Segmentação de Imagens Histológicas da Cavidade Oral: Um Estudo sobre Variantes da U-Net e *Lightweight Backbones*

Luana R. Borges[1], Vitória F. C. Silva[1], Davi Soares[1], Gustavo C. Miranda[1], Luis Felipe G. S. Paim[1], Daniel B. Gonçalves[1], Adriano B. Silva [1], Domingos L. L. de Oliveira[1,2], Leandro A. Neves [3] , Marcelo Z. do Nascimento[1]

[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

[3]Departamento de Ciência da Computação e Estatística (DCCE) – UNESP
São José do Rio Preto – SP – Brasil

luana.borges1@ufu.br

Abstract. *Oral cavity cancer presents a high incidence, making early diagnosis essential for improving patient prognosis. In this context, the automatic analysis of histopathological images has gained increasing attention; however, their high structural complexity and morphological variability pose significant challenges. Semantic segmentation is a fundamental step in computer-aided diagnosis systems, as it enables the automatic delineation of regions of interest. This work evaluates different architectures from the U-Net family for histological image segmentation, including conventional models, enhanced variants, and versions incorporating lightweight pre-trained backbones. Experiments were conducted on two public datasets, achieving Dice coefficients of 0.912 and 0.871 for the OCDC and OralEpitheliumDB datasets, respectively. Furthermore, the impact of data augmentation, pre-training, and the trade-off between segmentation performance and computational efficiency were analyzed. The results demonstrate the potential of the evaluated architectures and contribute to a better understanding of more robust and efficient strategies in this domain.*

Resumo. *O câncer da cavidade oral apresenta elevada incidência, tornando o diagnóstico precoce essencial para a melhoria do prognóstico. Nesse contexto, a análise automática de imagens histopatológicas tem ganhado destaque, embora a elevada complexidade estrutural e a variabilidade morfológica imponham desafios significativos. A segmentação semântica constitui uma etapa fundamental em sistemas de diagnóstico assistido por computador, pois permite a delimitação automática de regiões de interesse. Este trabalho avalia diferentes arquiteturas da família U-Net para a segmentação de imagens histológicas, incluindo modelos tradicionais, variantes aprimoradas e versões com lightweight pre-trained backbones. Os experimentos foram conduzidos em dois conjuntos de dados públicos, alcançando coeficientes Dice de 0,912 e 0,871 para os bancos OCDC e OralEpitheliumDB, respectivamente. Além disso, foram analisados o impacto do aumento de dados, do pré-treinamento e o compromisso entre*

desempenho de segmentação e eficiência computacional. Os resultados evidenciam o potencial das arquiteturas avaliadas e contribuem para a compreensão de estratégias mais robustas e eficientes nesse domínio.

1. Introdução

Segundo o Instituto Nacional do Câncer (INCA), o câncer é a segunda causa de óbito no mundo, com taxa de incidência em aumento. No Brasil, no triênio de 2023 a 2025, estimou-se aproximadamente 704 mil novos diagnósticos de câncer, dos quais 15.100 casos de câncer da cavidade oral. Esse tipo de câncer ocupa o oitavo lugar em termos de mortalidade global [Santos 2023]. Uma das formas de minimizar a ocorrência deste tipo de câncer é a realização de diagnósticos precoces como ferramenta para melhorar as chances de cura e o prognóstico dos pacientes acometidos.

As imagens histológicas desempenham um papel central no diagnóstico e prognóstico de diversas patologias, incluindo câncer e lesões potencialmente malignas, uma vez que permitem a análise detalhada da organização celular e das estruturas teciduais. No entanto, essas imagens apresentam alta complexidade estrutural, caracterizada por variações morfológicas, sobreposição de tecidos e diferenças de coloração, o que dificulta sua análise automática por sistemas computacionais [Basu et al. 2024, Xu et al. 2024]. Esses fatores tornam a análise automática de imagens histopatológicas um desafio significativo para sistemas computacionais, impactando na confiabilidade de métodos computacionais aplicados nesse domínio [Silva et al. 2023].

A etapa de segmentação tem sido empregada como uma estratégia para organizar as informações visuais presentes nas imagens histológicas, possibilitando a delimitação de regiões de interesse e favorecendo análises quantitativas mais consistentes em sistemas de diagnóstico assistido por computador (do inglês, *computer-aided diagnosis* – CAD) [Moscalu et al. 2023]. Entretanto, a segmentação manual dessas imagens permanece um processo dispendioso, fortemente dependente de especialistas e suscetível à variabilidade interobservador. Embora as anotações realizadas por especialistas constituam o padrão-ouro para o treinamento e validação dos modelos, a necessidade de grandes volumes de dados rotulados reforça a busca por novas soluções [Madabhushi and Lee 2016].

Nos últimos anos, abordagens baseadas em redes neurais convolucionais (do inglês, *convolutional neural networks* – CNNs) para a segmentação de imagens médicas têm apresentado resultados relevantes em muitos cenários em relação as abordagens clássicas baseadas em processamento de imagens [Srinidhi et al. 2021, Springenberg et al. 2023]. Em particular, arquiteturas *encoder–decoder* tornaram-se amplamente adotadas pela capacidade de integrar informações espaciais e semânticas em múltiplas escalas [Litjens 2017]. Dentre essas arquiteturas, a U-Net destaca-se como uma das mais utilizadas na segmentação semântica de imagens biomédicas, apresentando resultados importantes mesmo em cenários com conjuntos de dados reduzidos [Ronneberger et al. 2015].

A partir da U-Net, diversas variações foram propostas com o objetivo de aprimorar a representação de contexto, o refinamento de bordas e o delineamento de estruturas complexas. Exemplos incluem a UNet++, que explora conexões densas para reduzir o *gap* semântico entre *encoder* e *decoder* [Zhou et al. 2018], a Attention U-Net, que incorpora mecanismos de atenção para enfatizar regiões relevantes da imagem

[Oktay et al. 2018], e a Sharp U-Net, voltada ao realce de bordas e transições entre regiões segmentadas [Zunair and Hamza 2021]. Apesar do desempenho promissor dessas arquiteturas, suas aplicações em imagens histológicas ainda enfrentam desafios relacionados à limitação de dados anotados, à variabilidade entre domínios histopatológicos e ao elevado custo computacional de alguns modelos CNNs [Banerji and Mitra 2022].

Estratégias como o aumento de dados e o uso de redes pré-treinadas em grandes bases de imagens naturais, como o ImageNet [Deng et al. 2009], têm sido amplamente investigadas para investigar a escassez de dados e melhorar a capacidade de generalização dos modelos [Cheplygina 2019]. Paralelamente, modelos *lightweight backbones*, como MobileNet [Howard et al. 2017] e EfficientNet [Tan and Le 2019], surgem como alternativas para aplicações com restrições computacionais, permitindo um melhor equilíbrio entre desempenho e custo de processamento. Entretanto, ainda há desafios em relação a como essas estratégias influenciam o desempenho de diferentes variantes da U-Net quando aplicadas a bases histológicas com variações intrínsecas relacionadas aos padrões de coloração, à diversidade morfológica dos tecidos e às definições das regiões de interesse [Banerji and Mitra 2022].

Este trabalho investiga métodos baseados em CNNs para segmentação semântica de tecidos histológicos da cavidade oral, explorando a U-Net tradicional, suas variantes aprimoradas e versões com incorporação de *lightweight backbones* pré-treinados. Os experimentos foram conduzidos em dois conjuntos de dados públicos, representativos de diferentes contextos histopatológicos. Como principais contribuições, destacam-se a avaliação comparativa de diferentes variantes da U-Net sob condições experimentais padronizadas, a análise do impacto do aumento de dados na capacidade de generalização dos modelos e a investigação do comportamento dessas arquiteturas em bases com diferentes características morfológicas das lesões histológicas.

2. Materiais e Métodos

Neste estudo, foram empregadas diferentes arquiteturas baseadas na U-Net para o treinamento, validação e teste de modelos de segmentação de imagens histológicas como é ilustrado na Figura 1. Os experimentos foram realizados em um computador equipado com processador Intel Core i5 de 12ª geração, 16 GB de memória RAM, GPU NVIDIA GeForce RTX 3050 com 4GB de VRAM e sistema operacional Windows 11.

2.1. Base de Imagens

Neste estudo, foram utilizadas duas bases de imagens histológicas. A primeira, OralEpitheliumDB [Silva et al. 2024a], é composta por 228 imagens de regiões de interesse (ROIs), sendo cada ROI considerada uma imagem individual, com magnificação de 400× e resolução de 450×250 pixels, categorizadas nas classes de tecido saudável e displasia severa. A segunda base, *Oral Cavity-Derived Cancer* (OCDC) [Santos et al. 2023b], contém 1.020 imagens coradas com hematoxilina e eosina (H&E), capturadas com ampliação de 200× e resolução de 640×640 *pixels*, incluindo tecidos normais e carcinoma oral de células escamosas. Ambas as bases incluem máscaras de padrão-ouro para segmentação, devidamente anotadas por patologistas e utilizadas como referência na avaliação. Na Figura 2 são exibidos exemplos de tecidos dessas bases de imagens.

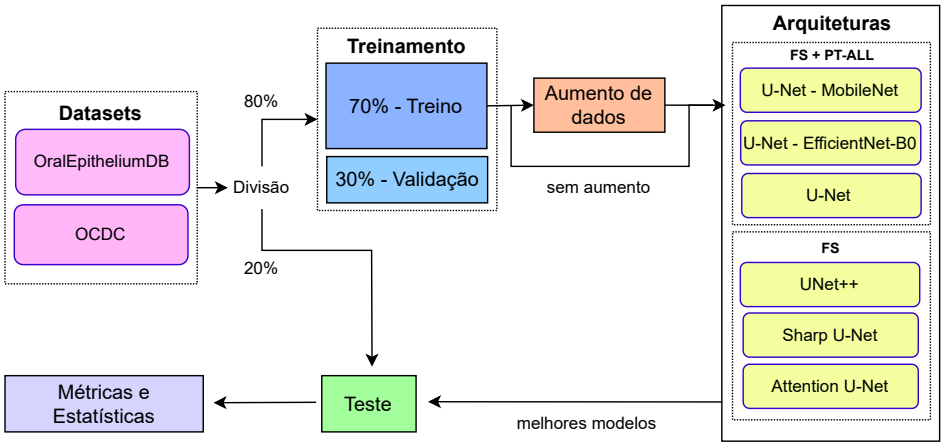

Figura 1. *Pipeline* da metodologia proposta, incluindo as etapas de divisão dos dados e avaliação dos modelos nos *datasets*.

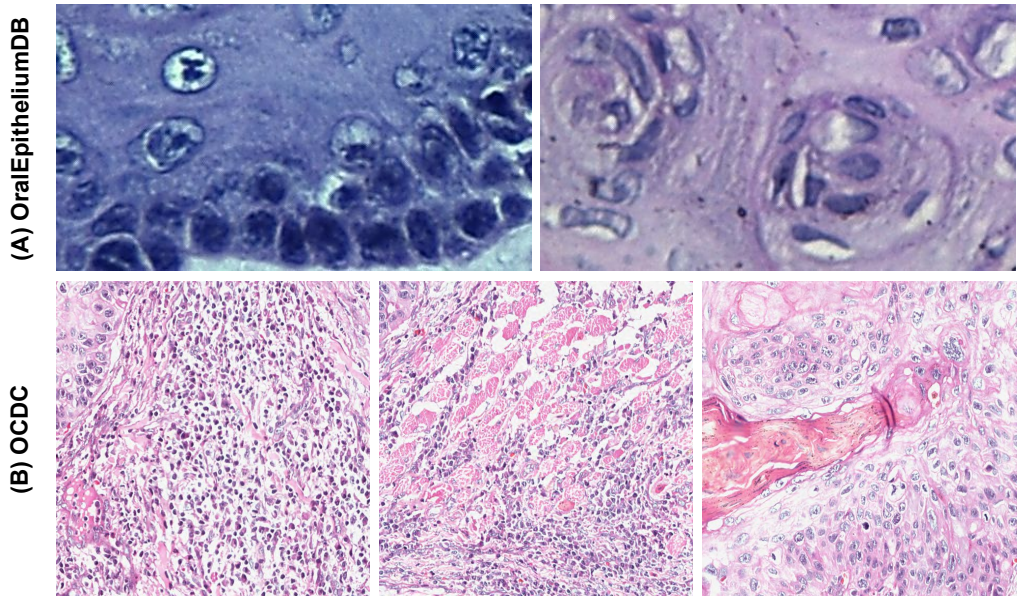

Figura 2. Exemplos de imagens histológicas utilizadas no estudo: (A) imagens de núcleos da base OralEpitheliumDB; (B) imagens da base OCDC com regiões tumorais da cavidade oral.

2.2. Segmentação dos Tecidos Histológicos

Para a análise proposta neste estudo, foram investigados, além da U-Net [Ronneberger et al. 2015], os modelos UNet++ [Zhou et al. 2018], Attention U-Net [Oktay et al. 2018], Sharp U-Net [Zunair and Hamza 2021] e variações da U-Net com *backbones* MobileNet [Howard et al. 2017] e EfficientNet-B0 [Tan and Le 2019]. Esse conjunto contempla modelos clássicos, versões aprimoradas com conexões densas, redes com mecanismos de atenção e arquiteturas voltadas ao refinamento de bordas, permitindo uma avaliação mais abrangente da eficiência dessas abordagens na segmentação de tecidos histológicos de cavidade oral.

A U-Net [Ronneberger et al. 2015] consiste em uma arquitetura em formato de "U", composta por um caminho de contração, responsável pela extração de características semânticas, e um caminho de expansão, que recupera a resolução espacial para produzir mapas de segmentação detalhados. As conexões de salto entre os dois caminhos permitem a preservação de informações espaciais relevantes, contribuindo para a precisão da segmentação. A UNet++ [Zhou et al. 2018] aprimora a arquitetura tradicional ao introduzir conexões densas entre os níveis do *encoder* e *decoder*. Essas conexões reduzem a diferença semântica entre as representações de diferentes níveis da rede, favorecendo a propagação de informações e melhorando o detalhamento das regiões segmentadas.

A Attention U-Net [Oktay et al. 2018] incorpora mecanismos de atenção nas conexões entre *encoder* e *decoder*, permitindo que o modelo destaque automaticamente regiões mais relevantes da imagem. Essa abordagem auxilia na segmentação de estruturas histológicas com limites pouco definidos, reduzindo a influência de informações irrelevantes. Já a Sharp U-Net [Zunair and Hamza 2021] introduz mecanismos voltados ao realce de bordas e refinamento das transições entre regiões segmentadas. Essa característica é particularmente relevante para imagens histopatológicas, nas quais frequentemente existem variações sutis entre diferentes estruturas dos tecidos, exigindo maior precisão na delimitação das regiões de interesse.

Também foram investigadas variações da U-Net utilizando *lightweight backbones* pré-treinados com a MobileNetV2 e EfficientNet-B0. A MobileNetV2 é uma arquitetura projetada para aplicações com restrições computacionais, empregando convoluções separáveis em profundidade para reduzir significativamente o número de parâmetros e operações necessárias. Já a EfficientNet-B0 utiliza uma estratégia de escalonamento composto, que equilibra profundidade, largura e resolução da rede. A variante B0 corresponde ao modelo base da família e, em geral, apresenta menor custo computacional do que versões maiores, sendo apropriada para cenários com restrições de latência e memória.

2.3. Treinamento dos Modelos de Segmentação

O *dataset* OralEpitheliumDB foi particionado em conjuntos de treinamento, validação e teste, utilizando 80% das imagens para treinamento e validação (70% e 30%, respectivamente) e 20% para teste, o qual foi mantido separado e utilizado exclusivamente na avaliação final. Já o *dataset* OCDC encontrava-se previamente organizado em diretórios de treinamento e teste, sendo o conjunto de treinamento subdividido em 70% para treinamento e 30% para validação.

Devido ao tamanho dos *datasets*, foi adotado um recurso de aumento de dados (do inglês: *data augmentation* - DA) com o objetivo de melhorar a capacidade dos modelos de generalizar padrões aprendidos, reduzindo o risco de *overfitting* e melhorando o desempenho em dados de teste. O aumento de dados foi aplicado exclusivamente ao conjunto de treinamento, por meio de transformações geométricas simples, como rotações e inversões. Foram empregadas operações de inversões horizontais aleatórias, com probabilidade de 50% e rotações aleatórias com ângulo variando entre -15° e $+15^{\circ}$. Essas operações promovem invariância e maior robustez dos modelos, permitindo gerar variações artificiais a partir das imagens originais sem alterar suas características semânticas [Shorten and Khoshgoftaar 2019, Cheplygina 2019].

Para os experimentos desse estudo, as variações da U-Net com *backbones* foram

pré-treinadas utilizando o banco de dados do ImageNet [Deng et al. 2009]. Os demais modelos foram treinados do zero. Os hiperparâmetros foram definidos de forma empírica, levando em consideração valores que reduzem o custo computacional da execução dos modelos sem comprometer o desempenho. Dessa forma, o número de épocas de treinamento foi fixado em 30, com base na convergência do coeficiente Dice e na redução da função de perda no conjunto de validação. Para a otimização, foi utilizado o otimizador Adam, combinando a função *Binary Cross-Entropy with Logits Loss* com a **Dice Loss**. Além disso, foi adotada uma taxa de aprendizado de 3×10^{-4} e *batch size* igual a 16. Os experimentos foram conduzidos com oito sementes aleatórias.

2.4. Métricas de Avaliação e Análise Estatística

Para a comparação entre os modelos testados, foram calculados o valor médio do coeficiente Dice por classe. Para a análise estatística, foram realizadas oito execuções com diferentes sementes aleatórias para cada arquitetura. Inicialmente, a normalidade dos dados foi verificada por meio do teste de Shapiro–Wilk. Conforme o comportamento dos dados, foram aplicados o teste paramétrico ANOVA de medidas repetidas ou o teste não paramétrico de Friedman, seguidos de análises *post-hoc* de Tukey ou Durbin–Conover, com o objetivo de verificar a significância das diferenças observadas entre os modelos avaliados. Testes t pareados também foram utilizados para avaliar o impacto das estratégias de aumento de dados nas arquiteturas com melhor desempenho.

3. Resultados e Discussão

Os experimentos foram conduzidos considerando os cenários com e sem DA para os *datasets* OCDC e OralEpitheliumDB. A Tabela 1 detalha o coeficiente Dice médio e o respectivo desvio padrão para todas as arquiteturas, com os melhores desempenhos globais destacados em negrito.

Tabela 1. Coeficiente Dice médio das arquiteturas nos *datasets* OCDC e OralEpitheliumDB. Valores em negrito indicam o melhor desempenho dentre todos os cenários avaliados.

Arquitetura	OCDC Sem DA	OCDC Com DA	OralEpitheliumDB Sem DA	OralEpitheliumDB Com DA
U-Net (MobileNet)	$0,904 \pm 0,002$	$\mathbf{0,912 \pm 0,002}$	$0,832 \pm 0,003$	$\mathbf{0,871 \pm 0,009}$
U-Net (EfficientNet-B0)	$0,895 \pm 0,007$	$\mathbf{0,910 \pm 0,002}$	$0,833 \pm 0,007$	$\mathbf{0,868 \pm 0,004}$
U-Net (Base)	$0,858 \pm 0,037$	$0,847 \pm 0,021$	$0,828 \pm 0,006$	$0,852 \pm 0,003$
Attention U-Net	$0,850 \pm 0,020$	$0,862 \pm 0,013$	$0,824 \pm 0,008$	$0,839 \pm 0,008$
Sharp U-Net	$0,874 \pm 0,008$	$0,873 \pm 0,008$	$0,826 \pm 0,005$	$0,850 \pm 0,005$
U-Net++	$0,866 \pm 0,014$	$0,864 \pm 0,015$	$0,825 \pm 0,006$	$0,839 \pm 0,008$

Em relação aos modelos investigados, observou-se uma diferença de desempenho com base em suas estruturas e na forma de treinamento. Arquiteturas pré-treinadas, como a U-Net combinada aos *backbones* MobileNetV2 e EfficientNet-B0, alcançaram os maiores valores de desempenho em quase todos os cenários, exibindo coeficientes Dice de 0,912 e 0,871 para as bases OCDC e OralEpitheliumDB, respectivamente, sem uso de DA. Por outro lado, arquiteturas de CNN com maior profundidade e complexidade paramétrica, ou treinadas *from scratch*, como a U-Net++ e a Attention U-Net, apresentaram valores inferiores, com a U-Net (Base) apresentando redução entre 0,48% e 7,13%. Essa

diferença no desempenho está relacionada à capacidade de representação do modelo e aos dados disponíveis. Arquiteturas com *backbones* pré-treinados aproveitam os pesos do ImageNet para mitigar os efeitos da escassez de dados. A U-Net MobileNet destacou-se por utilizar convoluções separáveis, que reduzem a carga de parâmetros sem sacrificar a eficiência. Por outro lado, modelos com mecanismos densos ou de atenção possuem um número elevado de parâmetros, o que os torna suscetíveis ao *overfitting* em bases restritas.

Além das diferenças arquiteturais, o desempenho foi influenciado pelas diferenças estruturais entre os *datasets*. O OCDC é caracterizado por padrões morfológicos amplos e segmentação em nível de regiões de lesão, que apresentam características visualmente distintas de regiões saudáveis ou tecidos adjacentes. Já o OralEpitheliumDB envolve estruturas celulares internas a regiões de lesão, apresentando bordas menos definidas e alterações morfológicas complexas, tornando as predições mais sensíveis ao coeficiente Dice. Diante dessas limitações, o desempenho inicial foi homogêneo entre as CNNs, com os modelos U-Net MobileNet e EfficientNet-B0 atingindo valores relevantes apenas após a aplicação do DA, que, combinado ao *transfer learning*, proporcionou a variabilidade para a delimitação de contornos desafiadores.

Em relação à análise estatística no *dataset* OCDC, o cenário sem aumento de dados revelou diferença significativa entre as arquiteturas, com $\chi^2 = 31,3$ e $p < 0,001$. A U-Net MobileNet e a U-Net EfficientNet-B0 apresentaram valores significantes em relação às demais opções com $p < 0,001$, e sem diferença significativa entre si, com $p = 0,117$. A aplicação do aumento de dados manteve a disparidade global, com estatística $F = 40,7$ para 5 e 35 graus de liberdade e $p < 0,001$. O teste de Tukey confirmou que a MobileNet e a EfficientNet-B0 foram superiores aos modelos *from scratch*, com $p < 0,001$, e permanecendo equivalentes entre si ($p = 0,567$). A avaliação pareada confirmou ganhos relevantes do aumento de dados para ambos os modelos, com t de $10,9$ para a MobileNet e $5,62$ para a EfficientNet-B0, ambas considerando 7 graus de liberdade e $p < 0,001$.

No *dataset* OralEpitheliumDB, embora a ANOVA e o teste de Friedman tenham indicado significância global, com $F = 3,14$ e $p = 0,019$, além de $\chi^2 = 13,1$, as médias de Dice variaram apenas entre $0,824$ e $0,833$. Contudo, o aumento de dados provocou uma separação entre os modelos, atestada por $\chi^2 = 33,1$ e $p < 0,001$. Sob essa condição, a U-Net MobileNet e a U-Net EfficientNet-B0 superaram significativamente as arquiteturas sem *transfer learning*. O impacto do aumento de dados foi comprovado pelos testes pareados, com estatísticas t de $10,6$ para a MobileNet e $12,9$ para a EfficientNet-B0, ambas considerando 7 graus de liberdade e $p < 0,001$.

Para complementar a avaliação, nesta etapa foi explorada uma análise visual dos resultados. Nas Figuras 3 e 4 são demonstrados os resultados qualitativos para os *datasets* OCDC e OralEpitheliumDB de imagens histológicas da cavidade oral. Com base nos melhores resultados quantitativos, as análises focam nas três arquiteturas que obtiveram os melhores desempenhos: a U-Net padrão, a U-Net EfficientNet-B0 e a U-Net MobileNet-V2. O objetivo foi analisar o comportamento desses modelos, contrastando seus perfis de erro com e sem a aplicação de técnicas de aumento de dados em relação ao padrão-ouro. Marcações visuais foram inseridas nas predições: círculos contínuos em cor laranja destacam as regiões de falsos positivos, enquanto círculos tracejados em cor roxa evidenciam as áreas de falsos negativos. Os modelos com *backbones* EfficientNet-B0 e

MobileNet-V2 são referenciados pelas abreviações U-Net Eff e U-Net Mob, respectivamente.

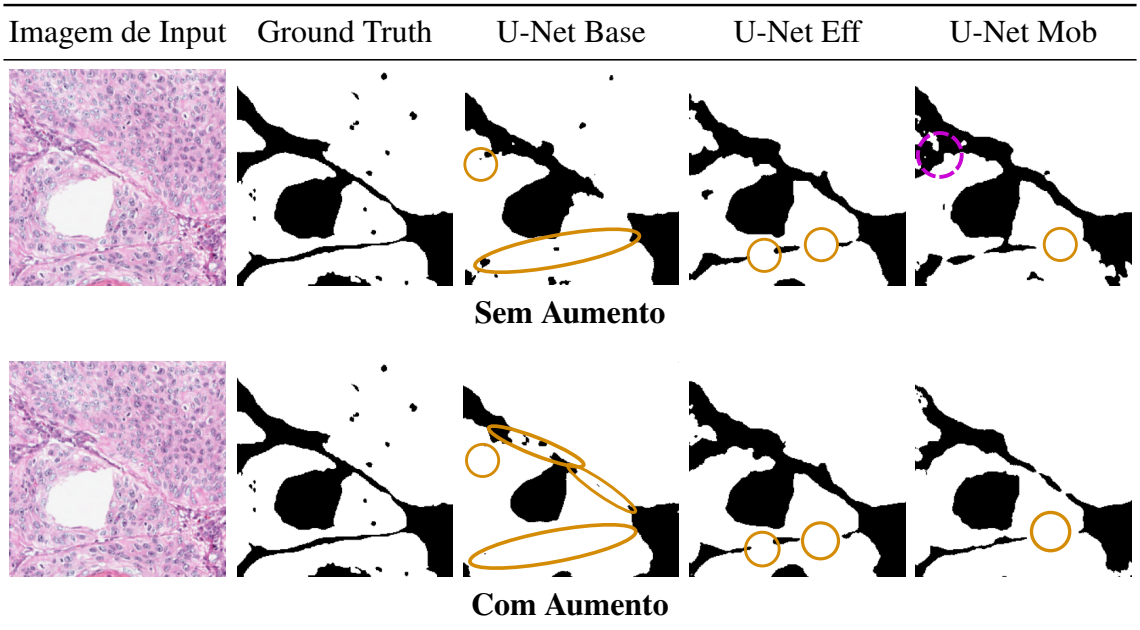

Figura 3. Comparação qualitativa dos modelos de segmentação para carcinoma, demonstrando a presença de falsos positivos e negativos com e sem aumento de dados.

Na avaliação das predições geradas sem o uso de aumento de dados (ver Figura 3), o modelo U-Net Base padrão exibe a maior incidência de falsos positivos entre as arquiteturas testadas, fenômeno evidenciado principalmente nas duas regiões destacadas pelos círculos de cor laranja. A introdução do *backbone* EfficientNet-B0 atenua essas rotulações de fundo indevidas na região, enquanto a U-Net MobileNet-V2 demonstra a menor ocorrência nesse contexto. Em relação aos falsos negativos, as falhas são menores, com apenas o modelo MobileNet-V2 apresentando uma omissão sutil no canto esquerdo da amostra (círculo roxo). A aplicação de aumento de dados revela mudanças no comportamento dos modelos. Para esta amostra, a U-Net padrão apresentou mais falhas na detecção de regiões marcadas no padrão-ouro. Em contrapartida, a U-Net EfficientNet-B0 demonstrou uma redução substancial dessas áreas mantendo o comportamento sem aumento de dados. O modelo MobileNet-V2 registrou apenas uma pequena área nesse mesmo contorno crítico e minimizou a região falso positivo. De modo geral, a análise qualitativa de regiões de tecido OCDC indica que as arquiteturas têm maior dificuldade em delimitar bordas finas e contornos irregulares. Além disso, mesmo com o treinamento enriquecido pelo aumento de dados, todos os modelos mostram falhas, indicando que características morfológicas muito específicas dessas regiões continuam desafiando a precisão de detalhes.

Na análise visual das imagens da base OralEpitheliumDB apresentadas na Figura 4, o modelo U-Net padrão (sem aumento) demonstra dificuldade na identificação de certas estruturas de núcleo, evidenciada pela presença de falso negativo (cor roxa), quadrante superior esquerdo da amostra. A adoção do *backbone* EfficientNetB0 altera as características da região de erro da rede. O modelo U-Net integrado ao *backbone* Mo-

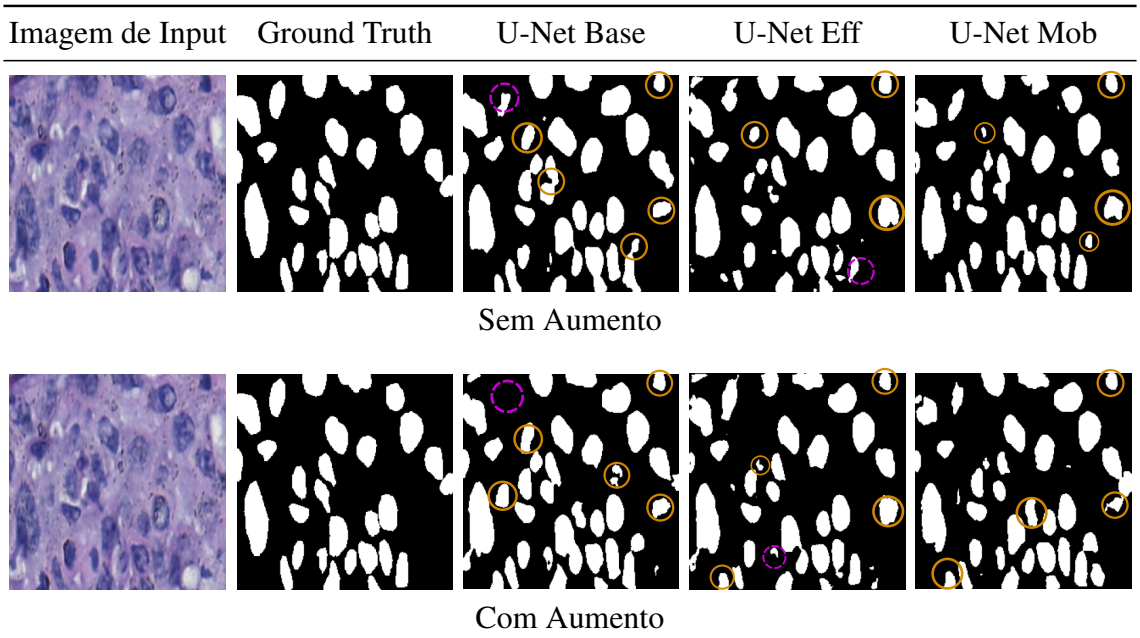

Figura 4. Comparação qualitativa dos modelos de segmentação para displasia, demonstrando a presença de falsos positivos e negativos com e sem aumento de dados.

bileNetV2 apresentou os resultados visualmente mais robustos quanto à completude das máscaras. Os pixels falsos negativos são consideravelmente menos evidentes; embora as bordas de alguns núcleos não estejam perfeitamente delimitadas em 100% de sua área. Apesar dos falsos positivos ainda estarem presentes, é nesta arquitetura que a aplicação do aumento de dados revela seu maior benefício, o modelo MobileNetV2 refina significativamente sua capacidade de separação espacial.

Além da qualidade da segmentação, o uso de *lightweight backbones* traz um ganho na eficiência computacional. A U-Net baseada em EfficientNet-B0 operou com 1,61M de parâmetros e 6,90 GFLOPs, seguida pela variante com MobileNetV2, com 3,73M de parâmetros e 7,08 GFLOPs. Redes estruturalmente mais complexas, como a U-Net++, registraram um consumo muito superior, chegando a 9,16M de parâmetros e 34,90 GFLOPs. Essa diferença mostra que a integração de extratores leves garante uma redução de até 80% nos GFLOPs, mostrando que é possível obter resultados compatitivos com a literatura sem elevar o custo computacional.

A Tabela 2 compara a U-Net MobileNet treinada com aumento de dados, modelo de melhor desempenho neste estudo, com abordagens recentes da literatura que também convergem para o uso da família U-Net nos mesmos *datasets*. No OCDC, o modelo proposto alcançou coeficiente Dice de 0,912, superando as estratégias baseadas em manipulação de dados de [Santos et al. 2023a], com 0,901, e em mecanismos de atenção *Channel Mixing* de [Crepaldi et al. 2025], com 0,911, além de manter-se competitivo com a complexa topologia U-Net++ de [de Oliveira et al. 2025], que obteve 0,918. Simultaneamente, no OralEpitheliumDB, a marca de 0,871 superou os resultados de 0,843 e 0,830 de [Crepaldi et al. 2025] e [de Oliveira et al. 2025], aproximando-se do índice de 0,880 alcançado pelos *ensembles* de [Silva et al. 2024b]. Essa análise conjunta evidencia que, enquanto os trabalhos correlatos tentam aprimorar a generalização por meio de fusão de

modelos, manipulação do espaço de dados ou expansão estrutural, o modelo proposto se destaca por seu desempenho em múltiplos domínios e priorizando a eficiência computacional intrínseca aos *lightweight backbones*.

Tabela 2. Comparação do modelo U-Net (MobileNet) com estudos da literatura nos *datasets* OCDC e OralEpitheliumDB.

Estudo	OCDC	OralEpitheliumDB
[Santos et al. 2023a]	0,901	-
[Silva et al. 2024b]	-	0,880
[Crepaldi et al. 2025]	0,911	0,843
[de Oliveira et al. 2025]	0,918	0,830
U-Net (MobileNet)	0,912	0,871

4. Conclusão

Este trabalho avaliou experimentalmente arquiteturas baseadas na U-Net para segmentação de imagens histológicas da cavidade bucal e tecido tumoral. Foram avalidados modelos tradicionais e variações com acoplamento de versões *lightweight backbones* (MobileNet e EfficientNet-B0). Além disso o aumento de dados também foi uma estratégia relevante usada no trabalho. O estudo ressalta que quando analisado o *dataset* OCDC houve diferença significativa entre as arquiteturas, com U-Net MobileNet e U-Net EfficientNet-B0 tendo os melhores resultados nesse ponto. Da mesma forma, a U-Net MobileNet e U-Net Efficiente-B0 demonstraram melhora quando aplicado o aumento de dados sem diferença significativa entre os dois *backbones*. Quando examinado o banco OralEpitheliumDB foram obtidos resultados mais homogêneos entre as arquiteturas com o coeficiente Dice tendo pequena variação, mesmo com a ANOVA e Friedman indicando significativas diferenças globais. Já com o uso de aumento de dados é possível observar um distanciamento dos modelos com abordagem *scratch*. O aumento de dados evidenciou e corroborou para que o desempenho da segmentação fosse maximizado, melhorando a capacidade do modelo de generalizar, o que também ressalta a relevância de técnicas que atuam aumentando a diversidade de conjuntos de treinamento, sobretudo em domínios médicos com limitação na quantidade de amostras. Em estudos futuros pretende-se empregar estratégias mais robustas de validação, como validação cruzada, bem como abordagens sistemáticas de otimização de hiperparâmetros, visando uma avaliação mais abrangente dos modelos.

5. Agradecimentos

Os autores agradecem o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processos nº 305386/2024-7 e nº 302833/2025-0) e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (Processos nº APQ-00727-24 e nº APD-01428-25).

Referências

Banerji, S. and Mitra, S. (2022). Deep learning in histopathology: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1):e1439.

Basu, A., Senapati, P., Deb, M., Rai, R., and Dhal, K. G. (2024). A survey on recent trends in deep learning for nucleus segmentation from histopathology images. *Evolving Systems*, 15(1):203–248.

Cheplygina, V. e. a. (2019). Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis*.

Crepaldi, G. G., de Oliveira, D. L., Tosta, T. A., Neves, L. A., Silva, A. B., Martins, A. S., and do Nascimento, M. Z. (2025). Nuclear segmentation in histological images using multiple attention system mixing. In *2025 IEEE 38th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 154–159. IEEE.

de Oliveira, D. L., Tosta, T. A., Neves, L. A., Silva, A. B., Martins, A. S., de Fariall, P. R., and do Nascimento, M. Z. (2025). A u-net-based approach for histological tissue segmentation using rcaug data augmentation. In *2025 38th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 1–6. IEEE.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

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*.

Litjens, G. e. a. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*.

Madabhushi, A. and Lee, G. (2016). Image analysis and machine learning in digital pathology. *Medical Image Analysis*.

Moscalu, M., Moscalu, R., Dascălu, C. G., Țarcă, V., Cojocaru, E., Costin, I. M., Țarcă, E., and Șerban, I. L. (2023). Histopathological images analysis and predictive modeling implemented in digital pathology—current affairs and perspectives. *Diagnostics*, 13(14):2379.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.

Santos, D. F., de Faria, P. R., Travençolo, B. A., and do Nascimento, M. Z. (2023a). Influence of data augmentation strategies on the segmentation of oral histological images using fully convolutional neural networks. *Journal of Digital Imaging*, 36(4):1608–1623.

Santos, D. F. D., de Faria, P. R., Loyola, A. M., Cardoso, S. V., Travençolo, B. A. N., and do Nascimento, M. Z. (2023b). Hematoxylin and eosin stained oral squamous cell carcinoma histological images dataset.

Santos, M. d. O. (2023). Estimativa de incidência de câncer no brasil, 2023-2025. *Revista Brasileira de Cancerologia*, 69(1).

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*.

Silva, A. B., Martins, A. S., Tosta, T. A. A., Loyola, A. M., Cardoso, S. V., Neves, L. A., de Faria, P. R., and do Nascimento, M. Z. (2024a). Oralepitheliumdb: A dataset for oral epithelial dysplasia image segmentation and classification. *Journal of Imaging Informatics in Medicine*, 37(4):1691–1710.

Silva, A. B., Rozendo, G. B., Tosta, T. A., Martins, A. S., Loyola, A. M., Cardoso, S. V., Lumini, A., Neves, L. A., de Faria, P. R., and do Nascimento, M. Z. (2023). Cnn ensembles for nuclei segmentation on histological images of oed. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 601–604. IEEE.

Silva, A. B., Tosta, T. A., Neves, L. A., Martins, A. S., De Faria, P. R., and Do Nascimento, M. Z. (2024b). Ensemble of semantic segmentation models for oral epithelial dysplasia images. In *2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 1–6. IEEE.

Springenberg, M., Frommholz, A., Wenzel, M., Weicken, E., Ma, J., and Strodthoff, N. (2023). From modern cnns to vision transformers: Assessing the performance, robustness, and classification strategies of deep learning models in histopathology. *Medical Image Analysis*, 87:102809.

Srinidhi, C. L., Ciga, O., and Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical image analysis*, 67:101813.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., and Liu, F. (2024). Advances in medical image segmentation: a comprehensive review of traditional, deep learning and hybrid approaches. *Bioengineering*, 11(10):1034.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4*, pages 3–11. Springer.

Zunair, H. and Hamza, A. B. (2021). Sharp u-net: Depthwise convolutional network for biomedical image segmentation. *Computers in biology and medicine*, 136:104699.