

Um Modelo de Inteligência Artificial Para Detecção de Melanoma via Redes Neurais Convolucionais

Gabriel R. G. Pessanha¹, Eleanderson Campos¹

¹Centro de Inteligência Artificial (CIA) – Universidade Federal de Alfenas
Av. Celina Ferreira Ottoni, 4000 – Padre Vitor – Varginha – MG – Brasil

gabriel.pessanha@unifal-mg.edu.br, eleandersoncampos@gmail.com

Abstract. *We propose an artificial intelligence (AI) model based on convolutional neural networks (CNN) for the task of melanoma and nevus classification. Due to its high risk of causing metastasis, melanoma is the most lethal type of skin cancer. However, its early detection significantly improves the probability of a favorable prognostic. Given this scenario, we intend to study the performance of the proposed CNN on the early detection of melanoma. By obtaining reasonable results in terms of specificity ($0,96 \pm 0,03$) and sensibility ($0,81 \pm 0,02$), we show that the proposed CNN model can be a valuable clinical diagnosis tool.*

Resumo. *Neste artigo é proposto um modelo de inteligência artificial (IA) baseado em redes neurais convolucionais (RNC) para classificação de melanomas e nevos displásicos em imagens dermatoscópicas. Com um alto risco de provocar metástase, o melanoma figura como o mais grave tipo de neoplasia cutânea. No entanto, sua detecção precoce eleva consideravelmente a probabilidade de um prognóstico favorável, sendo essa a premissa que motiva este estudo. Com um desempenho satisfatório em termos de sensibilidade ($0,81 \pm 0,02$) e especificidade ($0,96 \pm 0,03$), concluímos que o modelo de RNC proposto pode ser uma ferramenta valiosa no apoio ao diagnóstico clínico de melanoma.*

1. Introdução

Graças a uma quantidade e variedade cada vez maiores de dados sendo coletados e armazenados digitalmente (Big Data) e melhoras sistemáticas de hardware nos computadores modernos, criou-se o ambiente perfeito para uma nova revolução nas áreas de gerenciamento e análise de dados (Murdoch e Detsky, 2013). Com essa rápida mudança de cenário surgem os novos desafios: Como transformar essa enorme quantidade de dados em informações e *insights* úteis nas diversas áreas do conhecimento?

Uma solução natural para o problema supracitado está no uso da inteligência artificial (IA), especificamente no uso dos modelos baseados em aprendizagem de máquina. Isso se deve a inerente capacidade desses modelos em lidar com grandes volumes de dados (estruturados ou não) e aprender padrões (Jordan e Mitchell, 2015). Na área médica, por exemplo, as redes neurais convolucionais (RNCs) têm se destacado pelo excelente desempenho como ferramentas de apoio ao diagnóstico clínico. Uma dessas aplicações envolve o uso desses modelos na classificação de imagens dermatológicas (e.g., lesão de pele benigna *versus* lesão maligna), que, de uma forma geral, têm demonstrado desempenho superior ao de dermatologistas na detecção de neoplasias cutâneas (Haenssle et al., 2018; Brinker et al., 2019).

A aplicação de técnicas de aprendizagem de máquina para detecção de neoplasias cutâneas é de enorme interesse, sobretudo como ferramentas para detecção precoce de melanomas. Embora não seja a neoplasia maligna que mais acomete a população brasileira, representando apenas 3% de todos os casos que afetam o órgão (neoplasias cutâneas malignas representam 30% do total de tumores malignos no país), o melanoma é uma das mais graves, dado seu alto risco de causar metástase (Instituto Nacional do Câncer, 2021). Apesar disso, uma detecção precoce desse tipo de câncer eleva consideravelmente as chances de um prognóstico favorável.

Buscando analisar o desempenho das RNCs na tarefa de detecção de melanoma, Brinker et al. (2019) e Haenssle et al. (2018) fizeram uso das arquiteturas ResNet50 e Inception v4, respectivamente, para classificar imagens dermatoscópicas de nevos displásicos e melanomas. Utilizando, respectivamente, um banco de dados com 12.738 imagens e outro com mais de 100.000, esses autores treinaram seus modelos de RNC e compararam seus desempenhos com os de dermatologistas. Os resultados de sensibilidade, especificidade e AUC (*Area under the ROC curve*) evidenciaram que, tanto em Brinker et al. (2019) quanto em Haenssle et al. (2018), os modelos de RNC obtiveram desempenho superior ao dos dermatologistas (em média).

Haja vista o enorme potencial das RNCs como ferramentas de apoio ao diagnóstico clínico, em especial na detecção precoce do melanoma, propõe-se, neste trabalho, o desenvolvimento de um modelo baseado em RNCs para detecção de melanoma em imagens dermatoscópicas.

Além da seção introdutória, este trabalho se divide nas seguintes partes: na seção 2 é apresentada a metodologia empregada e sequencialmente, na seção 3, a análise dos resultados obtidos. Finalmente, na seção 4, conclui-se este trabalho com uma discussão geral sobre a pesquisa e propostas de trabalhos futuros.

2. Metodologia

Os dados utilizados neste trabalho foram coletados do banco de dados de imagens dermatológicas do ISIC (International Skin Imaging Collaboration), e são compostos por uma amostra de 7.818 imagens dermatoscópicas referentes a melanomas e nevos displásicos (1:6) rotulados por biópsia (veja a Figura 1 para alguns exemplos). A etapa de tratamento e preparação dos dados consiste em redimensionar as imagens para o formato 225x225 *pixels* e convertê-las em *arrays* numéricos (RGB), bem como fazer a separação desses *arrays* para uma posterior validação cruzada 10-fold (para avaliação de performance em 10 etapas utilizando conjuntos disjuntos de 782 ou 781 imagens para validação e o restante para treino). Para o conjunto de teste, foram utilizadas as mesmas 100 imagens consideradas em Brinker et al. (2019). Todo o processo foi implementado via linguagem Python, e as bibliotecas utilizadas para redimensionar e converter as imagens foram, respectivamente, a PIL (função Image) e OpenCV (cv2.resize). Ademais, foi empregado um método de *Data Augmentation* via biblioteca Keras.

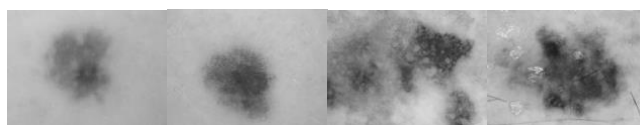


Figura 1. Imagens de nevos displásicos (duas primeiras) e melanomas (duas últimas) amostradas do banco de dados da ISIC.

Neste trabalho optou-se pela utilização da arquitetura da RNC *Inception-ResNet-v2* (Szegedy et al., 2017) como base para *transfer learning*. Essa arquitetura foi construída com base na família *Inception*, incorporando o conceito de conexões residuais da família *ResNet*. O Modelo resultante figura como estado-da-arte na classificação de imagens.

Visando atingir os objetivos deste trabalho, foi proposto um ajuste fino com base na arquitetura *Inception-ResNet-v2* para a tarefa de classificação de imagens dermatoscópicas de melanomas e nevos displásicos. Para isso, três camadas adicionais foram incorporadas ao modelo (sequencialmente às camadas do *Inception-ResNet-v2*), a saber: uma primeira de *Global Average Pooling 2D*, seguida por uma camada de *Dropout(0.5)* e outra de *Batch Normalization*. A camada de saída é composta por dois neurônios (classes melanoma e não melanoma) com função de ativação do tipo *Softmax*. A ideia por trás da utilização das camadas de *Global Average Pooling 2D* e *Dropout* é reduzir o risco de *overfitting* e incorporar algumas propriedades de invariância espacial (e.g., robustez à imagens em ângulos diferentes), enquanto que a camada de *Batch Normalization* tem a função de gerar uma distribuição de valores mais estável para posterior processamento via ativação *Softmax* nos neurônios de saída.

Todas as etapas de construção e treinamento do modelo proposto foram implementadas em linguagem Python, por meio das bibliotecas Keras e TensorFlow. Ademais, o algoritmo de otimização empregado foi o Adam, em conjunto com uma função perda do tipo *Binary Cross-Entropy*.

Como forma de avaliar o desempenho do modelo de RNC proposto, foram utilizadas as seguintes métricas: acurácia, especificidade, sensibilidade e AUC.

3. Análise dos resultados

Os resultados obtidos no processo de validação cruzada 10-fold foram os seguintes (média da métrica (desvio padrão)): acurácia 0.918 (0.023), especificidade 0.956 (0.031), sensibilidade 0.812 (0.017) e AUC 0.871 (0.029). No caso do conjunto de teste, a uma sensibilidade de 0,75, o modelo alcançou uma especificidade de 0,90 (*versus* 0,865 no modelo de Brinker et al., 2019 e 0,6 para os dermatologistas), e a uma especificidade de 0,6, alcançou uma sensibilidade de 0,9 (*versus* 0,875 em Brinker et al., 2019 e 0,75 para os dermatologistas). O valor de AUC foi de 0.86 (Link para o gráfico da curva ROC: <https://drive.google.com/file/d/1FXr6pOp99mxaVGvAVGmE46ffIKLbcWSu/view?usp=sharing>).

Os valores obtidos para as métricas consideradas evidenciam que o modelo de RNC proposto teve um desempenho bastante satisfatório, especialmente pelas baixas taxas de falso positivo/negativo e um elevado AUC. Vale destacar que o número de imagens utilizadas para treinar o modelo foi significativamente menor que em Brinker et al. (2019), e mesmo neste cenário a metodologia implementada apresentou desempenho levemente superior. Algumas imagens classificadas pelo modelo de RNC podem ser visualizadas na Figura 2.

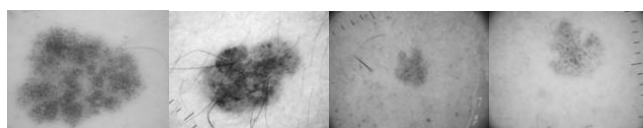


Figura 2. Amostra de imagens de nevos displásicos (duas primeiras) e melanomas (duas últimas) classificadas pelo modelo, sendo que a segunda é um falso positivo e a última é um falso negativo.

4. Considerações Finais

Neste trabalho objetivou-se desenvolver um novo modelo de RNC para detecção de melanoma em imagens dermatoscópicas. Por meio do método de *transfer learning*, utilizando como base a arquitetura Inception-ResNet-v2, foi proposto um novo modelo com três camadas adicionais, treinado especificamente para a tarefa de detecção de melanoma em imagens dermatoscópicas. A base de dados foi composta por 7.818 imagens dermatoscópicas do banco de dados da ISIC, relativas a melanomas e nevos displásicos.

Os resultados obtidos via validação cruzada 10-fold evidenciaram que, mesmo utilizando um conjunto pequeno de treino, o modelo de RNC proposto teve um desempenho satisfatório no problema de classificação binária, com valores elevados de sensibilidade, especificidade e AUC. Além disso, o modelo teve um desempenho bastante superior ao dos dermatologistas em Brinker et al. (2019). De tal forma, concluímos que o objetivo deste trabalho foi alcançado, demonstrando o potencial do modelo proposto como ferramenta de apoio ao diagnóstico clínico de melanoma.

Uma limitação deste estudo (considerando uma possível aplicação para detectar melanoma em brasileiros) é o fato de o modelo não ter sido treinado com imagens representativas dos diversos grupos étnicos brasileiros, além da não inclusão de variáveis relativas ao histórico do paciente (informações que podem melhorar o classificador). Sendo assim, uma proposta para trabalhos futuros é a de incluir *features* dos pacientes no modelo e, se possível, utilizar um banco de dados mais representativo da população brasileira, de tal sorte que o modelo possa ser aplicado nesse contexto (em pacientes brasileiros) e seu desempenho em termos de especificidade e sensibilidade seja melhorado.

Referências

- Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., ... e Schrüfer, P. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. In *European Journal of Cancer*, 113, pages 47-54.
- Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... e Zalaudek, I. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. In *Annals of oncology*, 29(8), pages 1836-1842.
- Instituto Nacional do Câncer (2021). *Tipos de Câncer: Melanoma*. In Instituto Nacional do Câncer (Brasil). Disponível em: < <https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-melanoma>>. Acesso em: 27 de maio de 2021.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. In *Science*, 349, pages 255-260.
- Murdoch, T. B. e Detsky, A. S. (2013). The inevitable application of big data to health care. In *Jama*, 309, pages 1351-1352.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1)*.