

Classificação de Displasias Epiteliais Orais utilizando MobileNetV3-Small: Uma Abordagem com *Learning Rate Decay* e *Cross-Validation*

Luis Felipe G. S. Paim¹, Domingos L. L. de Oliveira^{1,2}, Marcelo Z. do Nascimento¹

¹ Faculdade de Computação (FACOM) – Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brasil

² Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP)
São Paulo – SP – Brasil

luis.paim@ufu.br

Abstract. *The subjective histopathological diagnosis of oral epithelial dysplasias demands accurate Computer-Aided Diagnosis (CAD) systems. This paper investigates the MobileNetV3-Small architecture for classifying these lesions. To overcome data scarcity and prevent overfitting, we compared a baseline model with an optimized approach using transfer learning, 5-fold cross-validation, and data augmentation. Furthermore, the use of an adaptive learning rate improved model convergence. The baseline achieved 47.06% test accuracy, while the optimized approach reached up to 79.56%. The methodology provides a high-accuracy, computationally efficient solution for histopathology, mitigating interobserver variability.*

Resumo. *O diagnóstico histopatológico subjetivo de displasias epiteliais orais demanda sistemas CAD precisos. Este artigo investiga a arquitetura MobileNetV3-Small para a classificação dessas lesões. Para superar a escassez de dados e o sobreajuste, comparou-se um modelo baseline treinado do zero com uma abordagem otimizada via transfer learning, validação cruzada (5 folds) e data augmentation. O uso de uma taxa de aprendizado adaptativa também melhorou a convergência. O baseline obteve 47,06% de acurácia no teste, enquanto a abordagem otimizada alcançou até 79,56%. A metodologia fornece uma solução eficiente e de alta precisão para a histopatologia, mitigando a variabilidade interobservador.*

1. Introdução

O câncer de cavidade oral representa um desafio oncológico global, com mais de 389 mil diagnósticos em 2022 [International Agency for Research on Cancer (IARC) 2022] e 10.900 novos casos estimados no Brasil em 2023 [Instituto Nacional de Câncer (INCA) 2023]. Grande parte desses carcinomas evolui de desordens potencialmente malignas caracterizadas por Displasia Epitelial Oral (DEO) [Warnakulasuriya et al. 2007]. O diagnóstico padrão-ouro é a avaliação histopatológica, focada em alterações morfológicas e arquiteturais complexas, como pleomorfismo nuclear, perda de polaridade basal e hiperchromatismo [Cheung et al. 2022]. Contudo, a alta variabilidade interobservador e a subjetividade

intrínseca do método visual humano elevam os custos operacionais e o tempo de liberação dos laudos.

Sistemas de *Computer-Aided Diagnosis* (CAD) surgem como soluções viáveis para prover análises quantitativas e reprodutíveis, buscando alta sensibilidade para evitar o risco crítico de falsos negativos na prática clínica. Nesse cenário, a arquitetura MobileNetV3-Small destaca-se por conciliar eficiência computacional e desempenho preditivo competitivo, superando arquiteturas clássicas como a ResNet18 em *benchmarks* públicos com uma fração do custo de processamento [Shahriar 2025].

Entretanto, a aplicação direta de redes neurais profundas em domínios médicos enfrenta a restrição da escassez quantitativa de *datasets* anotados, o que frequentemente induz ao *overfitting* [Bradshaw et al. 2023]. Este trabalho investiga o desempenho da MobileNetV3-Small aplicada à DEO, integrando estratégias de *transfer learning*, validação cruzada estratificada (*cross-validation*), aumento de dados geométrico (*data augmentation*) e taxa de aprendizado adaptativa (*learning rate decay*) para garantir uma convergência robusta, minimizando a memorização de ruídos e maximizando a capacidade de generalização em tecidos não vistos.

2. Materiais e Métodos

A metodologia abrange o treinamento e a avaliação de generalização da MobileNetV3-Small em lâminas histológicas digitais, conforme ilustrado no *pipeline* estruturado na Figura 1. Os experimentos foram executados em hardware equipado com processador Intel Core i7 (13ª geração), 16 GB de RAM e placa de vídeo NVIDIA RTX 3050 com 6 GB de VRAM.

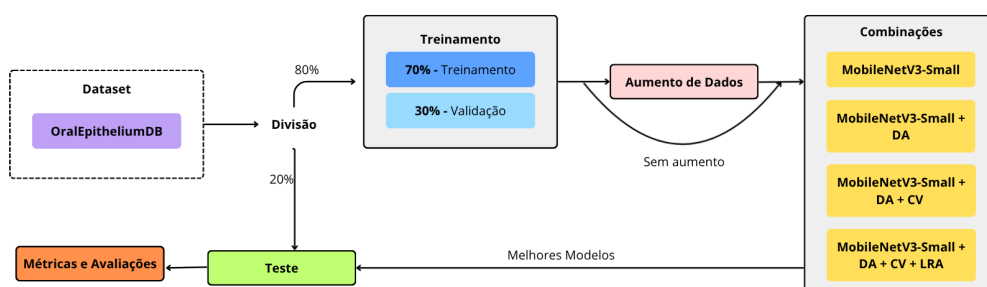


Figura 1. Pipeline da metodologia proposta, incluindo as etapas de divisão dos dados e avaliação dos modelos nos *datasets*.

2.1. Base de Imagens

Utilizou-se o *dataset* *OralEpitheliumDB* [Silva et al. 2024], composto por 456 regiões de interesse (ROIs) extraídas sob magnificação de 400x (resolução espacial de 450x250 pixels). As imagens dividem-se em quatro classes diagnósticas: tecido saudável, displasia leve, displasia moderada e displasia severa (Figura 2). A variabilidade de coloração do estroma e das estruturas celulares reflete as condições reais das lâminas analisadas laboratorialmente.

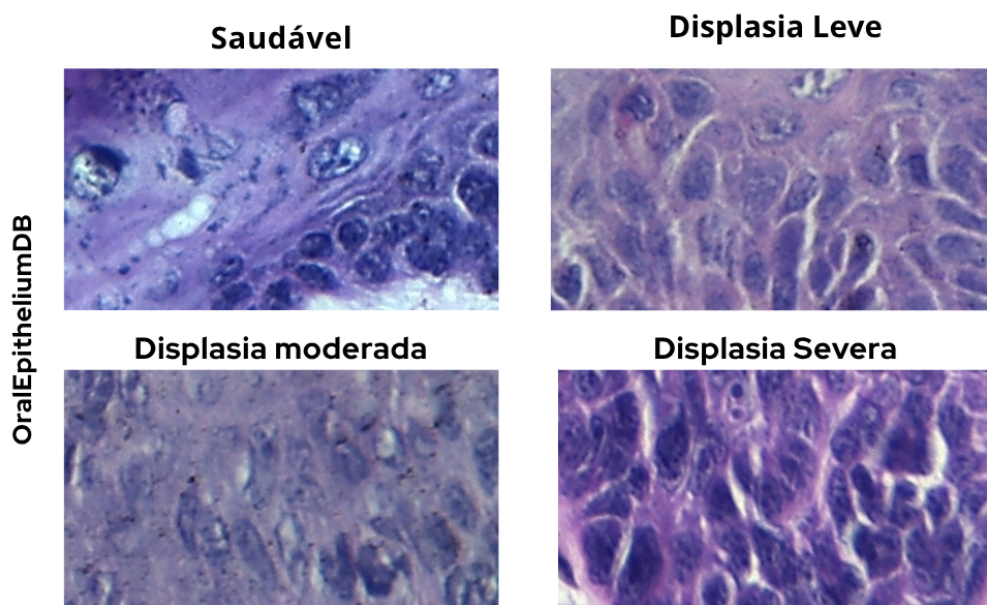


Figura 2. Exemplos de imagens histológicas extraídas do *dataset OralEpitheliumDB*, evidenciando a variação morfológica entre as classes.

2.2. Classificação e Estratégias de Otimização

A MobileNetV3-Small [Howard et al. 2019] utiliza blocos residuais invertidos, convoluções separáveis em profundidade (*Depthwise*) e ativação *Hard-Swish*. Adicionalmente, incorpora módulos de atenção *Squeeze-and-Excite* (SE) para recalibrar adaptativamente as respostas dos canais, tornando-a ideal para extrair texturas celulares complexas com restrições computacionais. Para a inicialização espacial, empregou-se *Transfer Learning* [Pan and Yang 2009] importando pesos mapeados na base *ImageNet* [Deng et al. 2009].

Para garantir rigor na avaliação estatística, o *dataset* foi dividido em uma partição *hold-out* com 20% das imagens reservadas exclusivamente para teste isolado. Os 80% restantes foram submetidos a uma validação cruzada estratificada com 5 *folds*, mantendo a proporção exata das classes em cada iteração de treino e validação. Aplicou-se *Data Augmentation* [Shorten and Khoshgoftaar 2019] restrito aos conjuntos de treinamento, utilizando inversão horizontal e rotações aleatórias ($\pm 15^\circ$) para simular variabilidades de posicionamento da câmera do microscópio.

A otimização dos gradientes utilizou a função *ReduceLROnPlateau*, que aplica um fator de redução multiplicativo na taxa de aprendizado caso a métrica de *Cross-Entropy Loss* [Goodfellow et al. 2016] na validação não apresente melhoria por 5 épocas consecutivas, permitindo um *fine-tuning* limpo e estabilizado.

3. Resultados e Discussão

A Tabela 1 sumariza o impacto quantitativo de cada técnica no diagnóstico automatizado. Todos os experimentos adotaram 100 épocas de processamento e uma taxa de aprendizado inicial de 1×10^{-4} . A acurácia de validação é reportada como média $\pm$ desvio padrão considerando os 5 *folds* da validação cruzada. A acurácia de teste, por sua vez,

corresponde à avaliação única do melhor modelo selecionado em cada configuração sobre o conjunto *hold-out*; por isso, o desvio padrão é apresentado apenas para a validação cruzada.

Tabela 1. Desempenho da MobileNetV3-Small sob diferentes estratégias de otimização no *dataset OralEpitheliumDB*.

Configuração Metodológica	Seed	TL	DA	Acc. Teste <i>hold-out</i> (%)	Acc. Validação (%) $\pm$ DP
Baseline (Treino do zero)	42	Não	Não	47,06	53,12 $\pm$ 0,02
Otimização Parcial (Sem DA)	42	Sim	Não	71,53	71,71 $\pm$ 8,79
	17	Sim	Não	71,53	73,08 $\pm$ 11,38
	12	Sim	Não	73,72	82,38 $\pm$ 5,83
Otimização Completa	42	Sim	Sim	79,56	73,00 $\pm$ 7,93
	17	Sim	Sim	71,01	68,51 $\pm$ 4,35
	12	Sim	Sim	75,36	73,51 $\pm$ 8,85

TL = *Transfer Learning*; DA = *Data Augmentation*; DP = Desvio Padrão.
A acurácia de validação é apresentada como média $\pm$ DP dos 5 *folds*; a acurácia de teste refere-se à avaliação única no conjunto *hold-out*.

O modelo *baseline* obteve apenas 47,06% de acurácia no teste, evidenciando que inicializar pesos de forma aleatória em *datasets* médicos pequenos impede a rede de aprender limites decisórios consistentes. A ativação do *Transfer Learning* elevou a acurácia global para o patamar de 73%. Contudo, na ausência de regularização espacial (*Data Augmentation*), a arquitetura demonstrou dependência das distribuições dos *folds*, resultando em uma instabilidade de validação com desvio padrão de até 11,38%. A integração de todas as técnicas atuou de forma sinérgica, atingindo o pico de 79,56% de assertividade e demonstrando que o modelo reduziu o mapeamento de ruídos de fundo para focar nas estruturas displásicas reais.

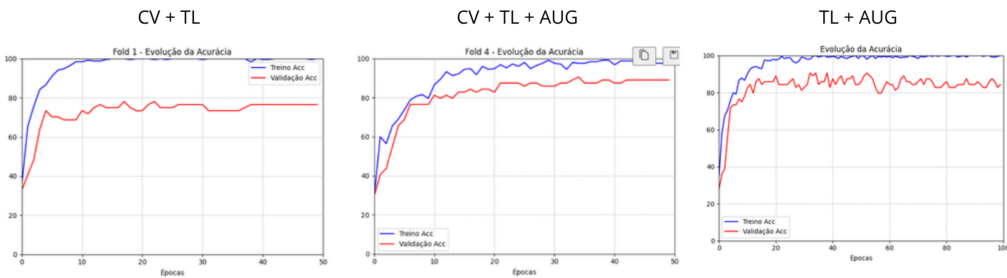


Figura 3. Curvas de treinamento: (A) *Overfitting* agudo sem DA (*Fold* 2); (B) Estabilização com DA (*Fold* 5); (C) Impacto visual do *Learning Rate Adaptive* na supressão da variância e estabilidade de perda.

A topologia dessas convergências (Figura 3) reflete as métricas tabuladas. A Figura 3A demonstra que, sem DA, a rede esgota sua capacidade de aprendizado antes da 15ª época, gerando o clássico *overfitting* manifestado pelo crescimento contínuo do custo de validação (efeito "U"). O DA (Figura 3B) exige da rede um aprendizado invariante à rotação, achatando o custo em um platô horizontal de segurança. A eficácia operacional do *Learning Rate Adaptive* é isolada na Figura 3C: a atuação exata do *ReduceLROnPlateau* cessa os saltos aleatórios de validação nas últimas épocas, ancorando a rede no

mínimo local da função de custo e resultando em acurácia no teste isolado de 77,37% naquele *fold* específico.

4. Conclusão

Este estudo confirmou que a MobileNetV3-Small, estruturada sob protocolos de regularização, é instrumental para a classificação de displasias epiteliais orais. O desempenho insatisfatório do treinamento *baseline* foi sobreposto pela arquitetura otimizada, que combinou a maturidade de extração do *Transfer Learning* com a robustez matemática provida pela Validação Cruzada, *Data Augmentation* e decaimento adaptativo de *Learning Rate*. Com acurácias validadas próximas a 80%, a metodologia apresenta maturidade para apoiar laudos patológicos minimizando discrepâncias visuais. Pesquisas subsequentes devem concentrar esforços em algoritmos de *Early Stopping* aliados a *K-Fold* e na extração de métricas de interpretabilidade, como mapas de calor (*Grad-CAM*), para integrar a validação clínica visual ao processo CAD.

Referências

- Bradshaw, T. J., McCradden, M. D., Jha, A. K., et al. (2023). A guide to cross-validation for artificial intelligence in medical imaging. *Radiology: Artificial Intelligence*, 5(4):e220232.
- Cheung, V., Hulme, K., and Gupta, R. (2022). Oral epithelial dysplasia: A review of diagnostic criteria for anatomic pathologists. *Advances in Anatomic Pathology*, 29(6):416–432.
- 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.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324.
- Instituto Nacional de Câncer (INCA) (2023). Estimativa 2023: incidência de câncer no brasil. Acesso em: 07 maio 2026.
- International Agency for Research on Cancer (IARC) (2022). Global cancer observatory: Cancer today. Acesso em: 07 maio 2026.
- Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- Shahriar, M. (2025). Comparative analysis of lightweight deep learning models for memory-constrained devices. *arXiv preprint arXiv:2505.03303*.
- Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.
- Silva, A. B. et al. (2024). Oralepitheliumdb: A dataset for oral epithelial dysplasia image segmentation and classification. *Journal of Imaging Informatics in Medicine*, 37(4):1691–1710.

Warnakulasuriya, S., Johnson, N. W., and van der Waal, I. (2007). Nomenclature and classification of potentially malignant disorders of the oral mucosa. *Journal of Oral Pathology & Medicine*, 36(10):575–580.