

DESENVOLVIMENTO DE UM SISTEMA BASEADO EM REDES NEURAIIS PARA O SUPORTE AO DIAGNÓSTICO DE CÂNCER DE PELE

Laura Ribeiro Bernardino – d2023009575@unifei.edu.br
Kevin Henrique Monfredini Silva – d2022006962@unifei.edu.br
Matheus Brendon Francisco – matheus_brendon@unifei.edu.br

Palavras-chave: Câncer de Pele; Redes Neurais Convolucionais; Aprendizado de Máquina; Imagens Dermatoscópicas; Diagnóstico Auxiliado por Computador.

1. INTRODUÇÃO

O câncer de pele é uma preocupação proeminente na saúde global. Em 2020, o Observatório Global do Câncer (GLOBOCAN) estimou 1,20 milhão de novos casos de câncer de pele não melanoma e 325.000 de melanoma, este último com alta mortalidade. No Brasil, o Instituto Nacional de Câncer (INCA) o aponta como o tipo mais incidente no país. O diagnóstico precoce é o fator determinante para a sobrevida, especialmente em casos de melanoma.

O padrão diagnóstico baseia-se na avaliação clínica e dermatoscópica, uma técnica que melhora a acurácia. Contudo, a interpretação das imagens é complexa e sujeita à variabilidade interobservador. Neste contexto, a Inteligência Artificial (IA), especificamente as Redes Neurais Convolucionais (CNNs), emergiu como uma tecnologia disruptiva, com estudos demonstrando desempenho comparável ao de dermatologistas.

Apesar dos avanços, o estado da arte revela lacunas metodológicas. Muitos estudos avaliam seus modelos apenas em conjuntos de validação, sem o rigor de um conjunto de teste independente, o que pode gerar estimativas otimistas de performance. Além disso, a natureza de "caixa-preta" (black box) da maioria dos modelos é um desafio, pois não há garantia de que aprendem características clinicamente relevantes, uma etapa essencial para a adoção clínica.

Este trabalho aborda diretamente estas lacunas ao detalhar o desenvolvimento de um sistema de classificação de lesões cutâneas (benignas vs. malignas). As contribuições se centram em: (1) um protocolo de validação robusto com avaliação final em um conjunto de teste independente; (2) o uso da arquitetura moderna EfficientNet-B3 com ponderação de classes para tratar desbalanceamento; e (3) a incorporação de Inteligência Artificial Explicável (XAI) via Grad-CAM para interpretabilidade visual.

2. DESENVOLVIMENTO

2.1 Referencial teórico

A aplicação de deep learning para Diagnóstico Auxiliado por Computador (CAD) em dermatologia busca superar os desafios da interpretação visual humana. A literatura recente consolidou o uso de transferência de aprendizado (transfer learning), onde arquiteturas de CNNs pré-treinadas em datasets massivos (como o ImageNet) são adaptadas para tarefas específicas. Arquiteturas como ResNet e InceptionV3 são comuns, mas modelos mais recentes como a família EfficientNet têm se destacado por otimizar a relação entre precisão e custo computacional.

Um desafio central no domínio médico é o desbalanceamento de classes (mais casos benignos que malignos). Abordagens tradicionais como subamostragem (undersampling) podem descartar informações valiosas. O referencial teórico aponta que técnicas de ponderação de classes na função de perda (como a CrossEntropyLoss ponderada) são mais eficazes, pois utilizam todos os dados, atribuindo maior penalidade a erros na classe minoritária.

Finalmente, para superar a barreira da "caixa-preta", o campo da Inteligência Artificial Explicável (XAI) tornou-se fundamental. Técnicas como o Grad-CAM (Gradient-weighted Class Activation Mapping) permitem a geração de mapas de calor que destacam quais regiões da imagem foram mais influentes para a decisão da rede. Isso não só aumenta a confiança no modelo, mas permite a análise qualitativa de seus acertos e erros, sendo crucial para a validação clínica.

2.2 Metodologia

A metodologia foi estruturada para garantir robustez, performance e interpretabilidade.

Conjunto de Dados: Utilizou-se o acervo público "Skin Cancer (ISIC Images)" da plataforma Kaggle, composto por 3.297 imagens dermatoscópicas (224x224 pixels), divididas em 1.800 benignas e 1.497 malignas. A classificação binária alinha-se ao objetivo de um sistema de triagem.

Protocolo de Validação: Para uma avaliação imparcial, o conjunto completo foi dividido de forma estratificada (preservando a proporção das classes) em três subconjuntos independentes: Treinamento (70% - 2.307 imagens), Validação (15% - 495 imagens) e Teste (15% - 495 imagens). O conjunto de teste foi usado apenas para a avaliação final do modelo.

Pré-processamento e Aumento de Dados: As imagens foram redimensionadas para 300x300 pixels (entrada da EfficientNet-B3). Foi aplicada uma robusta pipeline de aumento de dados (data augmentation) em tempo real no conjunto de treinamento, incluindo rotações aleatórias (+/- 20 graus), espelhamentos horizontal e vertical, e variação de cor (brilho, contraste, saturação).

Modelo e Treinamento: Adotou-se a transferência de aprendizado com a arquitetura EfficientNet-B3 pré-treinada na ImageNet. A camada classificadora final foi substituída por uma nova camada linear com duas saídas (benigna, maligna). Para tratar o desbalanceamento, aplicou-se a ponderação de classes (class weighting) na função de perda CrossEntropyLoss, calculando pesos inversamente proporcionais à frequência de cada classe. O treinamento usou o otimizador AdamW e callbacks de EarlyStopping (paciência de 15 épocas) e ReduceLROnPlateau, salvando sempre o modelo com melhor desempenho na validação.

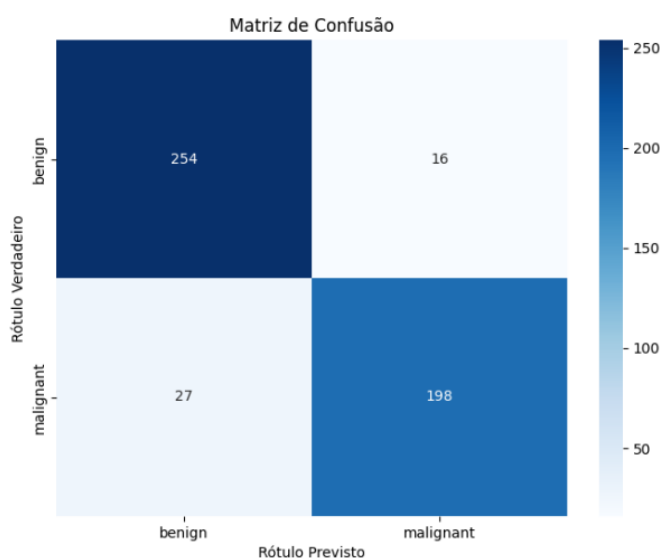
Avaliação e Interpretabilidade: A performance quantitativa (acurácia, precisão, sensibilidade, especificidade) foi aferida no conjunto de teste independente. Para a análise qualitativa e de interpretabilidade, foi implementada a técnica XAI Grad-CAM45.

3. RESULTADOS E DISCUSSÃO

A avaliação final do modelo otimizado foi realizada no conjunto de teste independente de 495 imagens (270 benignas, 225 malignas). O sistema alcançou uma acurácia geral de 91.31%.

A Matriz de Confusão (Figura 1) detalha o desempenho. O modelo classificou corretamente 254 das 270 lesões benignas (Verdadeiros Negativos) e 198 das 225 lesões malignas (Verdadeiros Positivos). Os erros foram baixos: 16 Falsos Positivos (lesões benignas classificadas como malignas) e 27 Falsos Negativos (lesões malignas não detectadas).

Figura 1 - Matriz de Confusão do modelo no conjunto de teste.



Fonte: Autoria própria

As métricas de classificação (figura 2) demonstram um excelente equilíbrio. Para a classe maligna, o modelo atingiu uma Sensibilidade (Revocação) de 88.00% e uma Especificidade de 94.07%. Isso indica que o sistema foi capaz de identificar corretamente 88% dos cânceres, ao mesmo tempo que descartou corretamente 94% das lesões benignas. O F1-Score, que balanceia precisão e revocação, ficou acima de 90% para ambas as classes.

Figura 2 - Resultados de Métricas de Classificação no Conjunto de Teste

Métrica	Classe Maligna	Classe Benigna	Média Ponderada
Precisão	0.9252 (92.52%)	0.9039 (90.39%)	0.9136 (91.36%)
Revocação	0.8800 (88.00%)	0.9407 (94.07%)	0.9131 (91.31%)
Especificidade	0.9407 (94.07%)	0.8800 (88.00%)	-
F1-Score	0.9021 (90.21%)	0.9220 (92.20%)	0.9129 (91.29%)

Fonte: Autoria própria

A acurácia de 91.31% é competitiva com trabalhos recentes, como Güler e Ağraz (2025) (90.17%). Mais importante, o equilíbrio entre 88.00% de sensibilidade e 94.07% de especificidade é um perfil clinicamente desejável, pois minimiza tanto os cânceres não detectados (falsos negativos) quanto as biópsias desnecessárias (falsos positivos).

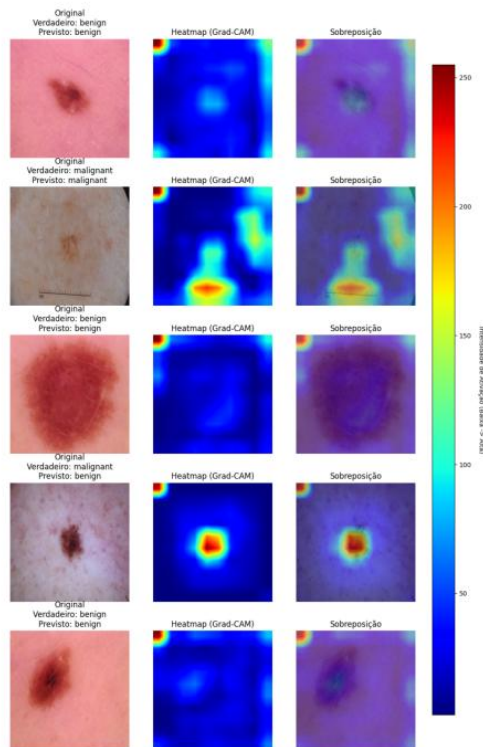
A análise qualitativa com Grad-CAM (Figura 3) forneceu insights sobre o comportamento do modelo. Nos casos de acerto (benignos e malignos), os mapas de calor mostraram que o modelo concentrou sua atenção sobre a área da lesão, e não em artefatos ou na pele saudável. Na análise crucial de um erro (um falso negativo), o Grad-CAM revelou que a rede localizou corretamente a lesão, mas as características visuais extraídas daquela região foram insuficientes para a classificação (provavelmente por ambiguidade), levando à predição incorreta. Isso seria impossível de saber apenas com métricas quantitativas e reforça a visão da IA como uma ferramenta de apoio, onde o julgamento clínico final permanece essencial.

Figura 3 - Análise com Grad-CAM para exemplos do conjunto de teste.



IEPG SUMMIT'25

Pensando o futuro com inteligência
artificial e consciência social



Fonte: Autoria própria

4. CONSIDERAÇÕES FINAIS

Este trabalho apresentou o desenvolvimento e a validação de um modelo de deep learning para suporte ao diagnóstico de câncer de pele. A principal contribuição foi a integração sinérgica de metodologias de ponta: a arquitetura EfficientNet-B3, um protocolo de validação rigoroso com teste independente, tratamento avançado de desbalanceamento de classes e, crucialmente, uma análise de interpretabilidade com Grad-CAM.

O sistema alcançou um desempenho de alta acurácia (91.31%) e, mais importante, um excelente equilíbrio entre sensibilidade (88.00%) e especificidade (94.07%). Metodologicamente, o estudo reforça a importância de protocolos de validação imparciais e da aplicação de XAI para uma análise profunda do comportamento do modelo. Na prática, os resultados demonstram o potencial significativo deste modelo como uma ferramenta de apoio à decisão clínica, capaz de auxiliar dermatologistas na triagem de lesões, otimizar fluxos de trabalho e contribuir para o diagnóstico precoce.

Apesar dos resultados, o estudo é limitado a um único dataset público. A pesquisa futura deve focar na validação em múltiplos datasets e em dados clínicos prospectivos para confirmar a generalização do modelo e pavimentar o caminho para sua translação segura para a prática médica.

AGRADECIMENTOS

Os autores desejam agradecer a Universidade Federal de Itajubá (UNIFEI) por toda a infraestrutura de pesquisa disponibilizada e as agências de fomentos, CAPES, CNPq e FAPEMIG pelo apoio financeiro na realização das atividades.

REFERÊNCIAS

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249-259.

Bennecer, I., Sili, M., Ben-Ayed, Y., Al-Maadeed, S., & El-Sabagh, A. (2023). Interpretable Skin Cancer Classification based on Incremental Domain Knowledge Learning. *IEEE Access*, 11, 21798-21811.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.

Alwakid, G., Gouda, W., Humayun, M., & Sama, N. U. (2022). Melanoma Detection Using Deep Learning-Based Classifications. *Healthcare (Basel)*, 10(12), 2481.

Gouda, W., Sama, N. U., Al-Waakid, G., Humayun, M., & Jhanjhi, N. Z. (2022). Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning. *Healthcare (Basel)*, 10(7), 1183.

Güler, I., & Ağraz, M. (2025). Investigation of Binary and Multiclass Classification Performance of Skin Cancer Images Using Transfer Learning Methods. *Journal of Clinical and Practical Research*, 47(3), 235-245.

Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). (2022). Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro: INCA.

Kittler, H., Marghoob, A. A., Argenziano, G., et al. (2016). Standardization of terminology in dermoscopy. *Journal of the American Academy of Dermatology*, 74(6), 1093-1106.

Rosendahl, C., Tschandl, P., Cameron, A., & Kittler, H. (2011). Diagnostic accuracy of dermoscopy for melanocytic and nonmelanocytic pigmented lesions. *Journal of the American Academy of Dermatology*, 64(6), 1068-1073.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 618-626).

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209-249.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML) (pp. 6105-6114).