementadas por imagens de rostos sem lesões, que funcionaram como classe de controle para que os modelos aprendessem a distinguir as dermatoses de pele normal. Por conta das variações de condições de iluminação e enquadramento, com fundos e outras partes do corpo aparecendo, os dados passaram por etapas de pré-processamento, seleção e padronização.

Para a composição do conjunto de dados próprio, foi definido o uso de quatro *datasets*, sendo três retirados da plataforma *Kaggle* (2026) e um do banco SCIN (*Skin Condition Image Network*), que fornece imagens de capturas reais, e possui colaboração de médicos da *Stanford Medicine*.

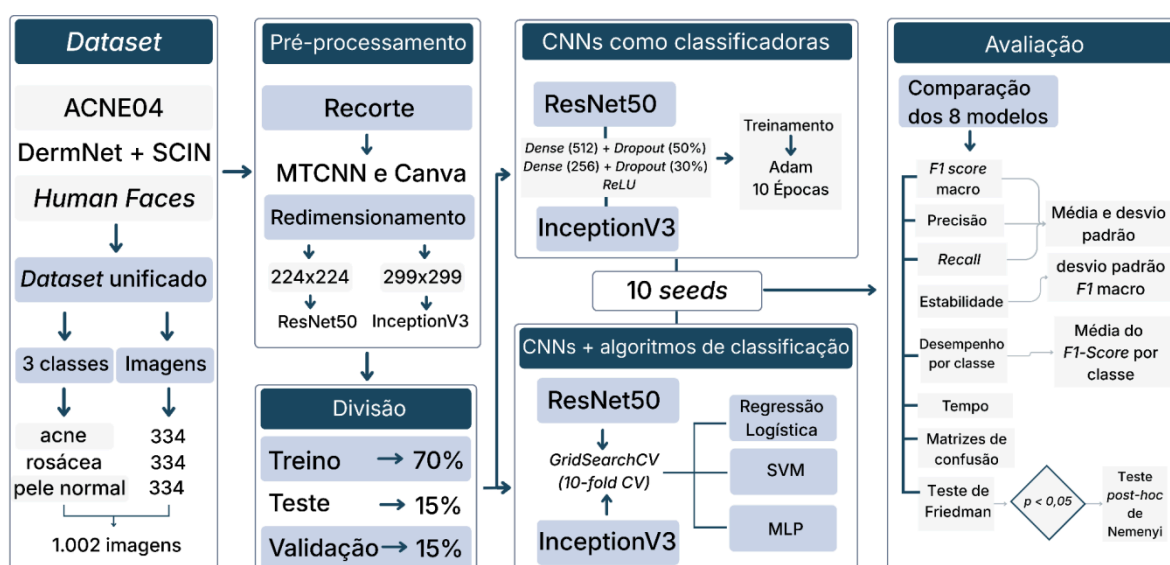
O desenvolvimento da classe de acne fez uso do *dataset* ACNE04 (*Kaggle*), composto por 1.457 imagens de indivíduos asiáticos distribuídas em quatro graus de severidade, da leve para a mais severa. Desse conjunto, foram selecionadas cerca de 80 imagens do grau 0, 94 do grau 1, 82 do grau 2, e 78 do grau 3, obtendo 334 imagens que foram unidas para compor o rótulo único 'acne'.

Com o objetivo de se obter um *dataset* balanceado, as classes foram limitadas ao número de imagens da menor classe, rosácea, que totalizou 334 amostras. Essas imagens foram extraídas do repositório DermNet (*Kaggle*), onde estavam agrupadas na categoria "Acne e Rosácea", e do banco de imagens SCIN. As imagens de acne provenientes do DermNet foram descartadas, pois não apresentavam a mesma padronização das imagens do conjunto ACNE04.

A composição da classe de controle correspondente à pele normal fez uso do conjunto *Human Faces* (Kaggle), um *dataset* estruturado com diversidade de etnias, faixas etárias e perfis demográficos, do qual foram selecionadas as devidas 334 imagens. Sendo assim, o conjunto totalizou 1.002 imagens igualmente distribuídas entre as três classes.

A organização das etapas, desde a obtenção dos dados até a avaliação dos modelos, está ilustrada no fluxograma metodológico do projeto (Figura 1).

Figura 1 – Fluxograma metodológico do projeto



Fonte: Elaborada pelos autores (2026)

O pré-processamento teve como objetivo eliminar o excesso de elementos de fundo, como paredes, roupas e cabelos, que introduzem ruído ao aprendizado das redes neurais. Para isso, foi utilizada a arquitetura MTCNN (*Multi-task Cascaded Convolutional Networks*), que conforme sugerido por Chandra *et al.* (2025), evidencia as características do rosto, permitindo que o foco de detecção e recorte seja apenas na região facial.

Desse modo, a rede foi configurada com limiar de confiança de 85% e margem de 20% ao redor da *bounding box* (caixa delimitadora), garantindo a inclusão das regiões centrais do rosto, por conta da natureza das dermatoses. As imagens que não foram detectadas pela rede passaram por recorte e remoção do fundo com auxílio da plataforma Canva (Canva, 2026).

Após o pré-processamento, o conjunto de 1.002 imagens foi dividido de forma estratificada em três partições, utilizando a biblioteca *Scikit-learn* (Pedregosa *et al.*, 2011), sendo 70% para treinamento, 15% para validação e 15% para teste. O processamento foi realizado no Google Colab (Google, 2026), ambiente de desenvolvimento em nuvem, utilizando a linguagem *Python* 3.12.13 (Python, 2026). As configurações do ambiente consistiram em 12.7 GB de memória RAM, uma GPU Tesla T4 com 14.6 GB VRAM, e 112.6 GB de armazenamento de disco.

Os rótulos das classes foram convertidos de forma automática em valores inteiros pelo *ImageFolder*, da biblioteca *Torchvision* (Torchvision, 2016), que organiza as classes a partir dos nomes das pastas do *dataset*. As imagens foram normalizadas com os parâmetros do *ImageNet* e padronizadas em tamanho por meio da função *Resize* da mesma biblioteca, sendo redimensionadas para 224×224 *pixels* para a ResNet50 e para 299×299 *pixels* para a InceptionV3, seguindo o padrão de cada rede.

O uso das CNNs como classificadoras envolveu a substituição das camadas finais por uma estrutura voltada para a classificação de acne, rosácea e pele normal. Essa estrutura foi composta por duas camadas densas, uma com 512 neurônios e outra com 256, com função de ativação *ReLU*, seguidas respectivamente de camadas de *Dropout* com taxas de 50% e 30% (Pereira, 2025). A camada de saída foi constituída por três neurônios, um para cada classe, porém no *framework PyTorch* (Paszke *et al.*, 2019), a ativação *softmax* dessa camada fica incorporada internamente pela função de perda *CrossEntropyLoss* durante o treinamento, o qual foi conduzido por 10 épocas com o otimizador Adam.

Na abordagem híbrida, as camadas convolucionais da ResNet50 e da InceptionV3 foram totalmente congeladas, atuando somente como extratoras de características. A camada de saída foi substituída por uma operação de identidade, atribuída a uma função do *PyTorch* (Paszke *et al.*, 2019), de forma que a saída da rede correspondesse a um vetor contendo as características extraídas das imagens. Como essas características possuem escalas diferentes, o vetor foi normalizado por meio do *StandardScaler* da biblioteca *Scikit-Learn* (Pedregosa *et al.*, 2011), e fornecido aos classificadores SVM, MLP e regressão logística.

Como forma de explorar combinações de parâmetros para identificar a configuração de melhor desempenho, os hiperparâmetros de cada classificador foram otimizados a partir do *GridSearchCV* com validação cruzada de 10 *folds* (Putra; Pradipta; Ayu, 2025). Nesse processo, os classificadores foram treinados com as características extraídas das CNNs, utilizando a biblioteca *Scikit-learn* (Pedregosa *et al.*, 2011).

Para avaliar a consistência dos modelos diante de diferentes divisões dos dados, os experimentos foram repetidos com múltiplas sementes aleatórias (*random seeds*), conforme recomendado por Picard (2021). Foram utilizadas 10 *seeds*, configuradas pelas bibliotecas *PyTorch* (Paszke *et al.*, 2019) e *NumPy* (Harris *et al.*, 2020), gerando 10 execuções para cada modelo. A partir delas, foram conduzidas análises de desempenho, estatísticas e computacionais.

A avaliação de cada modelo foi realizada sobre o conjunto de teste, tendo sido adotado o *F1-score* macro como métrica principal por calcular o *F1-score* de cada classe de forma individual e realizar a média entre elas, permitindo analisar o comportamento dos modelos em cada classe. Além disso, calculou-se a média e o desvio padrão dessa métrica junto às métricas macro de precisão e *recall*, bem como o *F1-score* médio por classe ao longo das 10 execuções.

O teste de Friedman ($\alpha = 0,05$) foi aplicado para comparar os 8 modelos sobre as mesmas execuções. Esse teste retorna um *p*-valor, que representa a probabilidade das diferenças observadas serem resultado do acaso. Assim, como o *p*-valor obtido foi inferior a 0,05, rejeitou-se a hipótese nula de equivalência entre os modelos, concluindo que havia diferenças estatisticamente significativas, o que levou à aplicação do teste *post-hoc* de Nemenyi para identificar quais pares diferiam entre si.

Por fim, para analisar os padrões de erro, foram geradas matrizes de confusão a partir da execução em que o *F1-score* ficou mais próximo da mediana das 10 execuções, retratando o desempenho mais comum do modelo. De forma complementar, os tempos de extração de características, treinamento e inferência foram medidos para avaliar a viabilidade computacional de cada abordagem.

3 RESULTADOS E DISCUSSÃO

A avaliação inicial concentrou-se na análise do desempenho das duas abordagens utilizadas, ou seja, as CNNs ResNet50 e InceptionV3 atuando de forma integral, como extratoras de características e classificadoras, e a abordagem híbrida, em que elas atuaram apenas como extratoras somadas aos classificadores SVM, MLP e regressão logística. A Tabela 1 mostra o desempenho médio dos 8 modelos, ordenados de forma decrescente pelo *F1-score* macro, juntamente com precisão, *recall* e respectivos desvios padrão.

Tabela 1 – Desempenho médio dos oito modelos

Modelo	<i>F1-Score</i>	Precisão	<i>Recall</i>
ResNet50	0,9883 ± 0,0081	0,9887 ± 0,0076	0,9883 ± 0,0082
InceptionV3	0,9851 ± 0,0126	0,9857 ± 0,0121	0,9850 ± 0,0127
InceptionV3 + SVM	0,9812 ± 0,0073	0,9819 ± 0,0070	0,9811 ± 0,0073
ResNet50 + Regressão Logística	0,9811 ± 0,0118	0,9815 ± 0,0116	0,9810 ± 0,0117
ResNet50 + SVM	0,9792 ± 0,0096	0,9800 ± 0,0096	0,9789 ± 0,0095
InceptionV3 + Regressão Logística	0,9760 ± 0,0109	0,9766 ± 0,0107	0,9759 ± 0,0107
ResNet50 + MLP	0,9752 ± 0,0145	0,9760 ± 0,0136	0,9751 ± 0,0145
InceptionV3 + MLP	0,9680 ± 0,0107	0,9684 ± 0,0108	0,9682 ± 0,0106

Fonte: Elaborada pelos autores (2026)

Observa-se que todos os modelos obtiveram desempenho elevado, alcançando um *F1-score* entre 0,9680 e 0,9883, uma diferença de apenas 2,03 pontos percentuais entre o desempenho mínimo e máximo. O resultado superior foi atingido pela ResNet50, com *F1-score* de 0,9883, desvio padrão de 0,0081, precisão de 0,9887 e *recall* de 0,9883. Esses valores próximos entre todas as métricas indicam uma classificação altamente equilibrada entre as três classes.

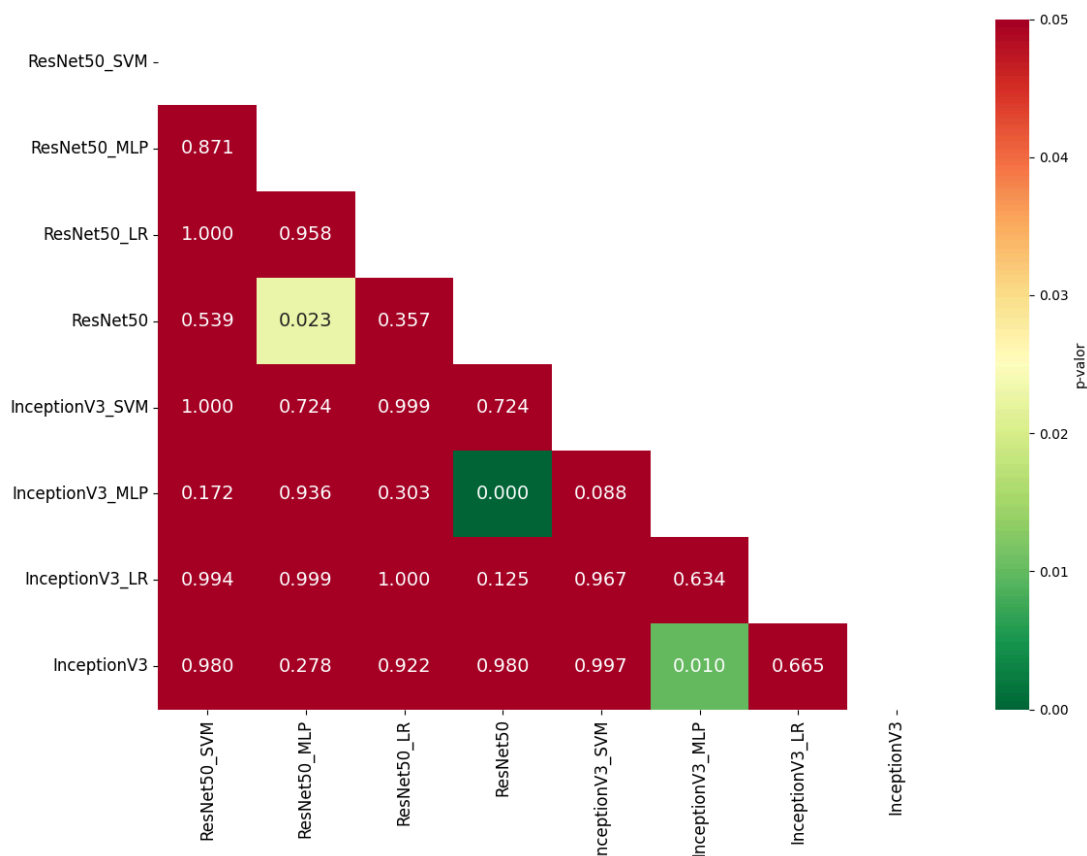
Na segunda colocação, encontra-se a InceptionV3, a qual alcançou *F1-score* de 0,9851 e desvio padrão de 0,0126. Como afirmam Hanif, Zim e Kaur (2024), essas arquiteturas alcançaram o que é chamado de estado da arte na classificação de imagens, o que explica a capacidade de ambas em extrair características discriminativas.

Em relação às CNNs combinadas aos classificadores, o SVM e a regressão logística apresentaram desempenhos muito similares aos das CNNs operando de forma integral. A InceptionV3 com o classificador SVM atingiu $F1$ -score de 0,9812 e desvio padrão de 0,0073, enquanto a ResNet50 com a regressão logística obteve $F1$ -score de 0,9811 e desvio padrão de 0,0118. Essa análise indica que as CNNs pré-treinadas no *ImageNet* foram capazes de extrair características visuais robustas o suficiente para que classificadores lineares como SVM e regressão logística separassem as classes de forma eficiente, o que tornou as fronteiras entre as classes mais claras no espaço de representação, dispensando o ajuste completo da rede para obter desempenho equivalente.

As CNNs somadas ao MLP apresentaram os menores valores da análise, ainda que com diferença de poucos pontos percentuais. A ResNet50 com MLP alcançou $F1$ -score de 0,9752 e desvio padrão de 0,0145, ao passo que a InceptionV3 com MLP obteve $F1$ -score de 0,9680 e desvio padrão de 0,0107, configurando a combinação de menor desempenho entre as avaliadas.

Embora as métricas tenham apontado desempenhos numericamente próximos, o teste de Friedman foi aplicado com base no $F1$ -score macro das 10 execuções. O p -valor obtido foi de 0,0003, indicando diferenças estatisticamente significativas entre os modelos. Logo, o teste *post-hoc* de Nemenyi foi realizado para identificar quais pares diferiram entre si, cujos resultados dos p -valores foram ilustrados na Figura 2.

Figura 2 - Resultados dos p -valores do teste *post-hoc* de Nemenyi



Fonte: Elaborada pelos autores (2026)

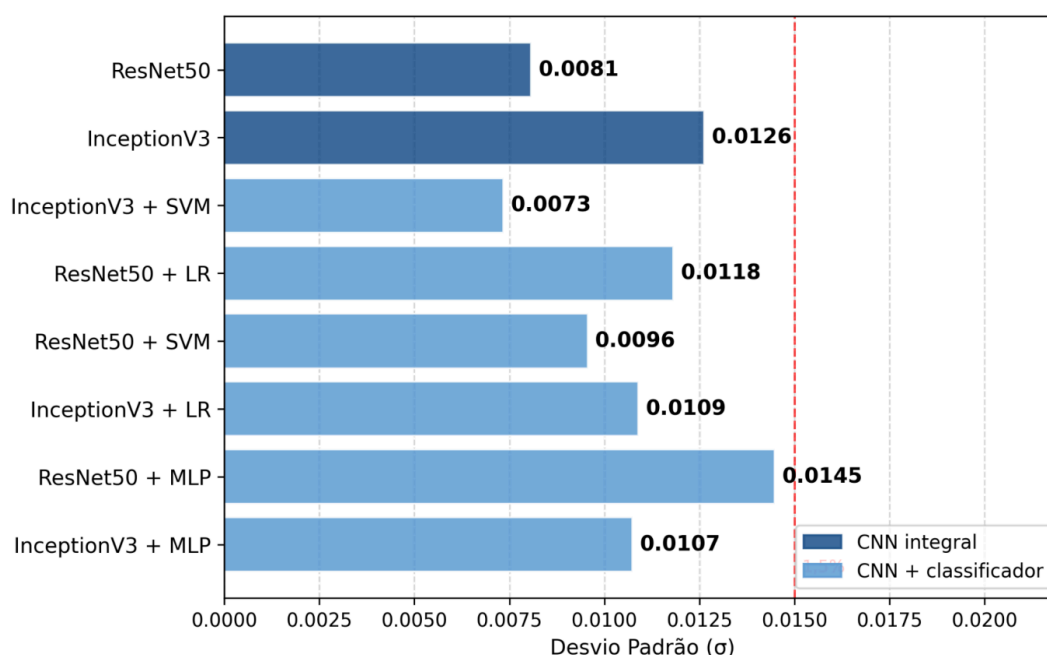
A análise do *heatmap* mostra que, apesar do teste de Friedman apontar diferença estatística entre os modelos, a maior parte dos pares apresenta p -valores muito superiores a 0,05, indicando ausência de diferença significativa entre eles. Esse resultado corrobora o que já havia sido observado na análise de desempenho, onde as CNNs operando de forma integral e as CNNs combinadas ao SVM e à regressão logística se comportaram de maneira equivalente.

No entanto, três pares apresentaram diferença significativa, sendo a ResNet50 operando de forma integral e a InceptionV3 com MLP ($p < 0,001$), a ResNet50 operando de forma integral e a ResNet50 com MLP ($p = 0,023$), e por fim, a InceptionV3 operando de forma integral e a InceptionV3 com MLP ($p = 0,010$). Em todos os casos, a diferença envolveu uma CNN operando de forma integral e uma rede combinada ao classificador MLP, reforçando que o MLP foi a única combinação que se distanciou estatisticamente das demais. Esse resultado pode ser atribuído à conhecida dificuldade de MLPs em cenários de alta dimensionalidade e baixo

número de amostras (HDLSS), nos quais a suscetibilidade ao *overfitting* compromete a capacidade de generalização (Silva *et al.*, 2024).

A Figura 3 ilustra a estabilidade dos modelos por meio do desvio padrão. Pode-se observar que todos os modelos apresentaram valores abaixo de 1,5%, indicando que os resultados se mantêm estáveis independentemente da divisão dos dados entre treino, teste e validação.

Figura 3 - Estabilidade dos modelos por meio do desvio padrão

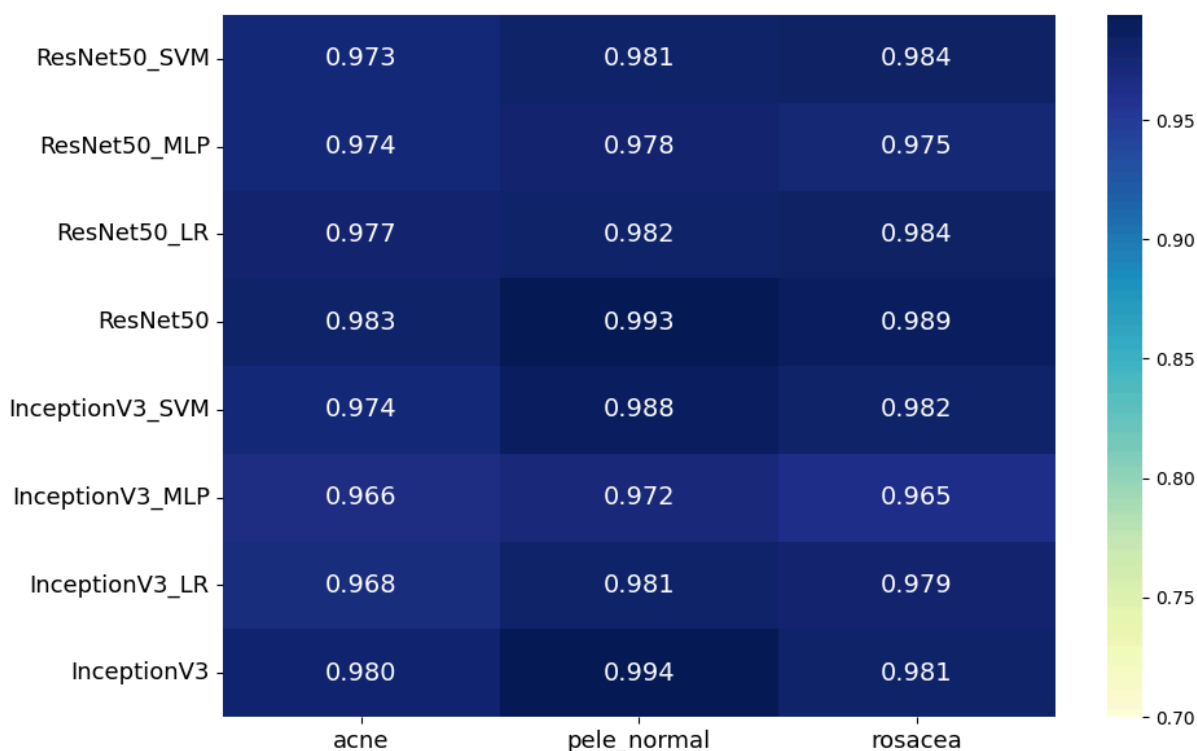


Fonte: Elaborada pelos autores (2026)

Dentre os modelos avaliados, três apresentaram desvio padrão inferior a 1%: a ResNet50 integral (*fine-tuning*), a ResNet50 combinada ao SVM e a InceptionV3 também combinada ao SVM. A presença recorrente do SVM entre os modelos mais estáveis pode ser atribuída ao seu princípio de maximização da margem entre classes, que tende a produzir fronteiras de decisão pouco sensíveis a variações nas divisões dos dados (Hao *et al.*, 2024). Ademais, segundo Lopez-Martinez *et al.* (2022), a instabilidade é um fator frequentemente negligenciado, mas que deve ser considerado antes da implantação de modelos em contextos clínicos. Assim, a estabilidade observada nessas configurações reforça sua relevância para aplicações na área da saúde.

A Figura 4 demonstra o desempenho de cada modelo por classe, considerando o $F1$ -score médio. Verifica-se que a ResNet50 e InceptionV3 operando de forma integral apresentaram os maiores valores de $F1$ -score nas três classes simultaneamente, com destaque para a classificação da classe pele normal.

Figura 4 - Desempenho de cada modelo por classe

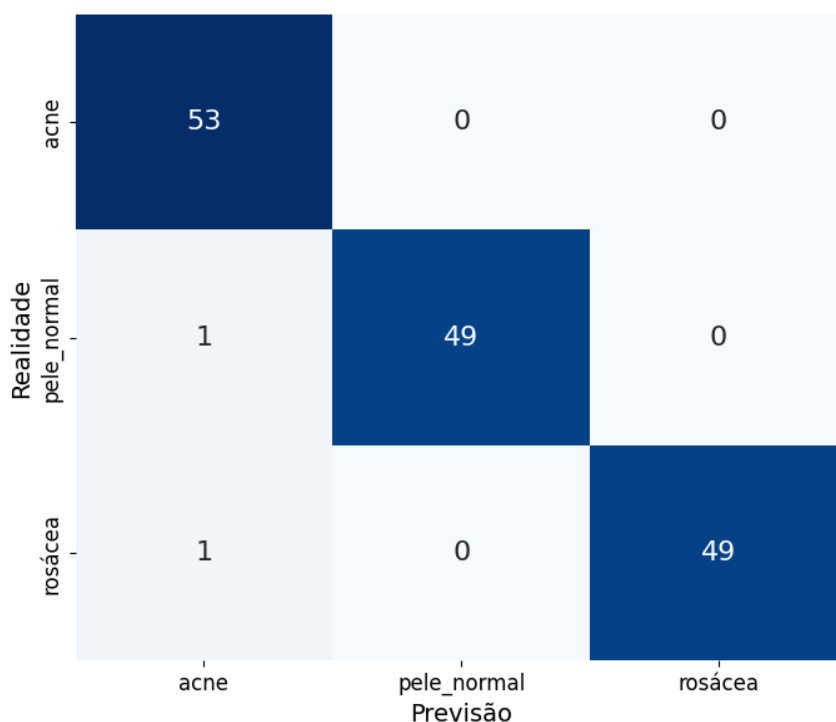


Fonte: Elaborada pelos autores (2026)

A classe acne mostrou ser a mais difícil de classificar, com $F1$ -score menor na maioria dos modelos, o que pode ser explicado pela grande variedade de formas das lesões de acne. Já a classe pele normal obteve valores maiores de classificação, pois a ausência de lesões torna seu padrão visual mais uniforme. Ademais, a arquitetura InceptionV3 junto ao classificador MLP alcançou os menores valores entre as três classes. Considerando que as diferenças entre as classes variam de 0,001 a 0,03, o que se pode concluir é apenas uma tendência de desempenho.

A Figura 5 aprofunda essa análise no nível das imagens, apresentando a matriz de confusão da ResNet50, que obteve o maior *F1-score*, enquanto a Figura 6 ilustra a matriz de confusão da InceptionV3 combinada ao MLP, que registrou o menor *F1-score*. Conforme o teste de Nemenyi, esses dois modelos também indicaram diferença estatística entre si.

Figura 5 - Matriz de confusão - ResNet50



Fonte: Elaborada pelos autores (2026)

Ao analisar os padrões de erro da ResNet50 operando de forma integral, percebe-se que houve apenas dois erros de previsão, sendo uma imagem de pele normal e uma de rosácea classificadas de forma incorreta como acne.

De modo semelhante, nota-se que a InceptionV3 combinada ao classificador MLP resultou em 3 erros, com uma imagem de acne classificada como pele normal, uma de pele normal classificada como acne e uma de rosácea classificada como acne. Em uma aplicação prática, classificar uma imagem com lesão, no caso acne, como pele normal pode ser considerado um erro crítico, pois significa que o modelo deixou de reconhecer a condição que deveria detectar. No entanto, como houve apenas um caso desse tipo, observa-se que ambas as abordagens foram capazes

de identificar a presença das dermatoses, assim como sua ausência, de forma consistente.

Figura 6 - Matriz de confusão - InceptionV3 + MLP

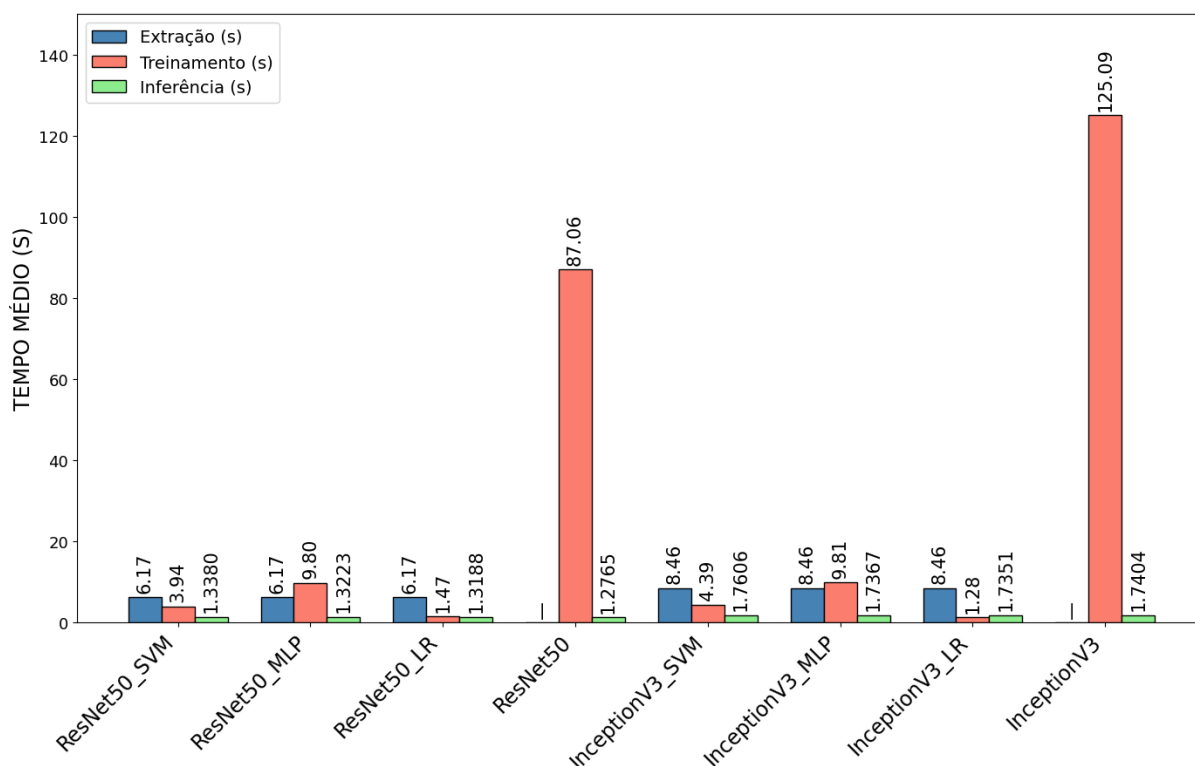
Realidade	acne	52	1	0
pele_normal	1	49	0	
rosácea	1	0	49	
	Previsão	acne	pele_normal	rosácea

Fonte: Elaborada pelos autores (2026)

Esses resultados se comparam ao observado por Nóbrega (2023), onde as confusões se concentraram entre a classe acne e pele normal, o qual é relacionado à variabilidade visual das lesões, que em casos mais leves podem se aproximar da pele normal. Somado a isso, os erros de classificação de rosácea como acne podem ser explicados pela similaridade visual entre elas (Gerges; Shih; Azar, 2022).

A Figura 7 apresenta os tempos médios das três etapas: extração de características, treinamento e inferência para cada uma das oito configurações avaliadas.

Figura 7 - Tempos médios de extração, treinamento e inferência por modelo



Fonte: Elaborada pelos autores (2026)

A ResNet50 mostrou-se mais rápida na extração, com tempo médio de 6,17 segundos, enquanto a InceptionV3 atingiu um tempo médio de 8,46 segundos, aproximadamente 37% de diferença, o que pode ser atribuído à maior complexidade arquitetural da InceptionV3, composta por módulos paralelos que aumentam o número de operações por imagem.

A maior diferença entre as abordagens se concentra na etapa de treinamento, visto que os classificadores de *machine learning* finalizaram o processo em poucos segundos, com a regressão logística levando apenas 1,28 a 1,47 segundos, o SVM entre 3,94 e 4,39 segundos e a MLP entre 9,80 e 9,81 segundos. Esse desempenho era esperado, pois segundo Thippanna e Kumar (2025), a regressão logística é reconhecida por sua simplicidade e interpretabilidade. A extração proveniente das CNNs e repassada para algoritmos de classificação corrobora o que afirmam Kim *et al.* (2022), em que o uso dessas arquiteturas como extratoras de características reduz o custo computacional em grande medida, sem degradar o poder preditivo.

Essa diferença ganha proporção significativa em cenários práticos, nos quais o sistema precisaria ser retreinado de forma periódica com volumes crescentes de imagens. Conforme Guimarães *et al.* (2023), em contextos de *big data*, modelos precisam ser frequentemente atualizados ou retreinados às custas de recursos computacionais significativos. De forma paralela, em aplicações de apoio ao diagnóstico em clínicas e hospitais, o crescimento contínuo da base de imagens exige atualizações constantes do sistema, tornando o custo computacional do treinamento um fator relevante na escolha da abordagem.

Na inferência, os tempos variaram entre 1,28 segundos para a ResNet50 operando de forma integral e 1,76 segundos para a InceptionV3 combinada ao SVM, revelando equivalência entre todos os modelos. O maior consumo de tempo está na passagem da imagem pelas camadas convolucionais da CNN, que ocorre nas duas abordagens, então a classificação final pelos algoritmos de *machine learning* é executada em tempo reduzido, sem gerar impactos no tempo total de inferência.

De modo geral, as oito configurações apresentaram desempenho elevado, em ambas as abordagens avaliadas, CNN atuando integralmente e CNN combinada aos classificadores, demonstrando o potencial de aplicação de técnicas de aprendizado profundo para auxiliar no diagnóstico diferencial entre acne, rosácea e pele normal.

CONSIDERAÇÕES FINAIS

As arquiteturas e combinações avaliadas apresentaram resultados significativos e estáveis, com equivalência estatística entre a maioria das abordagens, demonstrando viabilidade em sistemas de apoio ao diagnóstico dermatológico. As CNNs combinadas ao SVM e à regressão logística se destacaram como a alternativa mais equilibrada entre desempenho e eficiência computacional. Como limitações, ressaltam-se o *dataset* de tamanho moderado, resultante da união de diferentes bases, e o fato de as imagens de acne serem provenientes exclusivamente de indivíduos asiáticos, o que pode limitar a eficiência em outros fototipos de pele.

Para trabalhos futuros, propõe-se a inclusão de outras arquiteturas de CNN e classificadores na comparação, a ampliação do *dataset* com maior diversidade étnica para todas as classes, a incorporação dos graus de severidade da acne e de novas dermatoses, bem como o desenvolvimento de uma aplicação que classifique imagens enviadas pelos usuários.

REFERÊNCIAS

- AZIZ, M. T.; MAHMUD, S. M. H.; ELAHE, M. F.; JAHAN, H.; RAHMAN, M. H.; NANDI, D.; SMIRANI, L. K.; AHMED, K.; BUI, F. M.; MONI, M. A. **A Novel Hybrid Approach for Classifying Osteosarcoma Using Deep Feature Extraction and Multilayer Perceptron**. *Diagnostics*, v. 13, n. 12, p. 2106, 2023. Disponível em: <https://doi.org/10.3390/diagnostics13122106>. Acesso em: 18 mar. 2026.
- BONAMIGO, R. R. et al. **Clinical-demographic profile, aggravating factors, comorbidities, and quality of life in patients with Rosacea: a Brazilian multicenter study (GBPER: Brazilian Research and Studies Group on Rosacea)**. *Anais Brasileiros de Dermatologia*, v. 100, n. 5, p. 501160, 2025. Disponível em: <https://doi.org/10.1016/j.abd.2025.501160>. Acesso em: 07 fev. 2026.
- CACHINSKI, E. J.; SANTOS, A. H. L.; KIEUTIKA, I. F.; TANAKA, S. S. K.; THOMÉ, J. T. **Inteligência artificial em dermatologia: um panorama das aplicações tecnológicas e seus avanços**. *Brazilian Journal of Implantology and Health Sciences*, v. 7, n. 3, p. 318–329, 2025. Disponível em: <https://doi.org/10.36557/2674-8169.2025v7n3p318-329>. Acesso em: 06 fev. 2026.
- CANVA PTY LTD. **Canva**: Ferramenta de design gráfico online. Sydney: Canva, 2026. Disponível em: <https://www.canva.com>. Acesso em: 22 abr. 2026.
- CARRERA, Augusto Esteves; ALMEIDA, Gabriel Marques G.; OLIVEIRA, Ivan Carlos Alcântara de. **Estudo comparativo de modelos de Deep Learning para classificação de lesões cutâneas**. 2024. Artigo (Faculdade de Computação e Informática) – Universidade Presbiteriana Mackenzie, São Paulo, 2024. Disponível em: <https://dspace.mackenzie.br/items/f464b82c-3d57-4276-aa17-39e389b839f3>. Acesso em: 08 fev. 2026.
- CHANDRA, F. R. et al. **Analysis of the Use of MTCNN and Landmark Technology to Improve the Accuracy of Facial Recognition on Official Documents**. *Journal of Applied Informatics and Computing*, v. 9, n. 1, p. 16–22, 2025. DOI: 10.30871/jaic.v9i1.8814. Disponível em: <https://doi.org/10.30871/jaic.v9i1.8814>. Acesso em: 12 abr. 2026.
- DERMNET. **Dermnet**. Kaggle. Disponível em: <https://www.kaggle.com/datasets/shubhamgoel27/dermnet>. Acesso em: 11 fev. 2026.
- GERGES, F.; SHIH, F.; AZAR, D. **Automated Diagnosis of Acne and Rosacea using Convolution Neural Networks**. In: Proceedings of the 2021 4th International

Conference on Artificial Intelligence and Pattern Recognition. New York: Association for Computing Machinery, 2022. p. 607–613. Disponível em: <https://doi.org/10.1145/3488933.3488993>. Acesso em: 08 fev. 2026.

GOOGLE INC. **Google Colaboratory**. Mountain View: Google, 2026. Disponível em: <https://colab.research.google.com/>. Acesso em: 22 abr. 2026.

GUPTA, A. **Human Faces**. Kaggle. Disponível em: <https://www.kaggle.com/datasets/ashwingupta3012/human-faces>. Acesso em: 11 fev. 2026.

GUIMARÃES, Miguel et al. Predicting Model Training Time to Optimize Distributed Machine Learning Applications. *Electronics*, v. 12, n. 4, p. 871, 2023. DOI: 10.3390/electronics12040871. Disponível em: <https://doi.org/10.3390/electronics12040871>. Acesso em: 22 maio 2026.

HANIF, M. A.; ZIM, K. I.; KAUR, H. **ResNet vs Inception-v3 vs SVM: A Comparative Study of Deep Learning Models for Image Classification of Plant Disease Detection**. [S. l.: s. n.], 2024. p. 6. Disponível em: <https://doi.org/10.1109/IATMSI60426.2024.10502832>. Acesso em: 21 mar. 2026.

HAO, Zhezheng; NIE, Feiping; WANG, Rong. Multi-class Support Vector Machine with Maximizing Minimum Margin. arXiv:2312.06578v4, 2024. DOI: 10.48550/arXiv.2312.06578. Disponível em: <https://arxiv.org/abs/2312.06578>. Acesso em: 23 abr. 2026.

HARRIS, Charles R. et al. **Array programming with NumPy**. *Nature*, v. 585, p. 357–362, 2020. DOI: 10.1038/s41586-020-2649-2. Disponível em: <https://www.nature.com/articles/s41586-020-2649-2>. Acesso em: 12 abr. 2026.

IKEKCHUKWU, A. Victor et al. **ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images**. *Global Transitions Proceedings*, v. 2, n. 2, p. 375–381, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666285X21000558>. Acesso em: 26 mar. 2026.

JEONG, H. K.; PARK, C.; HENAO, R.; KHETERPAL, M. **Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations**. *JID Innovations*, v. 3, n. 1, p. 100150, 2023. Disponível em: <https://doi.org/10.1016/j.xjidi.2022.100150>. Acesso em: 28 mar. 2026.

KAGGLE. **ACNE04**. Kaggle. Disponível em: <https://www.kaggle.com/datasets/hamzaamanat/acne04>. Acesso em: 11 fev. 2026.

KIM, Hee E. et al. Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*, v. 22, n. 1, p. 69, 2022. DOI: 10.1186/s12880-022-00793-7. Disponível em: <https://doi.org/10.1186/s12880-022-00793-7>. Acesso em: 21 abr. 2026.

LOPEZ-MARTINEZ, D. et al. **Instability in clinical risk stratification models using deep learning**. Proceedings of Machine Learning Research, v. 193, p. 552–565, 2022. Disponível em: <https://proceedings.mlr.press/v193/lopez-martinez22a/lopez-martinez22a.pdf>. Acesso em: 28 mar. 2026.

MAEDA-GUTIÉRREZ, V. et al. **AI-Driven Plant Health Assessment: A Comparative Analysis of Inception V3, ResNet-50 and ViT with SHAP for Accurate Disease Identification in Taro**. Agronomy, v. 15, n. 1, p. 77, 2025. DOI: 10.3390/agronomy15010077. Disponível em: <https://doi.org/10.3390/agronomy15010077>. Acesso em: 09 fev. 2026.

MEHAN, Vineet. **Deep Inception Based Hybrid Machine Learning Framework for Binary Classification of Brain Tumor MRI Scans**. Complex Systems Informatics and Modeling Quarterly, n. 44, p. 1–16, 2025. DOI: 10.7250/csimq.2025-44.01. Disponível em: <https://doi.org/10.7250/csimq.2025-44.01>. Acesso em: 18 mar. 2026.

NÓBREGA, P. H. de M. **Classificação de queixas dermatológicas usando redes neurais convolucionais**. 2023. Disponível em: <https://dspace.sti.ufcg.edu.br/handle/riufcg/30477>. Acesso em: 07 fev. 2026.

PAIVA, Hermenegildo Woropo Albino. **Inteligência Artificial e Aprendizado de Máquina na Sociedade Moderna**. Educação, Research, Religion (ERR), v. 10, n. 2, 2025. DOI: 10.56238/ERR01v10n2-008. Disponível em: <https://doi.org/10.56238/ERR01v10n2-008>. Acesso em: 08 abr. 2026.

PAN, Y. et al. **Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases**. Frontiers in Physiology, v. 14, 2023. Disponível em: <https://doi.org/10.3389/fphys.2023.1126780>. Acesso em: 21 mar. 2026.

PASZKE, A. et al. **PyTorch: An Imperative Style, High-Performance Deep Learning Library**. arXiv:1912.01703, 2019. DOI: 10.48550/arXiv.1912.01703. Disponível em: <https://arxiv.org/abs/1912.01703>. Acesso em: 20 abr. 2026.

PEDREGOSA, F. et al. **Scikit-learn: Machine Learning in Python**. Journal of Machine Learning Research, v. 12, p. 2825–2830, 2011. Disponível em: <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>. Acesso em: 22 abr. 2026.

PEREIRA, Leticia Torres. **Um estudo comparativo de modelos de aprendizado de máquina para classificação de flores apícolas: integrando extratores de texturas e classificadores**. 2025. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal de Campina Grande, Campina Grande, 2025. Disponível em: https://repositorio.ufc.br/bitstream/riufcg/79979/1/2025_tcc_ltpereira.pdf. Acesso em: 19 mar. 2026.

PICARD, David. **Torch.manual_seed(3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision**. arXiv:2109.08203,

2021. DOI: 10.48550/arXiv.2109.08203. Disponível em: <https://arxiv.org/abs/2109.08203>. Acesso em: 12 abr. 2026.

PUTRA, R. Adhitama; PRADIPTA, G. Angga; AYU, P. Desiana Wulaning. **Gallbladder Disease Classification from Ultrasound Images Using CNN Feature Extraction and Machine Learning Optimization**. Journal of Electronics, Electromedical Engineering, and Medical Informatics, v. 7, n. 4, p. 1089–1111, set. 2025. DOI: 10.35882/jeeemi.v7i4.1030. Disponível em: <https://doi.org/10.35882/jeeemi.v7i4.1030>. Acesso em: 10 abr. 2026.

PYTHON SOFTWARE FOUNDATION. **Python Language Reference**, version 3.x. Wilmington: Python Software Foundation. Disponível em: <https://www.python.org/>. Acesso em: 22 abr. 2026.

ROCHA, M.; BARNES, F.; CALDERÓN, J.; FIERRO-ARIAS, L.; GOMEZ, C. E. M.; MUNOZ, C.; JANNELL, O.; TROIELI, P. **Acne treatment challenges – Recommendations of Latin American expert consensus**. Anais Brasileiros de Dermatologia, v. 99, n. 3, p. 414–424, 2024. Disponível em: <https://doi.org/10.1016/j.abd.2023.09.001>. Acesso em: 07 fev. 2026.

SAURAT, J.-H.; HALIOUA, B.; BAISSAC, C.; CULLELL, N. P.; BEN HAYOUN, Y.; AROMAN, M. S.; TAIEB, C.; SKAYEM, C. **Epidemiology of acne and rosacea: A worldwide global study**. Journal of the American Academy of Dermatology, v. 90, n. 5, p. 1016–1018, 2024. Disponível em: <https://doi.org/10.1016/j.jaad.2023.12.038>.

SCIN. **Skin Condition Image Network (SCIN)**. Google Health. Disponível em: https://storage.googleapis.com/dx-scin-public-data/dataset/scin_cases.csv. Acesso em: 12 fev. 2026.

SILVA, D. L. F. da; RUBINHO, R.; DENOFRE, A.; AVILA, S.; MAGALHÃES, R. F. **Inteligência artificial em dermatoses inflamatórias: onde estamos e para onde vamos?** Anais Brasileiros de Dermatologia, v. 100, n. 5, 2025. Disponível em: <https://www.anaisdedermatologia.org.br/pt-pdf-S2666275225001237>. Acesso em: 02 mar. 2026.

SILVA, D. R. da; ALMEIDA FILHO, A. T. de. **Hybrid Penalized Regression–MLP Models for Outcome Prediction in HDLSS Health Data**. arXiv:2409.12345, 2024. DOI: 10.48550/arXiv.2409.12345. Disponível em: <https://arxiv.org/abs/2409.12345>. Acesso em: 24 maio 2026.

SOCIEDADE BRASILEIRA DE DERMATOLOGIA. **Sociedade Brasileira de Dermatologia divulga inquérito atualizado sobre doenças de pele no Brasil**. Rio de Janeiro: SBD, 26 jun. 2025. Disponível em: <https://www.sbd.org.br/sociedade-brasileira-de-dermatologia-divulga-inquerito-atualizado-sobre-doencas-de-pele-no-brasil/>. Acesso em: 07 fev. 2026.

THIPPANNA, G.; KUMAR, G. Logistic Regression for Multiclass Classification of Dermatology Disorders: A Comparison of One-vs-One and One-vs-Rest Approaches. International Journal of Darshan Institute on Engineering Research and Emerging

Technologies, v. 14, p. 51–54, 2025. Disponível em:
<https://doi.org/10.32692/IJDI-ERET/14.2.2025.2509>. Acesso em: 02 abr. 2026.

TORCHVISION MAINTAINERS AND CONTRIBUTORS. **TorchVision**: PyTorch's Computer Vision library. [S. l.]: GitHub, 2016. Software. Disponível em:
<https://github.com/pytorch/vision>. Acesso em: 24 maio 2026.