

Investigação de Arquiteturas de Aprendizagem Profunda para Segmentação de Lesões em Imagens Histológicas

Domingos L. L. de Oliveira^{1,2}, Adriano B. Silva², Leandro A. Neves³,
Thaína A. A. Tosta⁴, Alessandro S. Martins⁵, Paulo R. de Faria⁶,
Marcelo Z. do Nascimento²

¹Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP)
São Paulo – SP – Brasil

²Faculdade de Computação (FACOM) – Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brasil

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

⁴Instituto de Ciência e Tecnologia (ICT), Universidade Federal de São Paulo - UNIFESP
São José dos Campos – SP – Brasil

⁵Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro (IFTM)
Ituiutaba – MG – Brasil

⁶Faculdade de Odontologia (FO) - Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brasil

domingos.oliveira@ifsp.edu.br

Abstract. *This study evaluates the performance of convolutional neural networks (CNN)-based models and Transformers for histological image segmentation using public datasets: OCDC and GlaS. Four models were analyzed: UNet, UNet++, SharpUNet, and TransUNet, with and without data augmentation. Data augmentation proved crucial for improving model generalization, with SharpUNet and UNet++ achieving the best results in Dice coefficient and accuracy. TransUNet, despite its hybrid architecture, underperformed, possibly due to its complexity and need for large training data. The results indicate that CNN-based models, such as UNet++ and SharpUNet, are effective for histological image segmentation, especially when combined with data augmentation.*

Resumo. *Este estudo avalia o desempenho de modelos baseados em redes neurais convolucionais (CNN) e Transformers para segmentação de imagens histológicas, utilizando os conjuntos de dados públicos OCDC e GlaS. Quatro modelos foram analisados: UNet, UNet++, SharpUNet e TransUNet, com e sem aumento de dados. O aumento de dados mostrou-se crucial para melhorar a generalização dos modelos, com SharpUNet e UNet++ alcançando os melhores resultados em coeficiente Dice e precisão. O TransUNet, apesar de sua arquitetura híbrida, teve desempenho inferior, possivelmente devido à sua complexidade e necessidade de grandes volumes de dados. Os resultados indicam que modelos baseados em CNN, como UNet++ e SharpUNet, são eficazes na segmentação de imagens histológicas, especialmente quando combinados com aumento de dados.*

1. Introdução

A segmentação de imagens histológicas desempenha um papel fundamental no diagnóstico assistido por computador (do inglês, *Computer-Aided Diagnosis - CAD*) [Moscalu et al. 2023], permitindo a detecção de estruturas celulares e teciduais em imagens médicas [Basu et al. 2024, Xu et al. 2024]. Na histopatologia, a crescente demanda por técnicas automatizadas de segmentação reflete a importância dessas abordagens na complementação da análise clínica e na categorização de tumores [Banerji and Mitra 2022]. Entre as abordagens existentes, a segmentação semântica tem se destacado por possibilitar a diferenciação automática de tecidos normais e anômalos em imagens histológicas, podendo auxiliar na identificação precoce e na quantificação de padrões histopatológicos relevantes [Silva et al. 2023, Silva et al. 2022]. No entanto, desafios persistem na área, incluindo a variabilidade morfológica das amostras, a escassez de dados anotados e a generalização dos modelos para diferentes condições patológicas [Banerji and Mitra 2022].

Redes neurais convolucionais (do inglês, *Convolutional Neural Networks - CNNs*) têm sido amplamente utilizadas para segmentação e classificação de imagens médicas [Srinidhi et al. 2021, Springenberg et al. 2023]. No contexto de segmentação semântica, a UNet [Ronneberger et al. 2015] tem sido uma das arquiteturas mais populares devido à sua capacidade de representar tanto informações de baixo nível (bordas e texturas) quanto de alto nível (formas e padrões globais). Com esse avanço, novos aprimoramentos surgiram como UNet++ [Zhou et al. 2018] e SharpUNet [Zunair and Hamza 2021]. Esses foram propostos para aprimorar a qualidade da segmentação, introduzindo conexões densas e mecanismos de realce de bordas. No entanto, modelos baseados em CNNs apresentam algumas limitações que impactam sua aplicação na segmentação de imagens médicas. Um dos principais desafios está na dificuldade de capturar relações espaciais de longo alcance [Maurício et al. 2023], o que pode comprometer a segmentação de estruturas complexas e altamente heterogêneas, como tecidos tumorais irregulares. Outra limitação é a tendência de CNNs dependerem de janelas locais de convolução, o que pode resultar na perda de informações globais essenciais para uma segmentação mais precisa.

Com o avanço dos Transformers, em visão computacional [Dosovitskiy et al. 2020], arquiteturas híbridas como a TransUNet [Chen et al. 2021] foram desenvolvidas para combinar os pontos fortes das convoluções com a capacidade dos mecanismos de autoatenção de capturar dependências entre regiões distantes da imagem. No entanto, essas abordagens frequentemente requerem grandes volumes de dados para treinamento eficaz [Xu et al. 2023], o que representa um desafio na área médica, onde a obtenção de imagens rotuladas é limitada por fatores como privacidade, custo e variabilidade interobservador [Banerji and Mitra 2022]. Assim, a aplicabilidade dessas arquiteturas em imagens histológicas ainda é um desafio, especialmente quando comparadas a abordagens baseadas exclusivamente em CNNs.

Este estudo apresenta uma investigação de modelos de aprendizagem profunda para a segmentação de imagens histológicas baseada nas arquiteturas UNet [Ronneberger et al. 2015], UNet++ [Zhou et al. 2018], SharpUNet [Zunair and Hamza 2021] e TransUNet [Chen et al. 2021]. A avaliação dos modelos possibilita identificar quais métodos são mais eficazes na segmentação de tecidos e células em imagens histológicas, fornecendo subsídios para o desenvolvimento de

soluções mais robustas em sistemas CAD. Essa pesquisa busca avaliar as abordagens utilizando as métricas coeficiente *Dice*, acurácia, sensibilidade e especificidade. O desempenho dos modelos e o impacto da estratégia de aumento de dados também foram investigados em dois conjuntos de dados: OCDC [Santos et al. 2023b] e GlaS [Sirinukunwattana et al. 2017].

2. Materiais e Métodos

A metodologia proposta para a investigação das abordagens de segmentação de imagens histológicas é ilustrada na Figura 1. O *pipeline* consistiu na utilização de duas bases de imagens de domínio público, cada uma dividida em conjuntos de treinamento, validação e teste. Também foram investigadas técnicas de aumento de dados para aprimorar a generalização dos modelos. Diferentes arquiteturas baseadas na UNet foram treinadas e avaliadas. As segmentações geradas foram comparadas com o padrão ouro marcado por especialistas, e métricas foram calculadas para avaliar o desempenho dos modelos. Os detalhes de cada etapa do método são descritos nas próximas subseções.

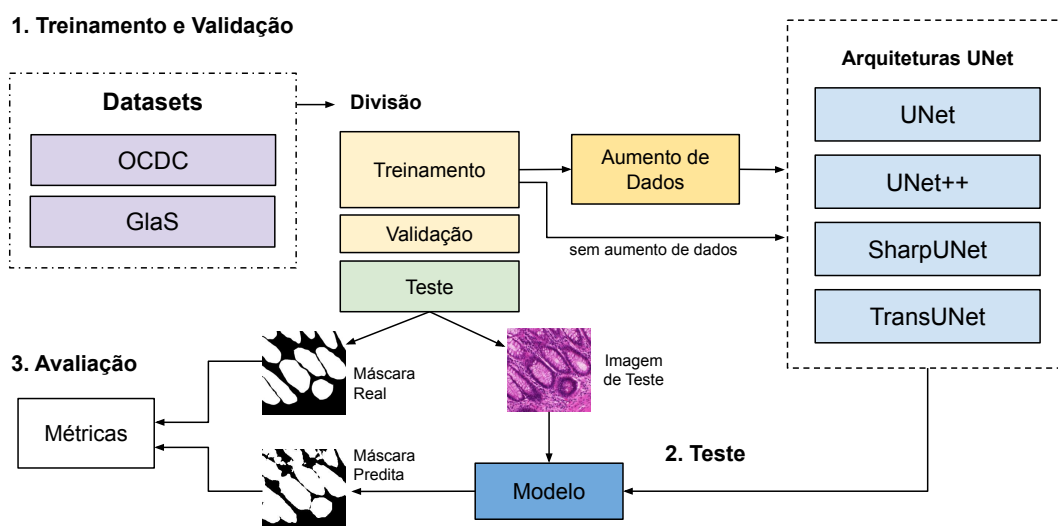


Figura 1. Pipeline da metodologia proposta, incluindo as etapas de divisão dos dados, treinamento com e sem aumento de dados, e avaliação dos modelos UNet, UNet++, SharpUNet e TransUNet nos datasets OCDC e GlaS.

2.1. Base de Imagens

Neste estudo, foram empregadas as seguintes bases de imagens (*datasets*): a) *Oral Cavity-Derived Cancer* (OCDC) [Santos et al. 2023b], que contém 508 imagens histológicas de carcinomas de células escamosas orais e tecidos normais, capturadas com ampliação de 20×. As imagens (640×640 pixels) foram anotadas por especialistas e divididas em 393 para treinamento e 61 para teste. No presente estudo, foi utilizada uma subdivisão composta por 278 imagens para treinamento, 115 para validação e 61 para teste. A base é empregada para treinar modelos de CNNs na segmentação e diagnóstico do câncer oral (Figura 2 (A)). b) A base *GlaS* [Sirinukunwattana et al. 2017] contém 165 imagens histológicas do cólon coradas com H&E, obtidas de adenocarcinomas colorretais em diferentes pacientes (Figura 2 (B)). As imagens possuem anotações manuais de glândulas,

servindo como referência para segmentação automática. A divisão adotada neste estudo compreende 60 imagens para treinamento, 25 para validação e 80 para teste.

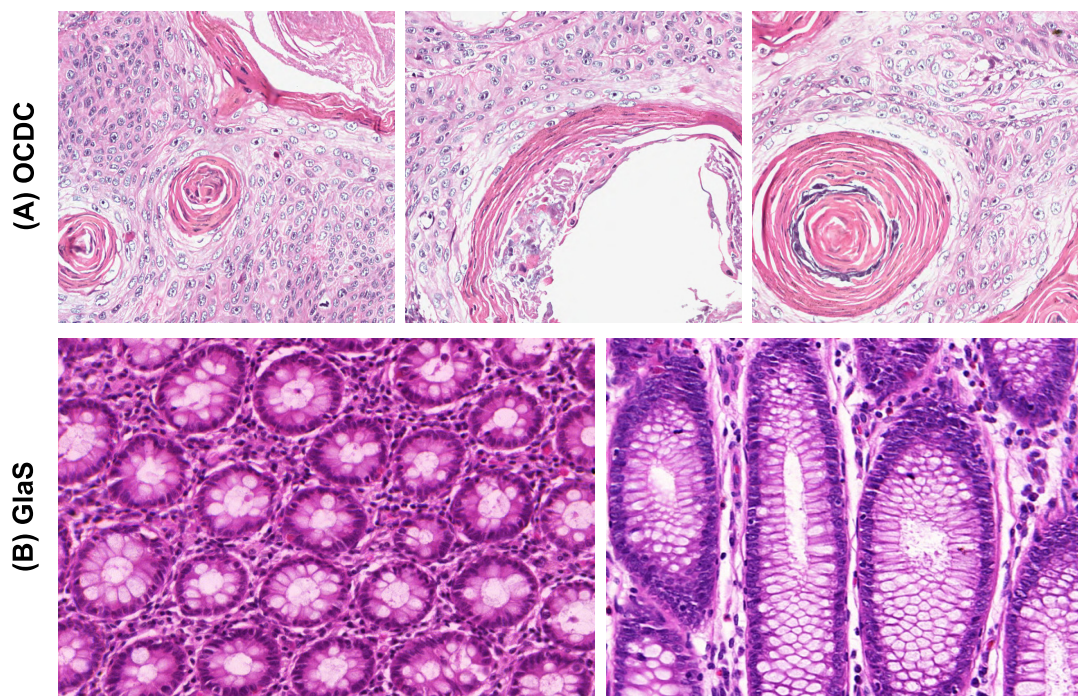


Figura 2. Exemplos de imagens histológicas utilizadas no estudo: (A) imagens da base OCDC com regiões tumorais da cavidade oral; (B) imagens da base GlaS contendo glândulas colorretais.

2.2. Segmentação dos Tecidos Histológicos

Para a análise proposta neste estudo, foram investigadas as arquiteturas UNet, UNet++, SharpUNet e TransUNet. Este conjunto é composto por modelos clássicos, leves e híbridos. Essa diversidade de arquiteturas contribui para mitigar vieses inerentes a modelos específicos e permite uma avaliação mais robusta da eficácia das técnicas de aumento de dados

A UNet [Ronneberger et al. 2015] é uma arquitetura em formato de “U”, que combina um caminho de contração, responsável pela extração de características, e um caminho de expansão, que refina a segmentação (Figura 3 (A)). A UNet++ [Zhou et al. 2018] introduz conexões densas entre os blocos da rede, melhorando a propagação de informações e a precisão da segmentação (Figura 3 (B)). O modelo SharpUNet [Zunair and Hamza 2021] incorpora mecanismos de realce de bordas, sendo útil para imagens histológicas onde há transições sutis entre estruturas dos tecidos (Figura 3 (C)).

Com o advento dos Transformers [Vaswani et al. 2017, Dosovitskiy et al. 2020], surgiram modelos híbridos que combinam convoluções com mecanismos de autoatenção para capturar relações de longo alcance nas imagens. A TransUNet [Chen et al. 2021] incorpora um codificador baseado em Transformers na estrutura da UNet, permitindo que o modelo aprenda dependências globais enquanto mantém a capacidade de capturar

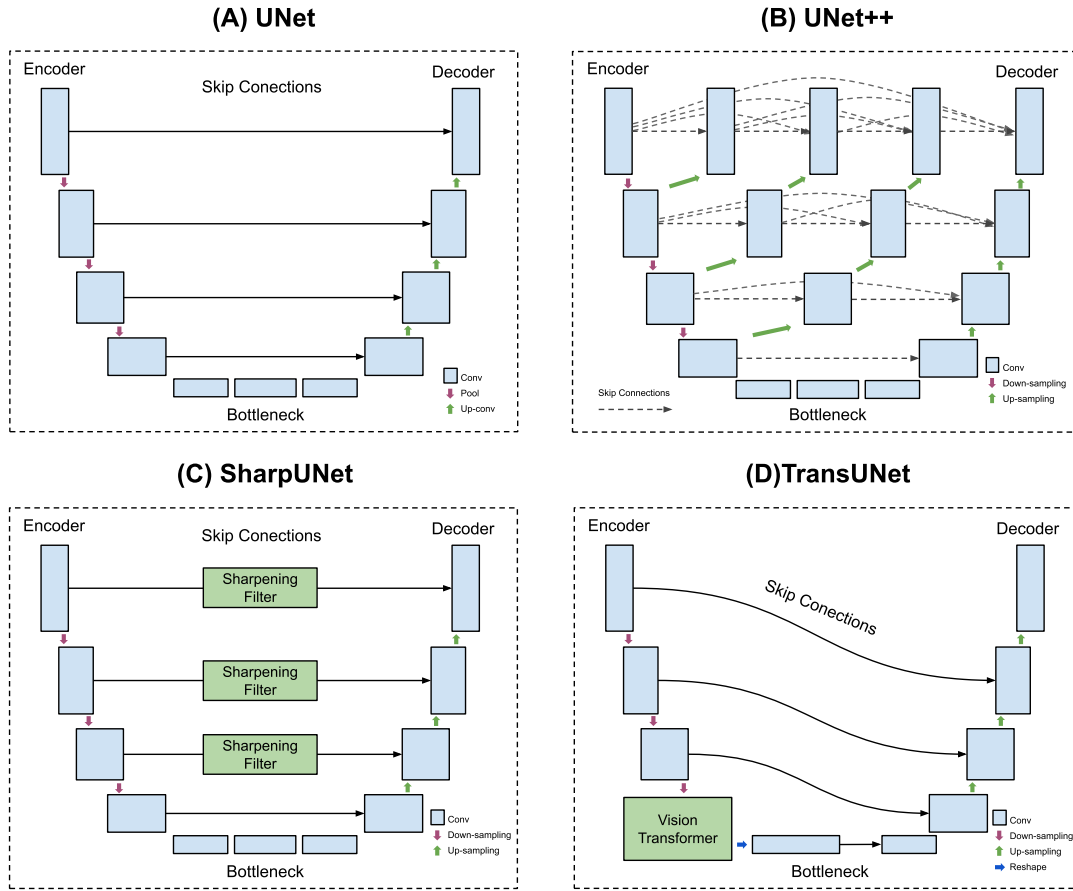


Figura 3. Representações esquemáticas das arquiteturas de segmentação avaliadas: (A) UNet original com caminho de contração e expansão; (B) UNet++ com conexões densas; (C) SharpUNet com módulos de realce de bordas; (D) TransUNet com codificador baseado em Transformer.

detalhes locais por meio das convoluções (Figura 3 (D)). Essa abordagem tem se mostrado eficaz na segmentação de imagens médicas, especialmente em cenários onde estruturas anatômicas apresentam grande variação espacial.

O aumento de dados é uma etapa essencial para melhorar a capacidade de generalização dos modelos e reduzir o risco de *overfitting*, especialmente em conjuntos de dados médicos, que costumam ser limitados em tamanho. Neste estudo, foram avaliadas três estratégias distintas de aumento de dados. A primeira estratégia seguiu a abordagem proposta por [Santos et al. 2023a], que combina transformações geométricas e deformações não lineares. A segunda estratégia empregou exclusivamente transformações geométricas, incluindo *flip* horizontal e vertical, rotação aleatória de até 90° e transposição. A terceira estratégia utilizou técnicas de distorção da imagem, como distorção elástica, de grade e óptica. Para cada imagem original do conjunto de treinamento, foram geradas três versões aumentadas, correspondentes a cada uma das três estratégias adotadas.

Os hiperparâmetros do treinamento foram ajustados de forma empírica para cada modelo e *dataset*, conforme mostrado na Tabela 1. O número de épocas foi definido em

150 com base na convergência das métricas coeficiente *Dice* e acurácia no conjunto de validação. Para evitar *overfitting*, foi usada uma taxa de *dropout* entre 0,0 e 0,3. Na otimização, foi empregado o Adam, combinando *Binary Cross-Entropy* com *Logits Loss* e *Dice Loss*. Os treinamentos foram realizados em ambiente Linux com suporte a CUDA e PyTorch 2.0, utilizando uma GPU NVIDIA RTX 4070 Ti SUPER (16 GB VRAM).

Tabela 1. Hiperparâmetros de treinamento para cada modelo e *dataset*.

<i>Dataset</i>	<i>Model</i>	<i>Batch Size</i>	<i>Learning Rate</i>	<i>Epochs</i>	<i>Dropout</i>
GlaS	UNet, UNet++, SharpUNet	8	0,001	150	0,2
	TransUNet	8	0,0001	150	0,1
OCDC	UNet, UNet++, SharpUNet	8	0,0005	150	0,0
	TransUNet	8	0,0001	150	0,1

O tempo de inferência permaneceu constante entre os conjuntos de dados para cada modelo, sendo de 0,3 ms para SharpUNet e UNet, 0,5 ms para UNet++ e 0,9 ms para TransUNet. Já o tempo de treinamento variou conforme a quantidade de amostras em cada *dataset*, com os menores tempos observados para o GlaS (SharpUNet: 2,2 min, TransUNet: 2,3 min, UNet: 2,1 min, UNet++: 5,9 min) e os maiores tempos para o OCDC (SharpUNet: 8,5 min, TransUNet: 8,9 min, UNet: 8,2 min, UNet++: 24,1 min).

2.3. Métricas de Avaliação e Análise Estatística

Para avaliar o desempenho dos modelos de segmentação foram utilizadas as métricas: coeficiente *Dice*, Acurácia, Sensibilidade e Especificidade. Essas métricas quantificam a qualidade das segmentações ao comparar as previsões do modelo com as máscaras do padrão ouro (anotações manuais), fornecendo uma análise detalhada do desempenho. Além disso, o teste de Friedman foi aplicado separadamente para cada *dataset*, considerando três repetições independentes por configuração.

3. Resultados e Discussão

A Tabela 2 apresenta as métricas de desempenho dos modelos de segmentação no *dataset* GlaS, comparando os resultados com e sem o uso de aumento de dados. Os melhores valores para cada métrica estão destacados em negrito. Observa-se que a aplicação de aumento de dados levou a melhorias expressivas, especialmente para o coeficiente *Dice* e a acurácia. O modelo SharpUNet com aumento de dados obteve os maiores valores para *Dice* ($89,47 \pm 8,82$) e acurácia ($86,34 \pm 9,17$). A maior especificidade foi alcançada pelo UNet++ com aumento ($89,76 \pm 9,22$), enquanto a maior sensibilidade foi observada no TransUNet com aumento ($91,95 \pm 8,19$), indicando uma melhor segmentação das regiões de interesse.

Na Figura 4 são ilustrados os resultados dos modelos avaliados, evidenciando os problemas de segmentação. Os valores de especificidade indicam que a UNet++ (89,76%) e a SharpUNet (89,26%) foram mais eficazes em evitar a segmentação indevida de áreas que não pertencem ao objeto de interesse, apresentando menor incidência de falsos positivos. Por outro lado, a UNet (86,87%) e a TransUNet (85,02%) demonstraram maior propensão a identificar erroneamente regiões do fundo como parte do objeto segmentado.

Tabela 2. Comparação dos modelos UNet, UNet++, SharpUNet e TransUNet no *dataset* GlaS [Sirinukunwattana et al. 2017] com e sem aumento de dados.

Modelo	Dice (%)	Acurácia (%)	Sensibilidade (%)	Especificidade (%)
UNet [†]	89,05 ± 7,98	85,94 ± 8,39	91,39 ± 10,27	86,87 ± 10,23
UNet [△]	86,85 ± 8,90	83,29 ± 9,89	<u>91,52 ± 11,37</u>	80,50 ± 16,79
UNet [◇]	85,98 ± 9,88	83,01 ± 8,52	<u>89,97 ± 9,82</u>	82,50 ± 13,68
UNet	82,98 ± 10,61	79,58 ± 9,96	89,64 ± 12,35	75,73 ± 19,35
UNet++ [†]	89,21 ± 8,13	86,16 ± 8,32	89,57 ± 11,08	89,76 ± 9,22
UNet++ [△]	88,76 ± 7,60	<u>85,69 ± 7,31</u>	90,86 ± 8,32	87,82 ± 10,47
UNet++ [◇]	87,24 ± 8,30	84,32 ± 7,85	88,80 ± 10,26	86,76 ± 10,90
UNet++	83,64 ± 10,72	80,00 ± 9,67	<u>91,48 ± 9,58</u>	74,98 ± 18,75
SharpUNet [†]	89,47 ± 8,82	86,34 ± 9,17	90,59 ± 11,53	89,26 ± 9,90
SharpUNet [△]	89,30 ± 8,52	86,11 ± 8,76	<u>90,67 ± 11,55</u>	88,62 ± 10,31
SharpUNet [◇]	88,58 ± 7,87	85,21 ± 8,42	<u>90,07 ± 10,08</u>	87,17 ± 12,45
SharpUNet	86,36 ± 8,54	82,60 ± 9,60	90,54 ± 10,57	80,63 ± 17,15
TransUNet [†]	88,42 ± 7,12	<u>85,00 ± 7,65</u>	91,23 ± 9,06	85,02 ± 13,12
TransUNet [△]	88,27 ± 7,18	84,70 ± 8,11	91,95 ± 8,19	83,38 ± 15,12
TransUNet [◇]	86,90 ± 7,68	83,40 ± 8,64	89,40 ± 10,25	83,77 ± 14,17
TransUNet	85,60 ± 7,99	82,03 ± 9,14	89,53 ± 10,19	80,70 ± 17,16

[†] Modelos treinados com aumento de dados proposto por [Santos et al. 2023a].

[△] Modelos treinados com aumento geométrico.

[◇] Modelos treinados com aumento por distorção.

Valores em negrito indicam o melhor desempenho entre todos os modelos para cada métrica, enquanto valores sublinhados representam o melhor desempenho dentro de cada arquitetura específica.

Esse comportamento é visível na Figura 4 (A), especialmente nas marcações em cor verde, onde a UNet e a TransUNet apresentam regiões indevidamente segmentadas. Além disso, a TransUNet demonstrou maior propensão a identificar erroneamente regiões do objeto segmentado como parte do fundo, conforme identificado em cor verde nas Figura 4 (B) e 4 (C). A superioridade da UNet++ e SharpUNet nesse aspecto pode estar relacionada à presença de conexões densas e mecanismos de realce de bordas, que contribuem para uma melhor discriminação entre estruturas relevantes e artefatos na imagem.

A Tabela 3 apresenta os resultados obtidos no *dataset* OCDC. O UNet++ com aumento alcançou o maior *Dice* (91,81 ± 5,87) e a maior acurácia (91,49 ± 4,12). O UNet com aumento com transformações geométricas teve a maior sensibilidade (94,45 ± 6,98), enquanto a SharpUNet com aumento com transformações de distorção apresentou a maior especificidade (90,87 ± 14,69).

Na Figura 5 são apresentados os resultados da segmentação no *dataset* OCDC. As áreas destacadas em cor verde evidenciam falhas na segmentação, como falsos positivos e falsos negativos. Observa-se que a segmentação é desafiada pela presença de regiões de fundo dentro das áreas segmentadas corretamente, o que resulta na inclusão indesejada de estruturas na região alvo. Além disso, algumas áreas segmentadas possuem texturas similares ao fundo, tornando difícil a distinção clara entre as estruturas de interesse e o tecido ao redor. Essa limitação pode estar relacionada à necessidade de aprimoramento nas técnicas de pré-processamento e aumento de dados, de forma a tornar os modelos

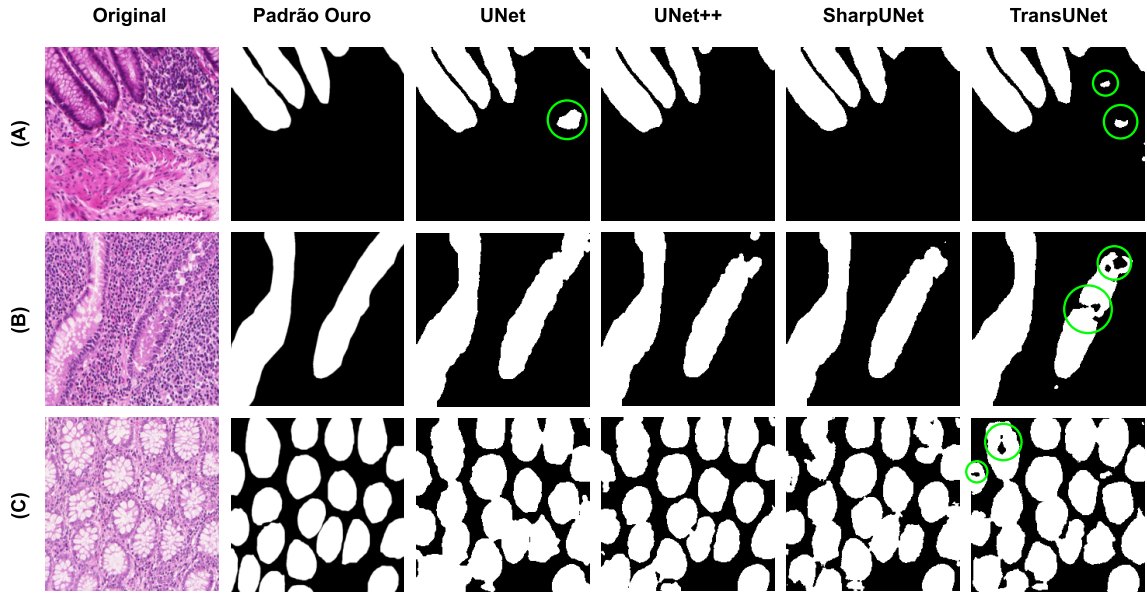


Figura 4. Comparação visual dos resultados de segmentação para amostras do *dataset* GlaS. As áreas destacadas em cor verde indicam problemas na segmentação, como falsos positivos (UNet e TransUNet (A)) e falsos negativos (TransUNet (B) e (C)).

Tabela 3. Comparação dos modelos UNet, UNet++, SharpUNet e TransUNet no *dataset* OCDC [Santos et al. 2023b] com e sem aumento de dados.

Modelo	Dice (%)	Acurácia (%)	Sensibilidade (%)	Especificidade (%)
UNet [†]	90,93 ± 9,00	91,30 ± 3,92	94,20 ± 10,70	88,01 ± 16,43
UNet [△]	<u>91,54 ± 5,76</u>	<u>91,41 ± 3,73</u>	94,45 ± 6,98	88,57 ± 15,74
UNet [◇]	89,39 ± 13,08	<u>90,82 ± 4,00</u>	<u>91,08 ± 13,77</u>	<u>89,16 ± 15,39</u>
UNet	88,41 ± 9,98	89,61 ± 4,80	91,37 ± 12,39	86,00 ± 16,69
UNet++ [†]	91,81 ± 5,87	91,40 ± 4,12	93,49 ± 6,85	89,51 ± 15,92
UNet++ [△]	91,61 ± 6,21	91,49 ± 3,60	<u>94,03 ± 7,95</u>	89,42 ± 14,26
UNet++ [◇]	89,85 ± 12,87	<u>90,93 ± 4,90</u>	<u>90,19 ± 14,01</u>	90,87 ± 14,69
UNet++	89,14 ± 9,12	90,04 ± 4,22	92,58 ± 10,93	86,62 ± 16,27
SharpUNet [†]	<u>91,00 ± 5,75</u>	<u>91,07 ± 3,95</u>	91,11 ± 7,63	<u>90,63 ± 13,99</u>
SharpUNet [△]	<u>90,59 ± 6,56</u>	<u>90,87 ± 3,97</u>	<u>91,39 ± 8,81</u>	<u>89,81 ± 15,05</u>
SharpUNet [◇]	90,46 ± 8,07	90,99 ± 3,98	<u>90,64 ± 10,49</u>	90,57 ± 13,82
SharpUNet	87,79 ± 8,96	89,36 ± 4,59	87,65 ± 11,58	89,44 ± 12,96
TransUNet [†]	90,54 ± 7,42	90,90 ± 4,10	92,87 ± 8,95	<u>88,13 ± 16,06</u>
TransUNet [△]	<u>90,72 ± 5,92</u>	<u>90,67 ± 4,61</u>	<u>93,44 ± 7,29</u>	<u>86,73 ± 17,98</u>
TransUNet [◇]	89,82 ± 7,21	90,24 ± 4,58	<u>92,57 ± 9,79</u>	86,04 ± 18,24
TransUNet	88,06 ± 9,19	89,20 ± 5,45	90,57 ± 12,01	85,59 ± 18,12

[†] Modelos treinados com aumento de dados proposto por [Santos et al. 2023a].

[△] Modelos treinados com aumento geométrico.

[◇] Modelos treinados com aumento por distorção.

Valores em negrito indicam o melhor desempenho entre todos os modelos para cada métrica, enquanto valores sublinhados representam o melhor desempenho dentro de cada arquitetura específica.

mais robustos a essas variações sutis.

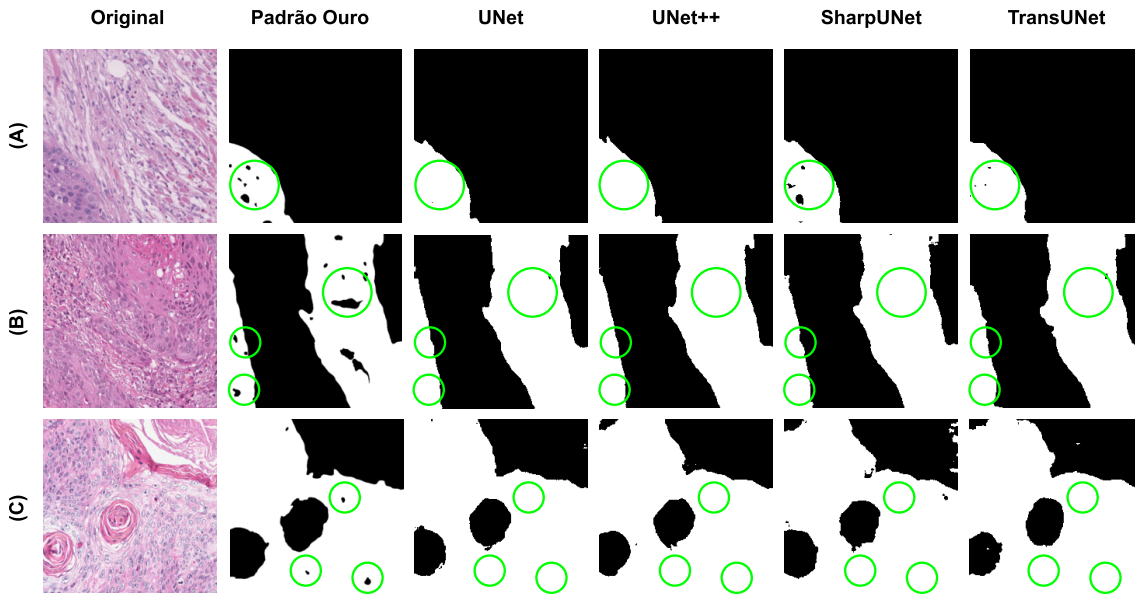


Figura 5. Comparação visual dos resultados de segmentação para amostras do *dataset* OCDC. As áreas destacadas em cor verde na segmentação indicam regiões do padrão ouro ((A), (B) e (C)) definidas como fundo e regiões segmentadas como falsos positivos em todos os métodos e amostras.

Entre os modelos avaliados, o TransUNet apresentou o menor desempenho nos bancos de imagens, possivelmente devido à sua maior complexidade computacional e à necessidade de grandes volumes de dados para generalização. Além disso, há dificuldades em capturar detalhes finos em imagens médicas, resultando em mais falsos positivos e segmentações fragmentadas. Em contrapartida, a UNet++ e a SharpUNet destacaram-se, com a UNet++ alcançando maior especificidade e reduzindo falsos positivos, enquanto a SharpUNet mostrou superioridade na segmentação de bordas e preservação de detalhes morfológicos, atribuída às suas conexões densas e mecanismos de realce de bordas. A UNet clássica, embora mais simples, teve alta sensibilidade, mas desempenho ligeiramente inferior em especificidade.

Os resultados indicam que o aumento de dados melhorou a generalização dos modelos, resultando em segmentações mais precisas. Arquiteturas treinadas com aumento alcançaram maiores valores de coeficiente *Dice* e acurácia, como a SharpUNet no *dataset* GlaS (89,47% e 86,34%) e a UNet++ no *dataset* OCDC (91,81% e 91,40%). Embora o aumento por distorção tenha melhorado os resultados em relação ao teste sem aumento, seu desempenho foi inferior às outras estratégias analisadas, possivelmente devido à introdução de variações que dificultam a preservação das características morfológicas essenciais para a segmentação. A análise estatística (teste de Friedman) confirmou o impacto significativo das estratégias de aumento de dados nos resultados de ambos os *datasets* ($p < 0,001$). As comparações *post-hoc* de Conover indicaram que todas as técnicas de aumento superaram a ausência de aumento, e que as abordagens de aumento geométrico e [Santos et al. 2023a] foram superiores ao aumento por distorção.

A Tabela 4 apresenta uma análise dos modelos em relação aos métodos de

segmentação propostos na literatura para as bases de imagens.

Tabela 4. Comparação entre modelos baseados em UNet e os métodos de segmentação da literatura utilizando o coeficiente *Dice*.

Modelo	OCDC	GlaS
[Santos et al. 2021]	92,00	
[Sirinukunwattana et al. 2017]		86,80
[Yuan and Cheng 2024]		90,56 ± 0,54
UNet	91,54 ± 5,76	89,05 ± 7,98
UNet++	91,81 ± 5,87	89,21 ± 8,13
SharpUNet	91,00 ± 5,75	89,47 ± 8,82
TransUNet	90,72 ± 5,92	88,42 ± 7,12

Entre as arquiteturas avaliadas, observa-se que a UNet++ obteve os melhores resultados no conjunto de dados OCDC (91,81%), enquanto a SharpUNet obteve a melhor desempenho no GlaS (89,47%), com resultados competitivos aos da literatura. Esses resultados indicam que arquiteturas com aprimoramentos estruturais, como conexões densas (UNet++) e mecanismos de realce de bordas (SharpUNet), contribuem para uma segmentação mais precisa em diferentes domínios histológicos. No entanto, é importante notar que o desempenho dos modelos varia de acordo com as características dos conjuntos de dados, incluindo a complexidade das estruturas segmentadas e a qualidade das anotações.

4. Conclusão

Este trabalho investigou modelos de aprendizagem profunda para a etapa de segmentação de estruturas histológicas em imagens da cavidade oral e colorretal. As abordagens foram investigadas em diferentes estruturas histológicas, tais como regiões tumorais e glândulas. Foram analisados os modelos baseados em segmentação semântica de arquiteturas de CNNs baseadas na arquitetura UNet (UNet, UNet++ e SharpUNet), assim como um modelo híbrido entre as abordagens CNN e Transformer, denominado TransUNet. O estudo destaca que a UNet++, SharpUNet e UNet obtiveram resultados consistentes, enquanto o TransUNet apresentou desempenho inferior, possivelmente devido à sua complexidade e necessidade de grandes volumes de dados para treinamento eficaz. O aumento de dados se mostrou essencial para melhorar a generalização dos modelos, reforçando a importância de técnicas que ampliem a diversidade dos conjuntos de treinamento, especialmente em domínios médicos com dados limitados. Trabalhos futuros podem explorar ajustes de hiperparâmetros, técnicas avançadas de pré-processamento, uso de *cross-validation* e a avaliação desses modelos em conjuntos de dados mais amplos e variados, visando aprimorar a precisão e a robustez da segmentação em cenários clínicos diversos.

5. Agradecimentos

Os autores agradecem o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Processos nº 313643/2021-0, nº 311404/2021-9 e nº 307318/2022-2), da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG (Processo nº APQ-00727-24) e da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (Processo nº 2022/03020-1).

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.
- Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *CoRR*, abs/2102.04306.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929.
- Maurício, J., Domingues, I., and Bernardino, J. (2023). Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Applied Sciences*, 13(9):5521.
- 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.
- 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. (2021). Automated detection of tumor regions from oral histological whole slide images using fully convolutional neural networks. *Biomedical Signal Processing and Control*, 69:102921.
- 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.
- Silva, A. B., Martins, A. S., Tosta, T. A. A., Neves, L. A., Servato, J. P. S., de Araújo, M. S., de Faria, P. R., and do Nascimento, M. Z. (2022). Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. *Expert Systems with Applications*, 193:116456.
- 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.

- Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., Wang, L. Y., Matuszewski, B. J., Bruni, E., Sanchez, U., et al. (2017). Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502.
- 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.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Xu, H., Xu, Q., Cong, F., Kang, J., Han, C., Liu, Z., Madabhushi, A., and Lu, C. (2023). Vision transformers for computational histopathology. *IEEE Reviews in Biomedical Engineering*.
- 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.
- Yuan, Y. and Cheng, Y. (2024). Medical image segmentation with unet-based multi-scale context fusion. *Scientific Reports*, 14(1):15687.
- 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.