

# Aprendizado profundo aplicado ao reconhecimento de células em exames de Papanicolau

Henrique Castro e Silva<sup>1</sup>, Leonardo Caetano Gomide<sup>1</sup>, Alexei Manso Correa Machado<sup>1,2</sup>

<sup>1</sup>Departamento de Ciência da Computação - Instituto de Ciências Exatas e Informática  
Pontifícia Universidade Católica de Minas Gerais (PUC-MG)

<sup>2</sup>Departamento de Anatomia e Imagem - Faculdade de Medicina  
Universidade Federal de Minas Gerais (UFMG)

henrique.castros@outlook.com, leocgomide@gmail.com, alexeimcmachado@gmail.com

**Abstract.** *This paper addresses the problem of cervix cancer detection using deep-learning methods on cells extracted from pap-smear tests. We present a methodology for cell classification, along with an evaluation of the effectiveness of different convolutional models in this classification task.*

**Resumo.** *Este artigo investiga o problema de apoio ao diagnóstico de câncer de colo de útero por meio da aplicação de aprendizado profundo para análise de células do exame Papanicolau. O trabalho apresenta um processo para classificação de células, além dos resultados de diferentes modelos convolucionais na tarefa de classificação.*

## 1. Introdução

O teste de Papanicolau, um procedimento crucial na detecção precoce de câncer cervical e outras anomalias do trato genital feminino, envolve a análise minuciosa de células epiteliais coletadas do colo do útero. No entanto, a interpretação dessas células pode ser complexa e requerer técnicas avançadas de processamento de imagem para classificar com precisão as células presentes.

No trabalho de [Rezende et al. 2021] é proposto um novo dataset, a coleção CRIC, que é composta por 400 imagens coletadas por meio do exame Papanicolau, cada imagem com resolução de 1376 por 1020 pixels. Cada imagem é composta por múltiplas células que foram classificadas de forma manual pelos autores, totalizando 11534 células classificadas junto da posição de seus centroides. As células são classificadas entre 6 classes, segundo a classificação Bethesda proposta em [Nayar and Wilbur 2015], onde a 5 classes indicam tipos de lesões e uma indica uma célula sem lesões, que neste artigo é referenciada como classe negativa.

Diversos trabalhos como os de [Ji et al. 2023], [Zak et al. 2022] e [Mosiichuk et al. 2023] fizeram uso de datasets de imagens para predição dos núcleos das células. Em [Ji et al. 2023] os modelos de U-Net, U-Net++ e outras arquiteturas foram usados na tarefa de segmentação de núcleos. [Zak et al. 2022] treinou um modelo classificador com ajuda de um GAN (*Generative Adversarial Network*) treinado no dataset. Mais recentemente, [Mosiichuk et al. 2023] propuseram usar o arquitetura RetinaNet para classificar imagens obtidas por meio de imagens microscópicas obtidas por meio de aparelhos celulares.

Este breve artigo propõe compartilhar resultados preliminares encontrados no processo de classificação automatizada das imagens de células que estão disponíveis no banco de dados apresentado em [Rezende et al. 2021]. Uma abordagem sistemática foi criada para classificação de células, bem como uma análise do desempenho das arquiteturas ResNet, MobileNet e EfficientNet, presentes em [Tan and Le 2019], [Howard et al. 2017] e [He et al. 2016], comparados a uma pequena arquitetura de rede convolucional, que foram treinados neste contexto. A escolha das arquiteturas foi determinada pelo sucesso em diferentes tarefas e a popularidade que cada arquitetura apresenta.

## 2. Metodologia

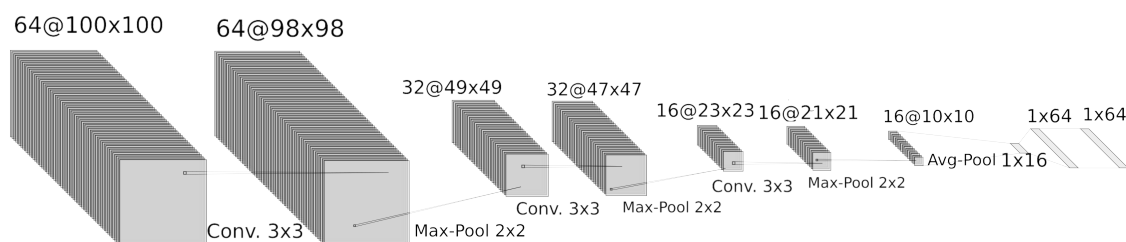
Os modelos foram inteiramente implementados em Python3, pelo qual se criou o módulo de predição, que faz o gerenciamento dos modelos treinados<sup>1</sup>.

Conforme descrito anteriormente, uma imagem do dataset CRIC possui mais de uma célula classificada. Dessa forma, antes de treinar os modelos, foram criadas imagens auxiliares de tamanho 100x100 pixels, onde o centro da imagem corresponde ao centro de uma célula, seguindo os valores anotados dos centroides.

Para células que ficavam próximas às bordas da imagem, que impossibilitavam o corte de tamanho 100x100, o recorte foi ampliado para maximizar o tamanho e o restante foi preenchido com pixels de valor 0.

Para treinamento e execução dos modelos de *deep-learning*, foi utilizado o módulo *Tensorflow*, que já possui os modelos pré-treinados no dataset ImageNet. Para cada um dos modelos pré-treinados, foi adicionado somente a função de pré-processamento equivalente e uma camada de saída com tamanho 1 e ativação sigmoide para a tarefa binária, ou tamanho 6 e ativação softmax para a multiclasse.

Para auxiliar o treinamento, foi criada uma estrutura de dados adicional, composta pela imagem da célula a ser analisada, o centroide da célula, a classe da célula (usada na tarefa multiclasse) e a classe binária (usada na tarefa binária). A classe binária é um mapeamento da classe da célula, onde a classe negativa é mapeada para 0 e todas as outras foram mapeadas para 1, indicando a presença de lesão.



**Figura 1. Arquitetura da rede Neural baseline**

Além dos modelos pré-treinados, foi desenvolvido um modelo próprio, chamado de *Baseline*, que tem a função de analisar como um modelo pequeno sem pré-treinamento prévio se comportaria na solução da tarefa. A arquitetura do modelo pode ser observado

<sup>1</sup>Os modelos assim como todo o código desenvolvido para este artigo estão disponíveis em <https://github.com/HenriqueCastros/pai-flower-nucleus>

na figura 1, nela nota-se que o modelo contém 3 camadas de convolução seguido de *max-pooling*, e 3 camadas densas no final, com a mesma camada de saída que os modelos pré-treinados.

### 3. Análise de resultados

Na tabela 1 podem ser observados os resultados obtidos para cada modelo em cada uma das tarefas, identificação se a célula é positiva ou não para uma lesão (binário) e qual lesão a célula possui (multiclasse). Também pode ser observada a quantidade de parâmetros de cada modelo.

**Tabela 1. Resultados**

Modelo Base	N. Param.	Tarefa	Acurácia	F1-Score
Baseline	30k	Binário	86%	0.86
MobileNetV2	2.2M	Binário	87%	0.87
<b>EfficientNetB3</b>	<b>10.6M</b>	<b>Binário</b>	<b>88%</b>	<b>0.87</b>
<b>Resnet 50</b>	<b>23.5M</b>	<b>Binário</b>	<b>88%</b>	<b>0.88</b>
Baseline	30k	Multiclasse	75%	0.46
MobileNetV2	2.2M	Multiclasse	78%	0.57
<b>EfficientNetB3</b>	<b>10.6M</b>	<b>Multiclasse</b>	<b>80%</b>	<b>0.58</b>
<b>Resnet 50</b>	<b>23.5M</b>	<b>Multiclasse</b>	<b>80%</b>	<b>0.58</b>

Nota-se que os modelos com mais parâmetros foram capazes de obter resultados melhores em ambas as tarefas, devido à sua maior capacidade de processamento. No entanto é importante ressaltar que o ganho não é linear, e que principalmente na tarefa mais simples, de detecção binária de lesão, o modelo baseline teve uma acurácia somente 2% menor que o modelo *resnet50*, que é por volta de 1000x maior em número de parâmetros. Isso indica que o modelo maior pode ser excessivo para a tarefa, possivelmente por codificar informações redundantes.

É interessante observar que, embora haja uma tendência geral de aumento de desempenho com modelos mais complexos em ambas as tarefas, essa tendência é mais pronunciada na tarefa de classificação multiclasse em comparação com a tarefa binária de identificação de lesões. Isso pode indicar que a complexidade do modelo é crucial em problemas de classificação multiclasse, onde há uma variedade maior de padrões que demandam a análise de maiores detalhes.

Na tarefa de classificação de lesão houve uma melhora de 5% na acurácia entre o modelo de referência e o maior modelo testado, que já é uma diferença mais significativa. Mas o fato do ganho de desempenho não ser linear ainda é mantido, sendo que os modelos *EfficientNetB3* e *Resnet 50* obtiveram o mesmo desempenho, apesar de o primeiro ser metade do tamanho do segundo. Reforçando a hipótese que modelos muito grandes podem estar sendo redundantes para a tarefa.

Finamente, é importante ressaltar que os modelos maiores também são mais sensíveis a *overfitting*, e portanto os números de camadas treinadas dos modelos necessitam ser ajustados e otimizados. Esse processo foi experimental, tendo sido escolhido o de melhor performance. No entanto, os modelos ainda sofreram algum *overfitting*.

## 4. Conclusão

A análise dos resultados revela insights valiosos sobre o desempenho e a eficiência dos modelos de detecção de lesões avaliados. Os modelos mais complexos tendem a apresentar melhores resultados em termos de acurácia e F1-Score. No entanto, essa melhoria não é linear, sugerindo que há um ponto de saturação em relação ao aumento da complexidade.

Além disso, observa-se que a eficiência computacional também é um fator importante a ser considerado ao se escolher um modelo, especialmente em cenários onde os recursos são limitados. Modelos como o EfficientNetB3 demonstraram ser capazes de alcançar desempenho competitivo com menos parâmetros em comparação com modelos mais complexos, como o ResNet50.

A variação no desempenho entre as tarefas de identificação binária e classificação multiclasse ressalta a importância de se adaptar o modelo à natureza específica do problema. Enquanto a complexidade do modelo desempenha um papel crucial em problemas multiclasse mais desafiadores, pode não ser tão crucial em tarefas binárias mais simples.

Como trabalhos futuros, pode ser feita uma análise e otimização sistemática de hiperparâmetros para obter melhores resultados com os modelos testados. Uma limitação deste que pode ser explorada seria a identificação automática de centroides, que no modelo proposto são anotados manualmente.

## Referências

- 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.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Ji, J., Zhang, W., Dong, Y., Lin, R., Geng, Y., and Hong, L. (2023). Automated cervical cell segmentation using deep ensemble learning. *BMC Medical Imaging*, 23(1):137.
- Mosiichuk, V., Sampaio, A., Viana, P., Oliveira, T., and Rosado, L. (2023). Improving mobile-based cervical cytology screening: A deep learning nucleus-based approach for lesion detection. *Applied Sciences*, 13(17).
- Nayar, R. and Wilbur, D. (2015). *The Bethesda System for Reporting Cervical Cytology. Definitions, Criteria, and Explanatory Notes*.
- Rezende, M. T., Silva, R., Bernardo, F. d. O., Tobias, A. H. G., Oliveira, P. H. C., Machado, T. M., Costa, C. S., Medeiros, F. N. S., Ushizima, D. M., Carneiro, C. M., and Bianchi, A. G. C. (2021). Cric searchable image database as a public platform for conventional pap smear cytology data. *Scientific Data*, 8(1):151.
- Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.
- Zak, J., Grzeszczyk, M. K., Pater, A., Roszkowiak, L., Siemion, K., and Korzynska, A. (2022). Cell image augmentation for classification task using gans on pap smear dataset. *Biocybernetics and Biomedical Engineering*, 42(3):995–1011.