
Segmentação Semântica de Núcleos em Leucoplasia Bucal por meio de Comitês de Modelos de Aprendizagem Profunda

Lucas Nardelli de Freitas Botelho Saar



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Lucas Nardelli de Freitas Botelho Saar

**Segmentação Semântica de Núcleos em
Leucoplasia Bucal por meio de Comitês de
Modelos de Aprendizagem Profunda**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência de Dados

Orientador: Marcelo Zanchetta do Nascimento

Coorientador: Adriano Barbosa Silva

Uberlândia

2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

S112 Saar, Lucas Nardelli de Freitas Botelho, 2000-
2026 Segmentação Semântica de Núcleos em Leucoplasia Bucal por
meio de Comitês de Modelos de Aprendizagem Profunda [recurso
eletrônico] / Lucas Nardelli de Freitas Botelho Saar. - 2026.

Orientador: Marcelo Zanchetta do Nascimento.

Coorientador: Adriano Barbosa Silva.

Dissertação (Mestrado) - Universidade Federal de Uberlândia,
Pós-graduação em Ciência da Computação.

Modo de acesso: Internet.

DOI <http://doi.org/10.14393/ufu.di.2026.622>

Inclui bibliografia.

1. Computação. I. Nascimento, Marcelo Zanchetta do, 1976-,
(Orient.). II. Silva, Adriano Barbosa, 1991-, (Coorient.). III.
Universidade Federal de Uberlândia. Pós-graduação em Ciência da
Computação. IV. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

Nelson Marcos Ferreira - CRB6/3074



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação, 26/2026, PPGCO				
Data:	11 de Agosto de 2026	Hora de início:	09:00	Hora de encerramento:	10:45
Matrícula do Discente:	12412CCP020				
Nome do Discente:	Lucas Nardelli de Freitas Botelho Saar				
Título do Trabalho:	Segmentação Semântica de Núcleos em Leucoplasia Bucal por meio de Comitês de Modelos de Aprendizagem Profunda				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Inteligência Artificial				
Projeto de Pesquisa de vinculação:	CNPq, Projeto 302833/2025-0, e FAPEMIG, Projetos APQ-00727-24 e APD-01428-25.				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Adriano Barbosa Silva - FACOM/UFU(Coorientador), Bruno Augusto Nassif Travençolo - FACOM/UFU, Rodrigo Pereira Ramos - UNIVASF e Marcelo Zanchetta do Nascimento - FACOM/UFU, orientador(a) do(a) candidato(a).

Os examinadores participaram desde as seguintes localidades: Adriano Barbosa Silva - Uberaba/MG, Rodrigo Pereira Ramos - Petrolina/PE. Os outros membros da banca participaram da cidade de Uberlândia. O aluno(a) Lucas Nardelli de Freitas Botelho Saar participou de Guarapari/ES.

Iniciando os trabalhos o(a) presidente da mesa, Prof. Dr. Marcelo Zanchetta do Nascimento, apresentou a Comissão Examinadora e o(a) candidato(a), agradeceu a presença do público, e concedeu ao(à) Discente a palavra para a exposição do seu trabalho.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato(a):

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos,

conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Marcelo Zanchetta do Nascimento, Professor(a) do Magistério Superior**, em 11/08/2026, às 13:44, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Bruno Augusto Nassif Travençolo, Professor(a) do Magistério Superior**, em 11/08/2026, às 17:54, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Adriano Barbosa Silva, Usuário Externo**, em 11/08/2026, às 18:39, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **RODRIGO PEREIRA RAMOS, Usuário Externo**, em 01/09/2026, às 10:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7552113** e o código CRC **8755F8A9**.

Dedico este trabalho à minha mãe, Érica, e ao meu pai, Valter, pelo amor, apoio e incentivo constantes ao longo da minha trajetória.

Agradecimentos

Agradeço, primeiramente, ao meu orientador, Prof. Dr. Marcelo Zanchetta do Nascimento, pela orientação dedicada, pela confiança em mim depositada e por todo o conhecimento compartilhado ao longo desta jornada. Seus ensinamentos e seu incentivo foram fundamentais para a realização deste trabalho. Estendo meus agradecimentos ao Adriano, pelo apoio, pelas contribuições e pela parceria ao longo deste percurso.

Agradeço também aos colegas do Laboratório Interdisciplinar de Processamento e Análise de Imagens (LIPAI), pela convivência, pelas trocas de ideias e pelo apoio constante durante o desenvolvimento desta pesquisa. O ambiente colaborativo do laboratório foi essencial em diversas etapas deste trabalho.

Sou imensamente grato à minha família e aos meus amigos, que me acompanharam e me incentivaram ao longo de toda esta caminhada. De modo especial, agradeço à minha mãe, Érica, e ao meu pai, Valter, pelo amor incondicional, pelo apoio em todos os momentos e por sempre acreditarem em mim.

Por fim, agradeço, de forma especial, o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processo 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).

Resumo

A avaliação histopatológica das desordens orais potencialmente malignas, com destaque para a leucoplasia bucal, depende fortemente da análise da morfologia nuclear. Esse processo, entretanto, é marcado por subjetividade e por expressiva divergência interobservador. Nesse contexto, a segmentação precisa dos núcleos celulares constitui uma etapa fundamental para sistemas computacionais de apoio à análise histopatológica. Apesar de sua relevância, a segmentação de núcleos especificamente em imagens de leucoplasia bucal permanece pouco explorada, assim como o uso de comitês construídos a partir de arquiteturas heterogêneas. Esta dissertação propõe e avalia comitês de modelos heterogêneos de aprendizagem profunda para a segmentação semântica de núcleos celulares em imagens histopatológicas de leucoplasia bucal. Foram investigadas dez arquiteturas representativas de três paradigmas: convolucional, híbrido e baseado em atenção. Os modelos foram avaliados individualmente e combinados por meio de cinco estratégias de *ensemble learning*, considerando diferentes configurações de aumento de dados. A avaliação empregou métricas tradicionais de segmentação e validação estatística. Como contribuição metodológica, foram propostas e formalizadas nesta dissertação três métricas específicas de contribuição individual: *kept rate*, *discarded rate* e *error mitigation*. Essas métricas permitem quantificar, respectivamente, a preservação dos acertos individuais, a perda desses acertos após a combinação e a correção, pelo comitê, dos erros cometidos pelos modelos isolados. Entre os modelos individuais, o melhor desempenho foi obtido pela ResNeSt-Unet++, com coeficiente Dice de 0,8853. As estratégias de combinação baseadas em média probabilística mostraram-se as mais eficazes, com o *soft voting* atingindo coeficiente Dice de 0,8924 e superando de forma estatisticamente significativa os modelos individuais. A aplicação das métricas propostas indicou que arquiteturas de paradigmas distintos apresentam padrões de erro complementares, favorecendo a preservação de acertos e a mitigação de erros individuais. Observou-se ainda que o ganho dos comitês tende a saturar a partir de três a quatro modelos. Esses resultados demonstram que comitês heterogêneos constituem uma estratégia eficaz para a segmentação de núcleos em leucoplasia bucal.

Palavras-chave: segmentação de núcleos celulares; leucoplasia bucal; aprendizagem profunda; comitês de modelos; imagens histopatológicas..

Abstract

The histopathological assessment of oral potentially malignant disorders, particularly oral leukoplakia, relies heavily on the analysis of nuclear morphology. However, this process is marked by subjectivity and substantial interobserver variability. In this context, accurate cell nuclei segmentation is a fundamental step for computer-aided histopathological analysis systems. Despite its relevance, nuclei segmentation specifically in oral leukoplakia images remains underexplored, as does the use of ensembles built from heterogeneous architectures. This dissertation proposes and evaluates ensembles of heterogeneous deep learning models for the semantic segmentation of cell nuclei in histopathological images of oral leukoplakia. Ten architectures representing three paradigms were investigated: convolutional, hybrid, and attention-based. The models were evaluated individually and combined using five *ensemble learning* strategies under different data augmentation configurations. The evaluation employed traditional segmentation metrics and statistical validation. As a methodological contribution, three individual-contribution metrics were proposed and formalized in this dissertation: *kept rate*, *discarded rate*, and *error mitigation*. These metrics quantify, respectively, the preservation of an individual model's correct predictions, the loss of those correct predictions after combination, and the correction of individual errors by the ensemble. Among the individual models, the best performance was achieved by ResNeSt-Unet++, with a Dice coefficient of 0.8853. Combination strategies based on probabilistic averaging proved to be the most effective, with *soft voting* achieving a Dice coefficient of 0.8924 and statistically significant gains over the individual models. The application of the proposed metrics indicated that architectures from distinct paradigms exhibit complementary error patterns, favoring the preservation of correct predictions and the mitigation of individual errors. In addition, the gain provided by the ensembles tended to saturate when three to four models were combined. These results demonstrate that heterogeneous ensembles constitute an effective strategy for nuclei segmentation in oral leukoplakia.

Keywords: cell nuclei segmentation; oral leukoplakia; deep learning; ensemble learning; histopathological images..

Lista de ilustrações

- Figura 1 – Progressão histológica do epitélio em direção à malignidade. Da esquerda para a direita: tecido normal, com arquitetura preservada; lesão pré-cancerosa (displasia), com alterações celulares e arquiteturais; carcinoma *in situ*, com células malignas ainda restritas ao epitélio; e carcinoma invasivo, no qual as células ultrapassam a membrana basal e invadem o tecido adjacente, com potencial de disseminação. Em cada estágio são apresentadas uma representação esquemática e a imagem histológica correspondente. Fonte: elaborado pelo autor com base em (KUMAR; ASTER; ABBAS, 2010; KADEMANI, 2019). 34
- Figura 2 – Exemplos clínicos de desordens orais potencialmente malignas. Da esquerda para a direita: leucoplasia, caracterizada por placa branca que não se destaca à raspagem; eritroplasia, área avermelhada e aveludada não atribuível a outra condição; e fibrose submucosa oral, condição crônica associada a risco de transformação maligna. Fonte: elaborado pelo autor com base em (WARNAKULASURIYA; JOHNSON; WAAL, 2007; WAAL, 2009). 35
- Figura 3 – Etapas de preparação e digitalização de tecidos histológicos. Na parte superior, a sequência do processamento: fixação, desidratação e diafanização, inclusão em parafina, microtomia, montagem em lâmina e coloração por hematoxilina e eosina (H&E) e, por fim, a análise ao microscópio ou a digitalização, que origina a imagem digital (WSI/RGB). Na parte inferior, as principais fontes de variabilidade do processo: variação de coloração, espessura do corte, equipamentos de digitalização e artefatos de preparo. Fonte: elaborado pelo autor com base em (JUNQUEIRA; CARNEIRO, 2013; FISCHER et al., 2008; TELLEZ et al., 2019). 37

Figura 4 – Fluxo típico de um sistema de diagnóstico auxiliado por computador em histopatologia, contemplando as etapas de pré-processamento, segmentação, extração de características, classificação e apoio à decisão. Fonte: elaborado pelo autor com base em (IRSHAD et al., 2014; TELLEZ et al., 2019; KUMAR et al., 2019).	38
Figura 5 – Comparação dos principais tipos de segmentação em histopatologia. Da esquerda para a direita: imagem original corada com H&E; segmentação semântica, na qual todos os núcleos recebem o mesmo rótulo de classe; segmentação por instâncias, em que cada núcleo é identificado como uma instância distinta; e segmentação panóptica, que combina a segmentação semântica do tecido/fundo com a identificação de instâncias dos núcleos. Fonte: elaborado pelo autor com base em (KIRILLOV et al., 2019; MINAEE et al., 2021).	40
Figura 6 – Paradigmas de aprendizado de máquina. À esquerda, o aprendizado supervisionado: o modelo é treinado com dados de entrada acompanhados de rótulos de referência e aprende a produzir a saída desejada, como em tarefas de classificação e de segmentação. À direita, o aprendizado não supervisionado: a partir de dados sem rótulos, o modelo busca descobrir estrutura, por exemplo por meio de agrupamento. Fonte: elaborado pelo autor com base em (MITCHELL, 1997; BISHOP, 2006).	41
Figura 7 – Estrutura básica de uma rede neural artificial. Em (1), o neurônio artificial: as entradas x_i são ponderadas por pesos w_i e combinadas, juntamente com um viés b , por uma soma ponderada, cujo resultado passa por uma função de ativação $f(\cdot)$ que produz a saída y . Em (2), uma rede multicamadas do tipo <i>feedforward</i> , composta por camadas de entrada, ocultas e de saída, com indicação dos fluxos de propagação direta e de retropropagação do erro. Fonte: elaborado pelo autor com base em (HAYKIN et al., 2009; RUMELHART; HINTON; WILLIAMS, 1986).	42
Figura 8 – Etapas típicas de uma rede neural convolucional aplicada a uma imagem histológica: a partir da imagem de entrada, filtros convolucionais geram mapas de características, sucedidos por funções de ativação não lineares (ReLU) e operações de <i>pooling</i> ; a repetição desses blocos produz representações progressivamente mais abstratas, combinadas na camada de saída para gerar a predição final. Fonte: elaborado pelo autor com base em (ALBAWI; MOHAMMED; AL-ZAWI, 2017; LECUN; BENGIO; HINTON, 2015).	43

Figura 9 – Processamento de uma imagem por um <i>Transformer</i> : a imagem é dividida em <i>patches</i> convertidos em uma sequência de <i>tokens</i> ; o mecanismo de autoatenção (<i>self-attention</i>), baseado em consultas, chaves e valores (Q, K e V), relaciona cada <i>token</i> a todos os demais, modelando dependências globais; os blocos subsequentes produzem representações contextualizadas. Fonte: elaborado pelo autor com base em (VASWANI et al., 2017; DOSOVITSKIY et al., 2021).	45
Figura 10 – Duas estratégias de hibridização entre redes convolucionais e <i>Transformers</i> para segmentação. À esquerda (CNN + <i>Transformer</i>), uma rede convolucional extrai características locais e um módulo <i>Transformer</i> modela o contexto global antes da geração da máscara. À direita (<i>encoder Transformer</i> + decodificador convolucional), um codificador baseado em <i>Transformer</i> fornece representações contextuais que alimentam um decodificador convolucional responsável pela recuperação espacial. Fonte: elaborado pelo autor com base em (CHEN et al., 2021; FAN et al., 2020).	46
Figura 11 – Representação esquemática de um comitê de modelos de aprendizagem profunda para segmentação de núcleos. A imagem histológica é processada por múltiplos modelos de arquiteturas distintas (convolucionais, híbridas e baseadas em atenção), cujas previsões individuais são combinadas por diferentes regras de fusão, gerando a máscara final. Fonte: elaborado pelo autor com base em (DIETTERICH, 2000; KITTLER et al., 1998).	47
Figura 12 – Fluxograma representando o pipeline da aproximação proposta para segmentação usando modelos <i>ensembles</i> com arquiteturas de aprendizagem profunda.	58
Figura 13 – Representação dos <i>datasets</i> : Leuco (A) e CryoNuSeg (B)	59
Figura 14 – Representação da arquitetura U-Net, composta por um caminho de contração (<i>encoder</i>) e um caminho de expansão (<i>decoder</i>), interligados por conexões de atalho que permitem combinar informações semânticas profundas com detalhes espaciais de maior resolução. Fonte: elaborado pelo autor com base em (RONNEBERGER; FISCHER; BROX, 2015).	62
Figura 15 – Representação da arquitetura UNet++, caracterizada por conexões de atalho densamente aninhadas entre o <i>encoder</i> e o <i>decoder</i> , permitindo uma fusão gradual de características em diferentes níveis de resolução e reduzindo a lacuna semântica entre os mapas extraídos e reconstruídos. Fonte: elaborado pelo autor com base em (ZHOU et al., 2018).	63

Figura 16 – Representação da arquitetura <i>Feature Pyramid Network</i> (FPN), baseada na combinação de um caminho <i>bottom-up</i> , um caminho <i>top-down</i> e conexões laterais para construção de mapas de características multi-escala, úteis para segmentar estruturas de diferentes tamanhos. Fonte: elaborado pelo autor com base em (LIN et al., 2017).	64
Figura 17 – Representação da arquitetura MANet, que combina uma estrutura do tipo <i>encoder-decoder</i> com mecanismos de atenção voltados à agregação multiescala de características. Neste trabalho, a arquitetura foi empregada com <i>encoder</i> MiT-B2, baseado em <i>Transformer</i> . Fonte: elaborado pelo autor com base em (FAN et al., 2020; XIE et al., 2021).	65
Figura 18 – Representação da arquitetura SegFormer, composta por um <i>encoder</i> hierárquico baseado em <i>Mix Transformer</i> (MiT) e um decodificador leve para agregação de características em múltiplas escalas e geração da máscara de segmentação. Fonte: elaborado pelo autor com base em (XIE et al., 2021).	66
Figura 19 – Representação da arquitetura TransUNet, que combina um codificador convolucional, responsável pela extração de características locais, com blocos <i>Transformer</i> , responsáveis pela modelagem de dependências de longo alcance, seguidos por um decodificador responsável pela reconstrução espacial da máscara. Fonte: elaborado pelo autor com base em (CHEN et al., 2021).	67
Figura 20 – Representação da arquitetura MaxViT com decodificador Deformable LKA, na qual o <i>backbone</i> MaxViT atua como codificador para extração de características locais e globais, enquanto o decodificador com atenção <i>Deformable Large Kernel Attention</i> refina as representações durante a reconstrução da máscara segmentada. Fonte: elaborado pelo autor com base em (TU et al., 2022; AZAD et al., 2024).	68
Figura 21 – Comparação qualitativa das segmentações no <i>dataset</i> de leucoplasia (cenário geométrico), para um exemplo representativo: imagem original, <i>ground truth</i> , resultado do <i>ensemble</i> por <i>soft voting</i> e previsões dos dez modelos individuais. As cores indicam verdadeiros positivos (branco), falsos positivos (vermelho), falsos negativos (azul) e verdadeiros negativos (fundo, em preto). Os rótulos em verde identificam os modelos que compõem o <i>ensemble</i> selecionado.	87
Figura 22 – Resultado da variação do limiar	90

Lista de tabelas

Tabela 1	– Síntese das arquiteturas de segmentação avaliadas, incluindo tipo de modelo, <i>encoder</i> utilizado e estratégia de modelagem.	60
Tabela 2	– Desempenho dos modelos no <i>dataset</i> de Leucoplasia, considerando diferentes estratégias de aumento de dados.	80
Tabela 3	– Desempenho dos modelos no <i>dataset</i> CryoNuSeg, considerando diferentes estratégias de aumento de dados.	82
Tabela 4	– Comparação entre técnicas de <i>ensemble</i> no <i>dataset</i> de leucoplasia considerando diferentes estratégias de aumento de dados.	84
Tabela 5	– Impacto do número de modelos no desempenho do <i>ensemble</i> utilizando <i>soft voting</i>	86
Tabela 6	– Análise de variação do limiar no <i>dataset</i> de leucoplasia (sem aumento de dados).	88
Tabela 7	– Análise de variação do limiar com aumento geométrico.	88
Tabela 8	– Análise de variação do limiar com estratégia RCAug.	89
Tabela 9	– Análise de contribuição dos modelos no <i>ensemble</i> utilizando <i>soft voting</i> com aumento geométrico.	91
Tabela 10	– Teste de normalidade de Shapiro-Wilk para os modelos avaliados. . . .	92
Tabela 11	– Teste de normalidade de Shapiro-Wilk para estratégias de aumento de dados.	92
Tabela 12	– Teste t pareado entre o <i>ensemble</i> e os modelos individuais.	92
Tabela 13	– Teste t pareado entre estratégias de aumento de dados.	93
Tabela 14	– Comparação da abordagem proposta com trabalhos recentes da literatura, considerando o coeficiente Dice (D_C) nos <i>datasets</i> de leucoplasia bucal e CryoNuSeg.	94

Lista de siglas

CAD *Computer-Aided Diagnosis*

CNN *Convolutional Neural Network*

DEO Displasia Epitelial Oral

DLKA *Deformable Large Kernel Attention*

FN *False Negative*

FP *False Positive*

FPN *Feature Pyramid Network*

GAN *Generative Adversarial Network*

H&E Hematoxilina e Eosina

INCA Instituto Nacional de Câncer

IoU *Intersection over Union*

LB Leucoplasia Bucal

LVP Leucoplasia Verrucosa Proliferativa

MiT *Mix Transformer*

NAS *Neural Architecture Search*

OPMD Desordens Oraís Potencialmente Malignas

ROI *Region of Interest*

RGB *Red, Green and Blue*

RNA Rede Neural Artificial

scSE *Spatial and Channel Squeeze-and-Excitation*

TN *True Negative*

TP *True Positive*

ViT *Vision Transformer*

WSI *Whole Slide Imaging*

Sumário

1	INTRODUÇÃO	25
1.1	Motivação	28
1.2	Objetivos e Desafios da Pesquisa	29
1.3	Hipóteses	30
1.4	Contribuições	30
1.5	Organização desta Dissertação	31
2	FUNDAMENTAÇÃO TEÓRICA	33
2.1	Câncer	33
2.2	Desordens Orais Potencialmente Malignas e Leucoplasia	34
2.3	Preparação e Digitalização de Tecidos Histológicos	36
2.4	Sistemas Computacionais de Apoio ao Diagnóstico	37
2.5	Segmentação de Imagens	39
2.6	Aprendizado de Máquina	40
2.7	Redes Neurais Artificiais	41
2.8	Redes Neurais Convolucionais	43
2.9	<i>Transformers</i> e Mecanismos de Atenção	44
2.10	Arquiteturas Híbridas	45
2.11	Comitês de Aprendizagem Profunda	46
3	ESTADO DA ARTE	49
3.1	Aprendizagem Profunda para Segmentação em Imagens Histológicas Orais	49
3.2	<i>Ensembles</i> de Aprendizagem Profunda para Segmentação de Imagens Histológicas	52
3.3	Considerações	54
4	METODOLOGIA	57

4.1	Bases de Imagens	58
4.2	Arquiteturas de redes para segmentação	59
4.2.1	Arquiteturas baseadas em U-Net	60
4.2.2	Arquiteturas UNet++	61
4.2.3	<i>Feature Pyramid Network</i>	63
4.2.4	MAnet	64
4.2.5	SegFormer	65
4.2.6	TransUNet híbrida	66
4.2.7	MaxViT com decodificador Deformable LKA	67
4.3	Aumento de Dados	68
4.4	Estratégias de <i>Ensemble Learning</i>	70
4.4.1	<i>Soft Voting</i>	71
4.4.2	<i>Hard Voting</i>	71
4.4.3	<i>Weighted Soft Voting</i>	72
4.4.4	Union Max	72
4.4.5	<i>Intersection Min</i>	73
4.5	Métricas de Avaliação	73
4.5.1	Coeficiente <i>Dice</i>	74
4.5.2	<i>Intersection over Union</i>	74
4.5.3	Acurácia	75
4.5.4	Sensibilidade	75
4.5.5	Especificidade	75
4.5.6	Métricas de contribuição individual no <i>ensemble</i>	75
4.5.7	<i>Kept Rate</i>	76
4.5.8	<i>Discarded Rate</i>	76
4.5.9	<i>Error Mitigation</i>	77
5	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	79
5.1	Resultados dos modelos individuais	79
5.1.1	Desempenho no <i>dataset</i> de leucoplasia	80
5.1.2	Desempenho no <i>dataset</i> CryoNuSeg	81
5.2	Resultados com <i>Ensemble Learning</i>	84
5.2.1	Comparação entre Técnicas de <i>Ensemble</i>	84
5.2.2	Impacto do Número de Modelos no <i>Ensemble</i>	85
5.2.3	Análise Qualitativa das Segmentações	86
5.2.4	Análise da Influência do Limiar de Binarização	87
5.2.5	Análise de Contribuição dos Modelos no <i>Ensemble</i>	90
5.3	Análise Estatística dos Resultados	91
5.4	Comparação com o Estado da Arte	93

6	CONCLUSÕES	97
6.1	Trabalhos Futuros	98
6.2	Contribuições desta Pesquisa	99
	REFERÊNCIAS	101

Introdução

O câncer figura entre os principais problemas de saúde pública em escala mundial. Do ponto de vista molecular, compreende um conjunto heterogêneo de doenças associadas ao acúmulo de alterações genéticas que conferem às células vantagens seletivas de crescimento e sobrevivência, permitindo que proliferem de maneira descontrolada e invadam tecidos adjacentes (KUMAR; ABBAS; ASTER, 2013; KADEMANI, 2019). A magnitude do problema é expressiva: estimou-se, para o ano de 2022, cerca de 20 milhões de casos novos e 9,7 milhões de óbitos por câncer no mundo, com projeções de crescimento contínuo nas próximas décadas (BRAY et al., 2024). Como consequência de sua natureza proliferativa, células neoplásicas costumam exibir alterações morfológicas marcantes, incluindo variações de tamanho, forma e padrão de coloração em relação às células normais, características que constituem importantes indicadores diagnósticos na análise dos tecidos (KUMAR; ASTER; ABBAS, 2010; KADEMANI, 2019).

Entre as diversas neoplasias malignas, o câncer da cavidade oral destaca-se por sua relevância epidemiológica. Em 2022, foram estimados aproximadamente 390 mil casos novos de câncer de lábio e cavidade oral no mundo, sendo a doença particularmente frequente em populações do Sul da Ásia e predominantemente representada pelo carcinoma de células escamosas (BRAY et al., 2024). No Brasil, o Instituto Nacional de Câncer (INCA) estima 15.100 casos novos de câncer da cavidade oral para cada ano do triênio 2023–2025, dos quais cerca de 10.900 em homens e 4.200 em mulheres, o que coloca a doença entre os dez tipos de câncer mais incidentes no país (SANTOS, 2023). Apesar de ser, em grande medida, prevenível e acessível ao exame clínico, o câncer bucal é frequentemente diagnosticado em estágios avançados, fato que compromete o prognóstico e mantém a taxa de sobrevida em cinco anos em torno de 50% (SANTOS, 2023; KADEMANI, 2019). Esse cenário evidencia a importância de estratégias que favoreçam a detecção precoce, especialmente das lesões que antecedem a malignidade.

Nesse contexto, ganham relevância as Desordens Orais Potencialmente Malignas (OPMD), alterações da mucosa associadas a risco aumentado de transformação maligna (WARNAKULASURIYA; JOHNSON; WAAL, 2007; WAAL, 2009). A leucoplasia bucal é a OPMD

mais prevalente, com prevalência mundial estimada entre 1,36% e 2,60% na população geral, sendo definida como uma placa predominantemente branca que não pode ser destacada por raspagem e que não corresponde a nenhuma outra doença conhecida (WAAL, 2009; WARNAKULASURIYA et al., 2008). Embora a maioria das lesões permaneça estável, uma parcela evolui para carcinoma, com taxas de transformação maligna que, em metanálises recentes, situam-se em torno de 7% a 10% (WARNAKULASURIYA et al., 2008; AGUIRRE-URIZAR; MENDOZA; WARNAKULASURIYA, 2021). Variantes como a leucoplasia verrucosa proliferativa apresentam comportamento ainda mais agressivo, com elevadas taxas de recidiva e malignização (BAGAN et al., 2010). A correta caracterização dessas lesões é, portanto, determinante para a definição de condutas terapêuticas e para o acompanhamento dos pacientes.

A avaliação dessas lesões apoia-se fundamentalmente no exame histopatológico, considerado o padrão de referência para a identificação das alterações celulares associadas à malignidade (KADEMANI, 2019). Após o preparo das lâminas e a coloração por hematoxilina e eosina (H&E), na qual a hematoxilina evidencia predominantemente os núcleos em tonalidade azul-arroxeadas e a eosina realça o citoplasma e a matriz extracelular em tons rosados (FISCHER et al., 2008), o patologista avalia a presença e a intensidade de displasia epitelial, conjunto de alterações arquiteturais e citológicas do epitélio que inclui variações no tamanho e na forma dos núcleos, hipercromasia e aumento da relação núcleo-citoplasma (KADEMANI, 2019; KUMAR; ASTER; ABBAS, 2010). Essa avaliação, contudo, depende fortemente da análise da morfologia nuclear e é marcada por considerável subjetividade. A graduação da displasia apresenta taxas expressivas de divergência interobservador, sendo um processo demorado e sujeito à fadiga e à experiência do avaliador (MÜLLER, 2018; SMITH et al., 2009). Tal variabilidade compromete a reprodutibilidade do diagnóstico e motiva a busca por ferramentas que tornem a análise mais objetiva.

Os sistemas de diagnóstico auxiliado por computador (do inglês, *Computer-Aided Diagnosis* (CAD)) surgem como resposta a essa demanda, combinando conhecimento médico e técnicas computacionais para atuar como uma segunda opinião ao especialista (SHAFI; PARWANI, 2023). Em histopatologia, o uso de CAD tem como objetivo aumentar a objetividade e a reprodutibilidade das análises, mitigando a divergência interobservador e os efeitos de ruídos e variações de coloração (SONG et al., 2023; SRINIDHI; CIGA; MARTEL, 2021). A análise computacional de imagens histológicas costuma ser organizada em etapas encadeadas, tais como pré-processamento, segmentação, extração de características e classificação (KADASKAR; PATIL, 2023). Nesse fluxo, a segmentação dos núcleos celulares ocupa posição crítica, pois a delimitação precisa dessas estruturas é etapa prévia indispensável para a extração quantitativa de atributos morfológicos. Como a qualidade das estruturas delimitadas condiciona diretamente o desempenho das etapas subsequentes, falhas na segmentação podem propagar-se por todo o sistema e compro-

meter a confiabilidade do apoio à decisão (KUMAR et al., 2019; KADASKAR; PATIL, 2023).

A segmentação de núcleos em imagens histopatológicas é, entretanto, uma tarefa desafiadora. A irregularidade morfológica dos tecidos, a frequente sobreposição e aglomeração de núcleos e a elevada variabilidade de coloração, decorrente de diferenças nos protocolos de preparo, na espessura dos cortes e nos equipamentos de digitalização, dificultam tanto a análise manual quanto a automática (GONZALEZ; WOODS, 2018; KUMAR et al., 2019; TELLEZ et al., 2019). Métodos clássicos baseados em limiarização, detecção de bordas ou regiões dependem fortemente de ajustes manuais e apresentam desempenho inconsistente diante dessa complexidade (OTSU, 1979; GONZALEZ; WOODS, 2018). Com o avanço do aprendizado profundo, modelos baseados em redes neurais, em especial as arquiteturas do tipo codificador-decodificador como a U-Net, tornaram-se uma abordagem predominante, por aprenderem automaticamente representações discriminativas a partir dos próprios dados (RONNEBERGER; FISCHER; BROX, 2015; MINAEE et al., 2021).

Apesar dos avanços proporcionados pelo aprendizado profundo, a segmentação de núcleos em histopatologia ainda apresenta lacunas importantes quando considerada no contexto específico das lesões orais potencialmente malignas. Muitos estudos concentram-se em bases públicas compostas por tecidos de múltiplos órgãos ou por domínios histológicos distintos, como mama, cólon, próstata e outros tecidos tumorais, que, embora relevantes para o desenvolvimento de métodos gerais de segmentação nuclear, não representam integralmente as particularidades morfológicas e cromáticas das lesões de leucoplasia bucal (KUMAR et al., 2019; GRAHAM et al., 2019; MAHBOD et al., 2021). No domínio oral, trabalhos recentes têm investigado a segmentação de estruturas em displasia epitelial oral e em outras alterações potencialmente malignas (SILVA et al., 2022; SILVA et al., 2024; SHEPHARD et al., 2023), mas a segmentação semântica de núcleos especificamente em imagens histológicas de leucoplasia bucal permanece pouco explorada, em parte devido à escassez de bases anotadas e validadas por especialistas para essa condição.

Além da limitação relacionada ao domínio de aplicação, há também uma questão metodológica associada à escolha das arquiteturas de segmentação. Modelos convolucionais, como a U-Net e suas variantes, apresentam forte viés indutivo local e são particularmente eficazes na representação de texturas, bordas e padrões espaciais próximos, características relevantes para a delimitação de estruturas celulares (RONNEBERGER; FISCHER; BROX, 2015; MINAEE et al., 2021). Por outro lado, arquiteturas baseadas em mecanismos de atenção e *Transformers* foram propostas para capturar dependências de maior alcance e incorporar informações contextuais mais amplas, podendo contribuir para a análise de padrões morfológicos distribuídos ao longo da imagem (VASWANI et al., 2017; DOSOVITSKIY et al., 2021; CHEN et al., 2021; XIE et al., 2021). Dessa forma, diferentes paradigmas arquiteturais podem explorar aspectos complementares das imagens histológicas, o que sugere que seus erros e acertos não necessariamente ocorrem nas mesmas

regiões. Essa complementaridade constitui uma motivação para investigar estratégias de combinação de modelos, nas quais a diversidade entre arquiteturas pode ser explorada para produzir segmentações mais robustas do que aquelas obtidas por um único modelo isolado (DIETTERICH, 2000; KUNCHEVA, 2003; SAGI; ROKACH, 2018).

Essa observação aponta para os comitês (do inglês, *ensembles*) como uma estratégia promissora para aumentar a robustez e a confiabilidade das segmentações. A combinação das predições de múltiplos modelos fundamenta-se na premissa de que os erros cometidos por um modelo podem ser compensados pelos acertos dos demais, desde que estes apresentem comportamentos suficientemente distintos (DIETTERICH, 2000; SAGI; ROKACH, 2018). A eficácia de um comitê depende, portanto, criticamente da diversidade entre seus membros (KUNCHEVA, 2003; GANAIE et al., 2022). Contudo, grande parte dos comitês aplicados à segmentação biomédica é construída a partir de modelos homogêneos, isto é, variações da mesma arquitetura treinadas com diferentes inicializações, subconjuntos de dados ou sementes, o que pode limitar a diversidade de vieses indutivos efetivamente explorada. A construção de comitês heterogêneos, que reúnem arquiteturas convolucionais, híbridas e baseadas em atenção, com mecanismos de modelagem distintos, permanece uma direção ainda pouco investigada na segmentação de núcleos e, em particular, no contexto das desordens orais potencialmente malignas. É nessa lacuna que se insere o presente trabalho, cujas motivações são detalhadas a seguir.

1.1 Motivação

A principal motivação deste trabalho reside na natureza subjetiva e pouco reproduzível da avaliação histopatológica das desordens orais potencialmente malignas. Como a graduação da displasia depende fortemente da avaliação visual da morfologia nuclear, está sujeita a divergências interobservador expressivas e à influência da fadiga e da experiência do avaliador (MÜLLER, 2018; SMITH et al., 2009). Em um cenário no qual o diagnóstico tardio compromete o prognóstico, ferramentas capazes de tornar essa análise mais objetiva e reproduzível têm potencial impacto clínico. A segmentação precisa dos núcleos é o alicerce dessas ferramentas, pois delimita as estruturas sobre as quais atributos morfométricos podem ser posteriormente calculados. Assim, qualquer aprimoramento nessa etapa tende a favorecer as análises subsequentes em sistemas computacionais de apoio ao diagnóstico (KUMAR et al., 2019; IRSHAD et al., 2014; SRINIDHI; CIGA; MARTEL, 2021). Aprimorar a segmentação de núcleos em leucoplasia é, portanto, uma etapa estratégica para a construção de sistemas confiáveis voltados à análise dessas lesões.

A estratégia para o uso de comitês decorre da constatação de que nenhuma arquitetura isolada necessariamente domina o problema em todos os cenários. Modelos convolucionais, híbridos e baseados em atenção possuem vieses indutivos distintos e, por isso, podem produzir erros em regiões diferentes da imagem (DIETTERICH, 2000; KUNCHEVA, 2003).

Essa complementaridade representa uma oportunidade: em vez de buscar uma única arquitetura ótima, é possível combinar modelos já existentes e pré-treinados para obter segmentações mais robustas, com menor custo de desenvolvimento em comparação ao treinamento de arquiteturas complexas a partir do zero (SAGI; ROKACH, 2018; KAMNITSAS et al., 2018). Tal estratégia é especialmente atraente em domínios como o da histopatologia de leucoplasia, marcados pela escassez de dados anotados e pelo alto custo da anotação por especialistas, condições em que aproveitar representações pré-treinadas e a diversidade entre modelos tende a ser uma alternativa relevante. Ainda assim, comitês construídos deliberadamente a partir de arquiteturas heterogêneas permanecem pouco explorados na segmentação de núcleos, o que reforça o interesse desta investigação.

Por fim, este trabalho se mostra relevante também pela necessidade de compreender não apenas *se* a combinação de modelos melhora os resultados, mas *como* esse ganho emerge. Grande parte dos estudos sobre comitês limita-se a reportar o desempenho final, tratando a combinação como uma caixa-preta. Investigar em que medida o comitê preserva os acertos individuais, descarta decisões corretas ou corrige erros oferece uma visão mais transparente de seu funcionamento, relevante tanto para a confiança em sistemas CAD quanto para o projeto de comitês mais eficazes.

1.2 Objetivos e Desafios da Pesquisa

O objetivo geral desta dissertação é desenvolver e avaliar uma abordagem baseada em comitês de modelos heterogêneos de aprendizado profundo para a segmentação semântica de núcleos celulares em imagens histopatológicas de leucoplasia bucal, investigando em que medida a combinação de arquiteturas com vieses indutivos distintos contribui para o ganho de desempenho e de robustez em relação aos modelos individuais.

A partir desse objetivo geral, definem-se os seguintes objetivos específicos:

- ❑ investigar, treinar e avaliar individualmente modelos de segmentação representativos de três paradigmas distintos: redes convolucionais, modelos híbridos que combinam convolução e *Transformers*, e modelos fortemente orientados por mecanismos de atenção;
- ❑ avaliar o impacto de diferentes estratégias de aumento de dados sobre o desempenho e a capacidade de generalização dos modelos;
- ❑ investigar e comparar diferentes estratégias de fusão para a combinação das predições dos modelos em comitê, incluindo abordagens baseadas em probabilidade e em operações sobre máscaras binárias;

- ❑ analisar a contribuição individual de cada modelo dentro do comitê, de modo a compreender como a combinação preserva acertos, descarta decisões corretas ou corrige erros previamente cometidos;
- ❑ validar estatisticamente os resultados obtidos, verificando a significância das diferenças entre os modelos individuais, os comitês e as estratégias de aumento de dados.

1.3 Hipóteses

A investigação conduzida neste trabalho orienta-se pelas seguintes hipóteses:

- ❑ **H1** – a combinação de múltiplas arquiteturas por meio de comitês tende a produzir segmentações de qualidade superior à obtida pelos modelos individuais;
- ❑ **H2** – a diversidade de paradigmas arquiteturais (convolucional, híbrido e baseado em atenção) gera previsões complementares entre os modelos, permitindo que o comitê corrija parte dos erros individuais sem comprometer significativamente os acertos;
- ❑ **H3** – estratégias de aumento de dados contribuem para a melhoria do desempenho dos modelos, sobretudo no domínio de treinamento, favorecendo a redução do sobreajuste;
- ❑ **H4** – estratégias de fusão que preservam a informação probabilística dos modelos superam aquelas baseadas exclusivamente em decisões binárias;
- ❑ **H5** – o ganho proporcionado pelo comitê não depende apenas da quantidade absoluta de modelos combinados, mas também da complementaridade entre suas previsões, tendendo a saturar a partir de um pequeno número de arquiteturas.

1.4 Contribuições

As principais contribuições desta dissertação podem ser descritas em:

- ❑ um estudo comparativo abrangente de dez arquiteturas de segmentação, organizadas em três paradigmas (convolucional, híbrido e baseado em atenção), aplicado à segmentação de núcleos celulares em lesões de leucoplasia bucal, contexto ainda pouco explorado na literatura;
- ❑ a investigação sistemática de comitês heterogêneos de aprendizado profundo, avaliando diferentes estratégias de fusão das previsões e o efeito do número de modelos combinados;

- ❑ a proposta, formalização e aplicação de três métricas de contribuição em nível de *pixel*, denominadas taxa de retenção (*kept rate*), taxa de descarte (*discarded rate*) e mitigação de erros (*error mitigation*). Essas métricas permitem quantificar, respectivamente, quanto dos acertos individuais é preservado, quanto é perdido e quanto dos erros individuais é corrigido pela decisão conjunta, tornando mais transparente o funcionamento interno do comitê;
- ❑ a avaliação do impacto de diferentes estratégias de aumento de dados sobre o desempenho e a generalização dos modelos;
- ❑ a análise da capacidade de generalização das abordagens propostas por meio de um conjunto de dados público e morfológicamente heterogêneo, distinto do domínio de treinamento;
- ❑ a validação estatística dos resultados, conferindo maior robustez às conclusões obtidas sobre o ganho dos comitês e das estratégias de aumento de dados.

1.5 Organização desta Dissertação

Além deste capítulo introdutório, o restante desta dissertação está organizado da seguinte forma:

- ❑ o Capítulo 2 apresenta a fundamentação teórica, abordando o contexto clínico do câncer oral, das desordens orais potencialmente malignas e da leucoplasia bucal, bem como os fundamentos computacionais relacionados a sistemas CAD, segmentação de imagens, redes neurais convolucionais, mecanismos de atenção, *Transformers*, arquiteturas híbridas e comitês de modelos;
- ❑ o Capítulo 3 apresenta o estado da arte, discutindo trabalhos relacionados à segmentação em imagens histológicas orais e ao uso de *ensemble learning* em tarefas de segmentação histopatológica;
- ❑ o Capítulo 4 descreve a metodologia proposta, detalhando as bases de imagens utilizadas, as arquiteturas de segmentação investigadas, as estratégias de aumento de dados, as técnicas de *ensemble learning* e as métricas de avaliação adotadas;
- ❑ o Capítulo 5 apresenta e discute os resultados experimentais, contemplando o desempenho dos modelos individuais, os resultados das estratégias de comitê, a análise de contribuição dos modelos, a influência do limiar de binarização e a validação estatística dos achados;
- ❑ o Capítulo 6 apresenta as conclusões da dissertação, sintetizando as principais descobertas, discutindo suas limitações e apontando direções para pesquisas futuras.

Fundamentação Teórica

2.1 Câncer

Lesões tumorais caracterizam-se pelo crescimento descontrolado de células que originam massas teciduais conhecidas como tumores. De acordo com seu comportamento biológico, essas lesões são classificadas como benignas ou malignas (KUMAR; ASTER; ABBAS, 2010). Tumores benignos apresentam crescimento limitado, permanecem restritos ao local de origem e raramente colocam a vida do paciente em risco. Já os tumores malignos, denominados câncer, exibem comportamento agressivo, com capacidade de invadir estruturas adjacentes e disseminar-se para regiões distantes por meio de metástases (KUMAR; ABBAS; ASTER, 2013).

Do ponto de vista molecular, o câncer é compreendido como uma doença genética originada pelo acúmulo de mutações no DNA, que podem surgir espontaneamente ou em decorrência da exposição a fatores ambientais, como o tabagismo, o consumo de álcool, infecções virais e a radiação (KADEMANI, 2019). Tais mutações conferem às células alteradas vantagens seletivas de crescimento e sobrevivência, permitindo que proliferem de maneira acelerada e substituam progressivamente as células saudáveis (KUMAR; ABBAS; ASTER, 2013). Como consequência, células neoplásicas costumam exibir alterações morfológicas marcantes, incluindo variações de tamanho, forma e padrão de coloração em relação às células normais, características que constituem importantes indicadores diagnósticos na análise histológica (KUMAR; ASTER; ABBAS, 2010; KADEMANI, 2019). Essas alterações acentuam-se à medida que a lesão progride. A Figura 1 ilustra essa evolução, do tecido normal à displasia e, posteriormente, ao carcinoma *in situ* e ao carcinoma invasivo.

Entre as neoplasias malignas, o câncer da cavidade oral destaca-se por sua relevância epidemiológica, figurando entre os tipos mais frequentes no mundo e sendo frequentemente diagnosticado em estágios avançados, o que compromete o prognóstico (KADEMANI, 2019; SANTOS, 2023). A taxa de sobrevida em cinco anos permanece em torno de 50%, evidenciando a importância de estratégias que favoreçam a detecção precoce (SANTOS,

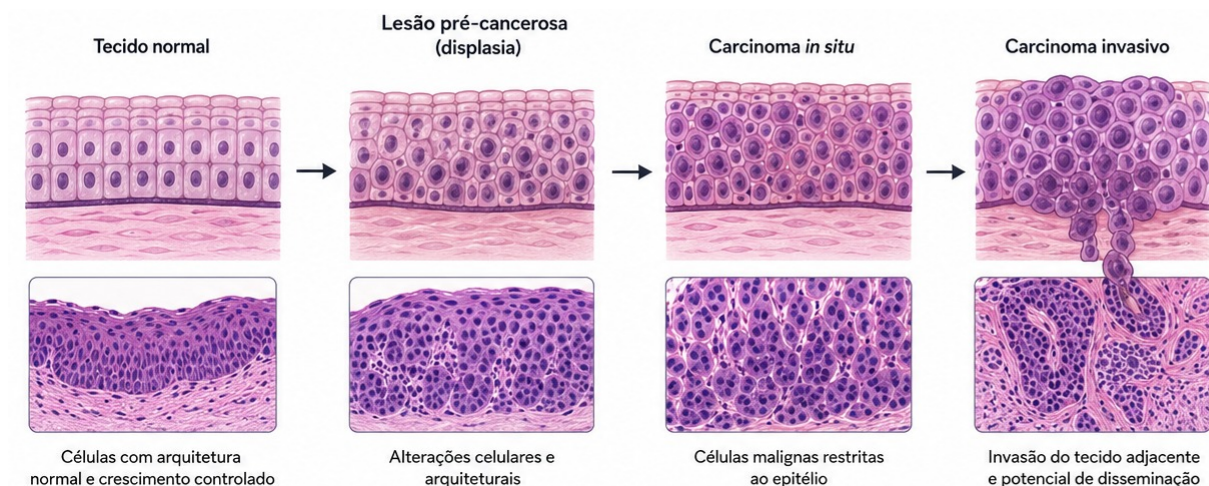


Figura 1 – Progressão histológica do epitélio em direção à malignidade. Da esquerda para a direita: tecido normal, com arquitetura preservada; lesão pré-cancerosa (displasia), com alterações celulares e arquiteturais; carcinoma *in situ*, com células malignas ainda restritas ao epitélio; e carcinoma invasivo, no qual as células ultrapassam a membrana basal e invadem o tecido adjacente, com potencial de disseminação. Em cada estágio são apresentadas uma representação esquemática e a imagem histológica correspondente. Fonte: elaborado pelo autor com base em (KUMAR; ASTER; ABBAS, 2010; KADEMANI, 2019).

2023). Esse cenário motiva, em particular, o estudo das lesões pré-cancerosas, alterações teciduais que, embora frequentemente estáveis, podem evoluir para malignidade sob a influência de estímulos externos, como traumas, infecções e exposição prolongada a agentes carcinogênicos (WARNAKULASURIYA et al., 2008; KUMAR; ASTER; ABBAS, 2010). A caracterização precisa dessas lesões é, portanto, determinante para a definição de condutas terapêuticas, e merece atenção especial no caso da cavidade oral, conforme detalhado a seguir.

2.2 Desordens Orais Potencialmente Malignas e Leucoplasia

As lesões da mucosa oral com risco aumentado de transformação maligna são atualmente agrupadas sob a denominação de desordens orais potencialmente malignas (do inglês, *oral potentially malignant disorders* – OPMD) (WARNAKULASURIYA; JOHNSON; WAAL, 2007). Essa nomenclatura reconhece que o risco de progressão não está restrito a uma lesão isolada, mas associa-se a um estado de predisposição da mucosa (WAAL, 2009). Entre as condições incluídas nesse grupo destacam-se a leucoplasia, a eritroplasia e a fibrose submucosa oral (WARNAKULASURIYA; JOHNSON; WAAL, 2007; WAAL, 2009), ilustradas na Figura 2.

A Leucoplasia Bucal (LB) é definida como uma placa predominantemente branca da mucosa oral que não pode ser destacada por raspagem e que não corresponde a nenhuma



Figura 2 – Exemplos clínicos de desordens orais potencialmente malignas. Da esquerda para a direita: leucoplasia, caracterizada por placa branca que não se destaca à raspagem; eritroplasia, área avermelhada e aveludada não atribuível a outra condição; e fibrose submucosa oral, condição crônica associada a risco de transformação maligna. Fonte: elaborado pelo autor com base em (WARNAKULASURIYA; JOHNSON; WAAL, 2007; WAAL, 2009).

outra doença conhecida, configurando, portanto, um diagnóstico de exclusão (WAAL, 2009). Trata-se da OPMD mais prevalente e, embora a maioria dos casos permaneça estável, uma parcela das lesões evolui para carcinoma (WARNAKULASURIYA et al., 2008). Uma variante de comportamento particularmente agressivo é a Leucoplasia Verrucosa Proliferativa (LVP), caracterizada por crescimento multifocal, aspecto verrucoso, resistência ao tratamento e elevadas taxas de recidiva e de transformação maligna (BAGAN et al., 2010). A distinção entre essas condições é clinicamente relevante, pois implica diferentes níveis de risco e estratégias de acompanhamento.

Do ponto de vista histopatológico, o risco de malignização das OPMD está fortemente associado à presença e à intensidade de displasia epitelial, definida como um conjunto de alterações arquiteturais e citológicas do epitélio (KADEMANI, 2019). Tais alterações incluem variações no tamanho e na forma dos núcleos, hipercromasia, aumento da relação núcleo-citoplasma e perda da maturação ordenada das camadas epiteliais (KUMAR; ASTER; ABBAS, 2010; WARNAKULASURIYA et al., 2008). A graduação da displasia, contudo, é marcada por considerável subjetividade, refletida em taxas expressivas de divergência interobservador (MÜLLER, 2018). Essa variabilidade decorre, em grande parte, do fato de que a avaliação depende fortemente da análise da morfologia nuclear, processo demorado e sujeito à fadiga e à experiência do avaliador (SMITH et al., 2009; MÜLLER, 2018).

É justamente essa centralidade da morfologia nuclear que motiva o desenvolvimento de métodos automáticos de segmentação de núcleos. A delimitação precisa dessas estruturas constitui etapa prévia indispensável para a extração quantitativa de atributos

morfológicos e, conseqüentemente, para a construção de ferramentas objetivas de apoio ao diagnóstico (KADEMANI, 2019; IRSHAD et al., 2014). Tal delimitação, entretanto, depende antes de tudo da qualidade das imagens histológicas, cujo processo de obtenção é descrito na seção seguinte.

2.3 Preparação e Digitalização de Tecidos Histológicos

A histologia consiste no estudo dos tecidos e células, de sua organização e de suas funções, sendo a microscopia óptica a principal ferramenta de observação (GARTNER, 2007; JUNQUEIRA; CARNEIRO, 2013). A análise histológica é central em oncologia, pois permite identificar as alterações celulares e morfológicas associadas a lesões pré-malignas e a tumores (KADEMANI, 2019). Para viabilizar essa análise, o material biológico precisa ser convertido em cortes finos que preservem, tanto quanto possível, o aspecto original do tecido (GARTNER, 2007).

O preparo das lâminas histológicas envolve uma sequência de etapas bem estabelecidas (JOHNSON, 1991; JUNQUEIRA; CARNEIRO, 2013), sintetizadas na Figura 3. Na fixação, o tecido é imerso em soluções químicas, como o formaldeído, com o objetivo de preservar sua integridade estrutural e evitar a degradação celular. Em seguida, a desidratação e a diafanização preparam o material por meio de banhos sucessivos em álcoois de concentração crescente e solventes que o tornam translúcido. Na etapa de inclusão, o tecido é embebido em parafina, formando um bloco rígido que possibilita o corte. A microtomia consiste no fatiamento desse bloco em seções de poucos micrômetros de espessura. Por fim, os cortes são montados em lâminas de vidro e submetidos à coloração.

A coloração mais empregada na rotina histopatológica é a Hematoxilina e Eosina (H&E) (FISCHER et al., 2008). A hematoxilina confere tonalidade azul-arroxeadada aos núcleos celulares, ao se ligar a estruturas ácidas como a cromatina, enquanto a eosina cora em tons rosados o citoplasma e os componentes extracelulares (FISCHER et al., 2008; JUNQUEIRA; CARNEIRO, 2013). Esse contraste é particularmente conveniente para a segmentação de núcleos, uma vez que o realce nuclear produzido pela hematoxilina facilita a distinção entre as estruturas de interesse e o restante do tecido.

Após a coloração, as lâminas podem ser analisadas diretamente ao microscópio ou digitalizadas por meio de escâneres equipados com câmeras de alta resolução, dando origem às imagens de lâmina completa, do inglês *Whole Slide Imaging* (WSI) (NIAZI; PARWANI; GURCAN, 2019). A consolidação da patologia digital ampliou as possibilidades de análise remota, ensino e colaboração diagnóstica, além de viabilizar a aplicação de técnicas computacionais de processamento de imagens (MADABHUSHI; LEE, 2016). As imagens resultantes são tipicamente armazenadas no espaço de cores *Red, Green and Blue* (RGB), com elevada resolução espacial.

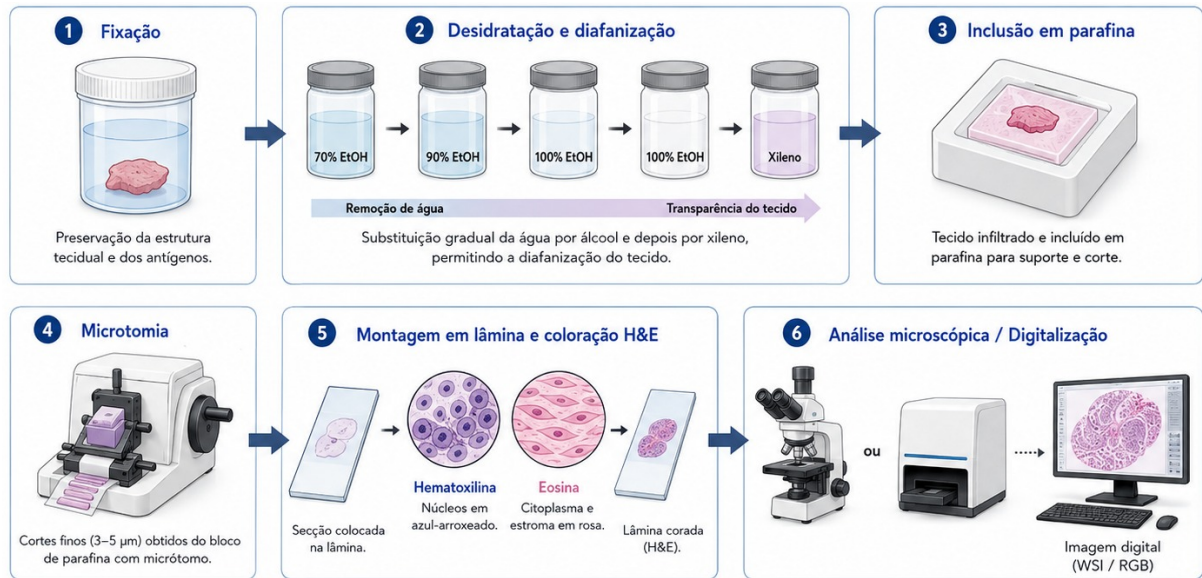


Figura 3 – Etapas de preparação e digitalização de tecidos histológicos. Na parte superior, a sequência do processamento: fixação, desidratação e diafanização, inclusão em parafina, microtomia, montagem em lâmina e coloração por hematoxilina e eosina (H&E) e, por fim, a análise ao microscópio ou a digitalização, que origina a imagem digital (WSI/RGB). Na parte inferior, as principais fontes de variabilidade do processo: variação de coloração, espessura do corte, equipamentos de digitalização e artefatos de preparo. Fonte: elaborado pelo autor com base em (JUNQUEIRA; CARNEIRO, 2013; FISCHER et al., 2008; TELLEZ et al., 2019).

Apesar dos avanços, o processo apresenta fontes relevantes de variabilidade. Diferenças nos protocolos de coloração, na espessura dos cortes, nos equipamentos de digitalização e nas condições de aquisição produzem variações de cor, brilho e contraste entre imagens de diferentes origens (MACENKO et al., 2009; TELLEZ et al., 2019). Artefatos de preparo também podem comprometer a qualidade das amostras. Essa heterogeneidade dificulta tanto a análise manual quanto a automática e motiva o uso de ferramentas computacionais robustas de apoio ao diagnóstico, apresentadas na seção seguinte.

2.4 Sistemas Computacionais de Apoio ao Diagnóstico

Os sistemas CAD combinam conhecimento médico e técnicas computacionais para apoiar a interpretação de dados clínicos, atuando como uma segunda opinião ao especialista (DOI, 2007). A complexidade do processo diagnóstico, o volume crescente de dados e os avanços em inteligência artificial impulsionaram a adoção desses sistemas na prática médica (LITJENS et al., 2017). Em histopatologia, especificamente, o uso de CAD tem como objetivo aumentar a objetividade e a reprodutibilidade das análises, mitigando a divergência interobservador e os efeitos de ruídos e variações de coloração (SRINIDHI;

CIGA; MARTEL, 2021; MADABHUSHI; LEE, 2016). A Figura 4 sintetiza o papel desses sistemas, seus principais benefícios e o fluxo típico de análise computacional de imagens histológicas, detalhado a seguir.

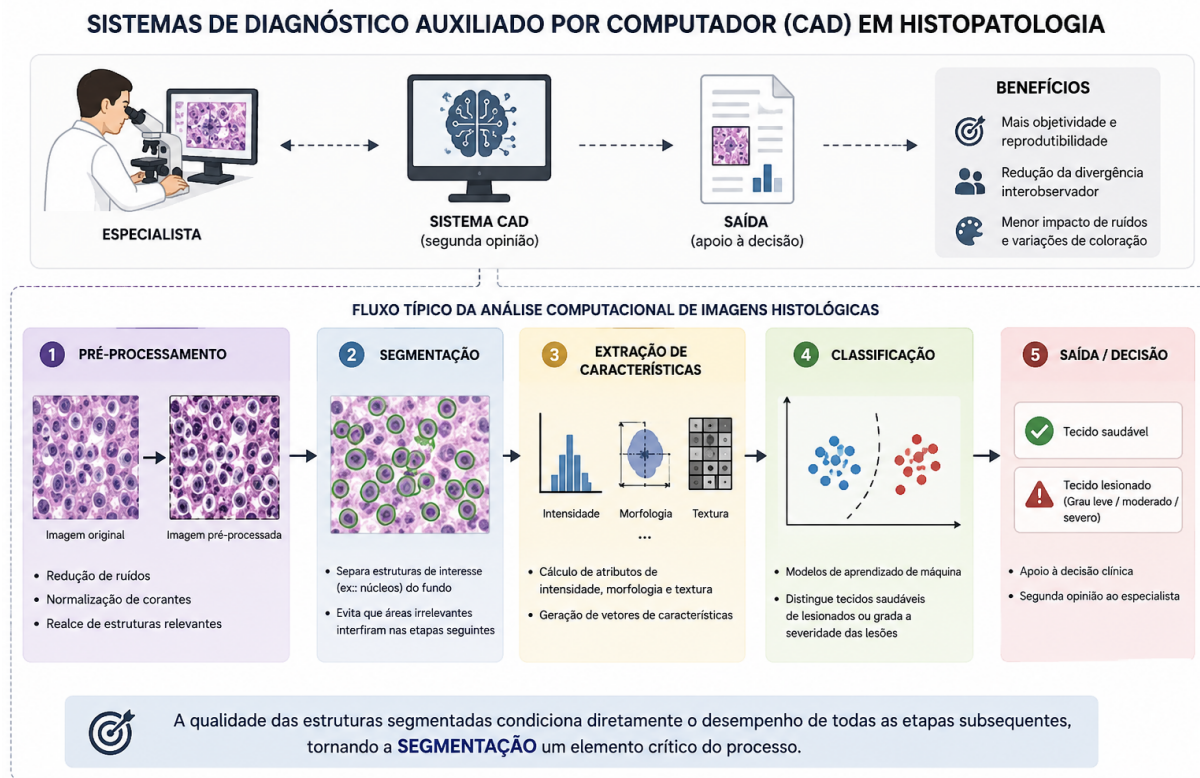


Figura 4 – Fluxo típico de um sistema de diagnóstico auxiliado por computador em histopatologia, contemplando as etapas de pré-processamento, segmentação, extração de características, classificação e apoio à decisão. Fonte: elaborado pelo autor com base em (IRSHAD et al., 2014; TELLEZ et al., 2019; KUMAR et al., 2019).

A análise computacional de imagens histológicas costuma ser organizada em etapas encadeadas: pré-processamento, segmentação, extração de características e classificação (IRSHAD et al., 2014). O pré-processamento prepara a imagem por meio da redução de ruídos, da normalização de corantes e do realce de estruturas relevantes (TELLEZ et al., 2019). A segmentação separa as estruturas de interesse, como os núcleos, da região de fundo, evitando que áreas irrelevantes interfiram nas etapas seguintes (KUMAR et al., 2019). A partir das estruturas segmentadas, a extração de características computa atributos de intensidade, morfologia e textura, posteriormente utilizados pela etapa de classificação para distinguir tecidos saudáveis de lesionados ou para graduar a severidade das lesões (IRSHAD et al., 2014). Nesse fluxo, a qualidade das estruturas delimitadas condiciona diretamente o desempenho de todas as etapas subsequentes, o que coloca a segmentação como elemento crítico do processo e foco da próxima seção.

2.5 Segmentação de Imagens

A segmentação consiste em particionar uma imagem em regiões ou objetos de interesse, atribuindo a cada *pixel* um rótulo que indica a categoria a que pertence (MINAEE et al., 2021). Em histopatologia digital, o diagnóstico depende fortemente da análise da forma, do tamanho e da disposição espacial dos núcleos, tornando a segmentação dessas estruturas uma etapa essencial para a caracterização de lesões (KUMAR et al., 2019). Trata-se, contudo, de uma tarefa desafiadora, dada a irregularidade morfológica dos tecidos, a sobreposição de núcleos e a variabilidade de coloração (GONZALEZ; WOODS, 2018; KUMAR et al., 2019).

Os métodos clássicos de segmentação podem ser agrupados em três grandes categorias. As abordagens baseadas em limiarização, como o método de Otsu, separam classes a partir de valores de intensidade, sendo eficazes apenas quando há contraste expressivo entre objetos e fundo (OTSU, 1979; GONZALEZ; WOODS, 2018). Os métodos baseados em detecção de bordas identificam mudanças abruptas de intensidade por meio de operadores de gradiente (GONZALEZ; WOODS, 2018). Já os métodos baseados em regiões empregam critérios de similaridade para agrupar ou dividir áreas da imagem (GONZALEZ; WOODS, 2018). Embora úteis, essas abordagens dependem fortemente de ajustes manuais e tendem a apresentar desempenho inconsistente diante da complexidade das imagens histológicas.

Com o avanço do aprendizado profundo, os modelos baseados em redes neurais tornaram-se a abordagem predominante para segmentação, por aprenderem automaticamente representações discriminativas a partir dos próprios dados, dispensando a engenharia manual de atributos (LONG; SHELHAMER; DARRELL, 2015; MINAEE et al., 2021). Em geral, essas redes adotam estruturas do tipo codificador-decodificador, nas quais a resolução espacial é progressivamente reduzida e depois recuperada para gerar a máscara final (RONNEBERGER; FISCHER; BROX, 2015). Esses modelos permitem abordar a segmentação sob três perspectivas: semântica, por instâncias e panóptica (KIRILLOV et al., 2019), ilustradas na Figura 5. Na segmentação semântica, todos os *pixels* de uma mesma classe recebem o mesmo rótulo, fornecendo uma visão global da estrutura do tecido, porém sem distinguir objetos individuais (MINAEE et al., 2021). A segmentação por instâncias gera máscaras individuais para cada objeto, permitindo isolar núcleos sobrepostos (HE et al., 2017), enquanto a segmentação panóptica unifica as duas abordagens (KIRILLOV et al., 2019).

Neste trabalho, adota-se a segmentação semântica binária, na qual cada *pixel* é classificado como núcleo ou fundo, concentrando a análise na capacidade dos modelos de delinear as estruturas nucleares. Esses métodos baseados em aprendizado profundo, por sua vez, derivam de princípios mais amplos de aprendizado de máquina, que fundamentam o restante deste capítulo.

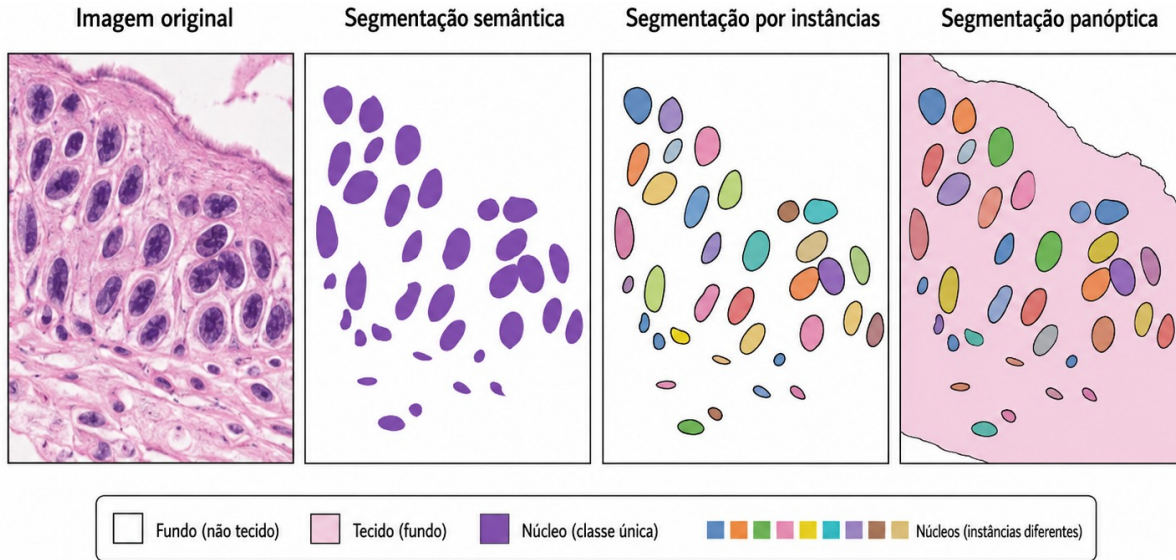


Figura 5 – Comparação dos principais tipos de segmentação em histopatologia. Da esquerda para a direita: imagem original corada com H&E; segmentação semântica, na qual todos os núcleos recebem o mesmo rótulo de classe; segmentação por instâncias, em que cada núcleo é identificado como uma instância distinta; e segmentação panóptica, que combina a segmentação semântica do tecido/fundo com a identificação de instâncias dos núcleos. Fonte: elaborado pelo autor com base em (KIRILLOV et al., 2019; MINAEE et al., 2021).

2.6 Aprendizado de Máquina

O aprendizado de máquina (do inglês, *machine learning*) é o campo da inteligência artificial que estuda algoritmos capazes de aprender padrões a partir de dados, em vez de seguir regras explicitamente programadas (MITCHELL, 1997). Em vez de codificar manualmente o conhecimento sobre um problema, esses algoritmos ajustam seus parâmetros a partir de exemplos, construindo um modelo que mapeia entradas em saídas e que, idealmente, generaliza para dados não vistos durante o treinamento (BISHOP, 2006).

As tarefas de aprendizado costumam ser divididas em paradigmas, ilustrados na Figura 6. No aprendizado supervisionado, o modelo é treinado com pares de entrada e rótulo de referência, aprendendo a prever a saída correta para novas entradas, problemas de classificação e de segmentação, como o tratado neste trabalho, enquadram-se nessa categoria (BISHOP, 2006). No aprendizado não supervisionado, busca-se descobrir estrutura nos dados sem rótulos, por exemplo por meio de agrupamento. Independentemente do paradigma, um objetivo central é alcançar boa capacidade de generalização, evitando o sobreajuste (*overfitting*), situação em que o modelo memoriza particularidades do conjunto de treinamento e perde desempenho em dados novos (MITCHELL, 1997; GOODFELLOW; BENGIO; COURVILLE, 2016).

Um aspecto determinante do desempenho desses modelos é a qualidade das representações utilizadas para descrever os dados. Em abordagens tradicionais, essas representações eram obtidas por engenharia manual de características, processo trabalhoso e dependente

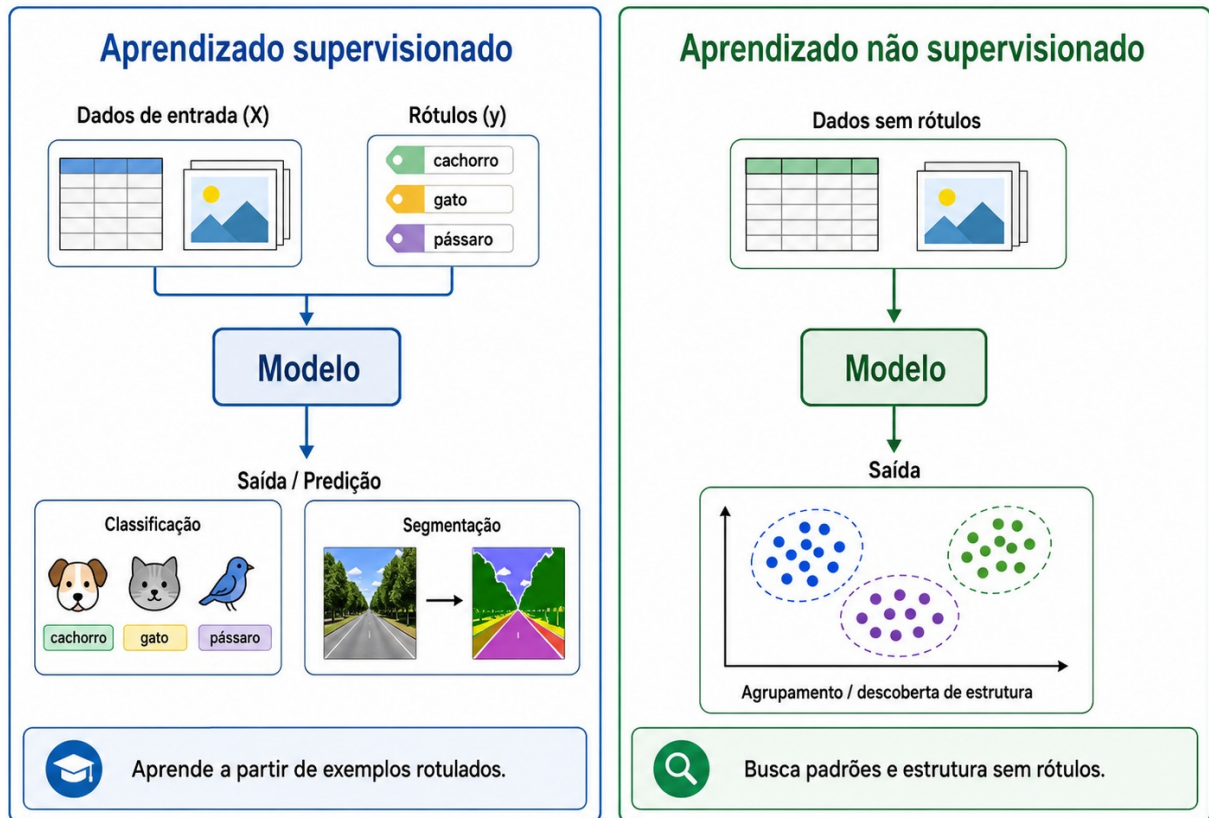


Figura 6 – Paradigmas de aprendizado de máquina. À esquerda, o aprendizado supervisionado: o modelo é treinado com dados de entrada acompanhados de rótulos de referência e aprende a produzir a saída desejada, como em tarefas de classificação e de segmentação. À direita, o aprendizado não supervisionado: a partir de dados sem rótulos, o modelo busca descobrir estrutura, por exemplo por meio de agrupamento. Fonte: elaborado pelo autor com base em (MITCHELL, 1997; BISHOP, 2006).

de conhecimento especializado (BISHOP, 2006). O aprendizado profundo (do inglês, *deep learning*) emergiu como um subcampo do aprendizado de máquina que supera essa limitação ao aprender, de forma automática e hierárquica, representações cada vez mais abstratas diretamente a partir dos dados brutos, utilizando modelos compostos por múltiplas camadas de processamento (LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016). Entre os diversos modelos de aprendizado de máquina, as redes neurais artificiais destacam-se por essa capacidade de aprender representações complexas e constituem a base do aprendizado profundo, sendo descritas a seguir.

2.7 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados na organização do sistema nervoso biológico, nos quais unidades de processamento, os neurônios artificiais, recebem, combinam e transformam sinais de entrada (HAYKIN et al., 2009). Cada neurônio é caracterizado por pesos sinápticos, que ponderam as entradas, por uma

função de agregação, que as combina, e por uma função de ativação, que introduz não linearidade à resposta, permitindo que a rede aprenda relações complexas (HAYKIN et al., 2009).

Os neurônios são organizados em camadas: uma camada de entrada, que recebe os dados, uma ou mais camadas ocultas, responsáveis por extrair representações progressivamente mais abstratas e uma camada de saída, que produz a predição final (GOODFELLOW; BENGIO; COURVILLE, 2016). O treinamento ajusta iterativamente os pesos da rede de modo a minimizar uma função de erro que mede a discrepância entre as predições e os valores de referência. Esse ajuste é realizado pelo algoritmo de retropropagação, que propaga o erro da saída em direção às camadas iniciais e o utiliza para atualizar os pesos por métodos de otimização baseados em gradiente (RUMELHART; HINTON; WILLIAMS, 1986). A Figura 7 sintetiza esses conceitos, apresentando a estrutura de um neurônio artificial e a organização de uma rede neural multicamadas, com seus fluxos de propagação direta e de retropropagação do erro.

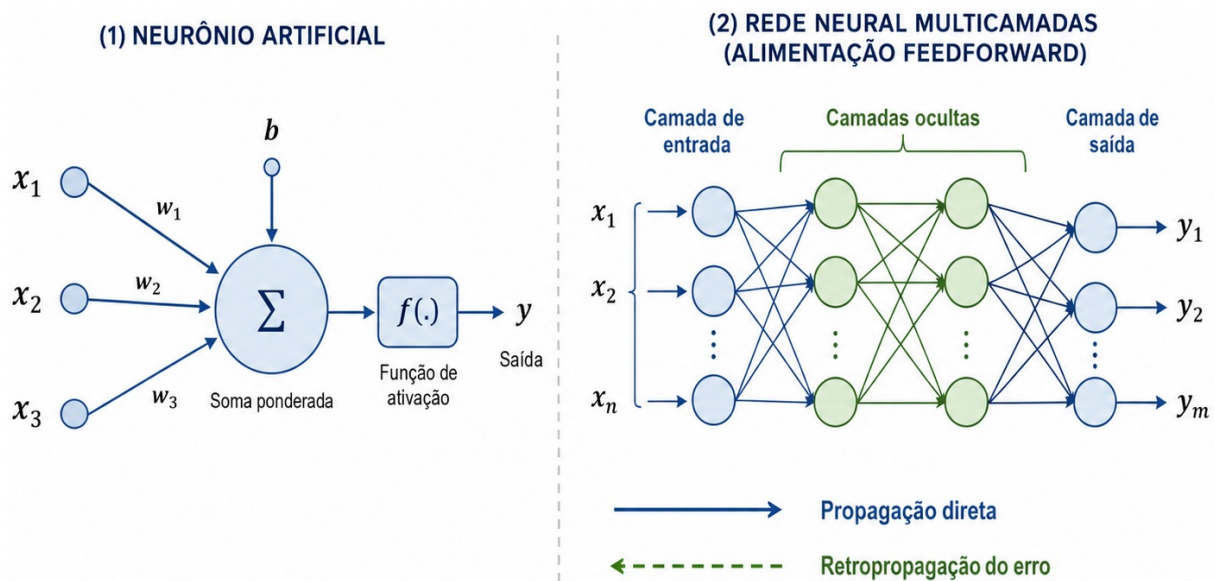


Figura 7 – Estrutura básica de uma rede neural artificial. Em (1), o neurônio artificial: as entradas x_i são ponderadas por pesos w_i e combinadas, juntamente com um viés b , por uma soma ponderada, cujo resultado passa por uma função de ativação $f(\cdot)$ que produz a saída y . Em (2), uma rede multicamadas do tipo *feedforward*, composta por camadas de entrada, ocultas e de saída, com indicação dos fluxos de propagação direta e de retropropagação do erro. Fonte: elaborado pelo autor com base em (HAYKIN et al., 2009; RUMELHART; HINTON; WILLIAMS, 1986).

Embora poderosas, as redes totalmente conectadas tornam-se impraticáveis quando aplicadas diretamente a imagens. Como cada neurônio se conecta a todos os *pixels* da entrada, o número de parâmetros cresce rapidamente, dificultando o treinamento e desconsiderando a estrutura espacial dos dados (GOODFELLOW; BENGIO; COURVILLE, 2016). Essa limitação é superada pelas redes neurais convolucionais, projetadas especifi-

camente para o processamento de imagens.

2.8 Redes Neurais Convolucionais

As *Convolutional Neural Networks* (CNNs) constituem uma especialização das RNAs voltada ao processamento de dados estruturados em grade, como imagens (LECUN et al., 1998; GOODFELLOW; BENGIO; COURVILLE, 2016). Em vez de conexões totalmente densas, as CNNs empregam camadas convolucionais, nas quais filtros percorrem a imagem extraíndo padrões locais e gerando mapas de características (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Esse mecanismo de compartilhamento de pesos reduz drasticamente o número de parâmetros e confere à rede invariância parcial à translação, além de explorar a correlação espacial entre *pixels* vizinhos.

A arquitetura típica de uma CNN combina três operações fundamentais (ALBAWI; MOHAMMED; AL-ZAWI, 2017), ilustradas na Figura 8. As camadas convolucionais extraem características locais, as funções de ativação, como a ReLU, introduzem não linearidade e atenuam o problema do desaparecimento do gradiente e as operações de *pooling* reduzem a resolução espacial dos mapas, preservando as informações mais relevantes e aumentando a robustez a pequenas variações. Pela composição sucessiva dessas operações, as camadas iniciais tendem a capturar elementos simples, como bordas e texturas, enquanto as camadas mais profundas combinam essas informações em representações semânticas complexas (LECUN; BENGIO; HINTON, 2015).

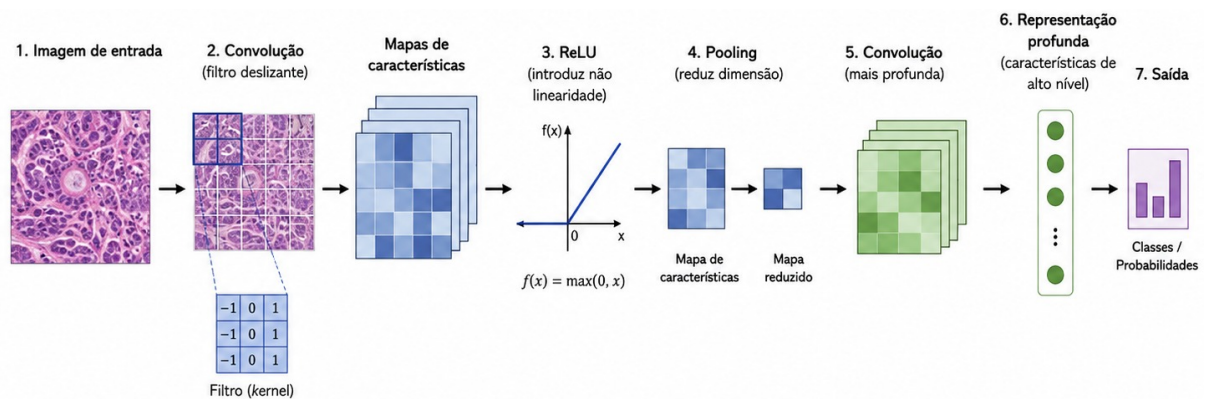


Figura 8 – Etapas típicas de uma rede neural convolucional aplicada a uma imagem histológica: a partir da imagem de entrada, filtros convolucionais geram mapas de características, sucedidos por funções de ativação não lineares (ReLU) e operações de *pooling*; a repetição desses blocos produz representações progressivamente mais abstratas, combinadas na camada de saída para gerar a predição final. Fonte: elaborado pelo autor com base em (ALBAWI; MOHAMMED; AL-ZAWI, 2017; LECUN; BENGIO; HINTON, 2015).

A consolidação das CNNs como abordagem dominante em visão computacional ocorreu a partir de marcos como a AlexNet, que demonstrou ganhos expressivos em classificação

de imagens em larga escala (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Desde então, a evolução das arquiteturas buscou treinar redes cada vez mais profundas e eficientes: as conexões residuais permitiram o treinamento de modelos muito profundos ao mitigar a degradação do gradiente (HE et al., 2016), o escalonamento composto de profundidade, largura e resolução melhorou a eficiência paramétrica (TAN; LE, 2019) e projetos convolucionais modernos incorporaram princípios inspirados em arquiteturas mais recentes, mantendo a natureza convolucional do processamento (LIU et al., 2022).

O treinamento de modelos profundos, entretanto, exige grandes volumes de dados rotulados e elevado custo computacional, restrições particularmente severas em imagens médicas, cuja anotação depende de especialistas (MINAEE et al., 2021). Uma estratégia amplamente adotada para mitigar essa limitação é a transferência de aprendizado (do inglês, *transfer learning*), na qual modelos previamente treinados em grandes bases de propósito geral são reaproveitados em tarefas com dados limitados (PAN; YANG, 2010). Na prática, é comum inicializar os codificadores com pesos obtidos na base ImageNet (DENG et al., 2009), aproveitando representações visuais genéricas. As características das primeiras camadas convolucionais mostram-se amplamente transferíveis entre domínios distintos (YOSINSKI et al., 2014), e o ajuste fino de redes pré-treinadas tende a superar o treinamento a partir do zero em aplicações de análise de imagens médicas (TAJBAKHSH et al., 2016).

Apesar de sua eficácia, as convoluções operam sobre vizinhanças locais, e seu campo receptivo cresce lentamente com a profundidade. Essa característica limita a captura de relações entre regiões distantes da imagem e motivou o desenvolvimento de mecanismos capazes de modelar dependências globais, como a atenção e os *Transformers*.

2.9 *Transformers* e Mecanismos de Atenção

Os mecanismos de atenção foram concebidos para permitir que um modelo pondere seletivamente as porções mais informativas de sua entrada, em vez de tratá-las de maneira uniforme. Em visão computacional, essa ideia manifesta-se de diferentes formas: mecanismos de atenção podem recalibrar a importância relativa dos canais de um mapa de características, enfatizando os filtros mais relevantes, bem como realçar regiões espaciais mais informativas (HU; SHEN; SUN, 2018). Tais módulos atuam como refinadores das representações aprendidas, favorecendo a supressão de respostas espúrias e o destaque das estruturas de interesse.

Uma forma especialmente influente desse princípio é a autoatenção (*self-attention*), introduzida pelos *Transformers* no contexto do processamento de linguagem natural (VASWANI et al., 2017). Diferentemente das convoluções, cujo campo receptivo é local, a autoatenção modela diretamente as relações entre todos os elementos de uma sequência, capturando dependências de longo alcance independentemente da distância. A adaptação desse me-

canismo às imagens ocorreu com o *Vision Transformer* (ViT), que divide a imagem em *patches* tratados como uma sequência de *tokens*, demonstrando desempenho competitivo quando treinado em larga escala (DOSOVITSKIY et al., 2021). A Figura 9 ilustra esse processo: a imagem é dividida em *patches* convertidos em uma sequência de *tokens*, sobre os quais a autoatenção modela as relações entre todos os elementos, produzindo representações enriquecidas com contexto global.

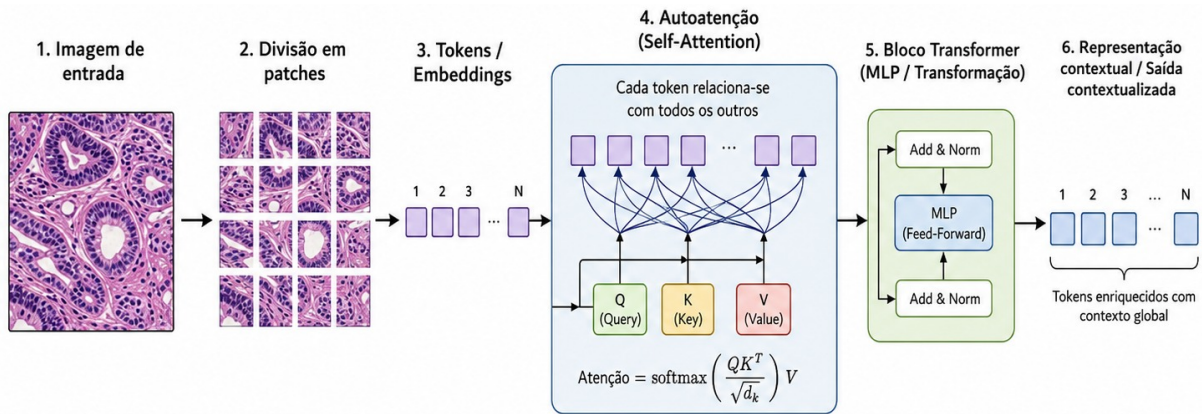


Figura 9 – Processamento de uma imagem por um *Transformer*: a imagem é dividida em *patches* convertidos em uma sequência de *tokens*; o mecanismo de autoatenção (*self-attention*), baseado em consultas, chaves e valores (Q, K e V), relaciona cada *token* a todos os demais, modelando dependências globais; os blocos subsequentes produzem representações contextualizadas. Fonte: elaborado pelo autor com base em (VASWANI et al., 2017; DOSOVITSKIY et al., 2021).

No contexto da segmentação, arquiteturas baseadas em *Transformers* exploram essa capacidade de modelar contexto global, mostrando-se vantajosas quando a tarefa depende não apenas da textura local, mas também do arranjo morfológico mais amplo das estruturas (XIE et al., 2021). Abordagens mais recentes propõem ainda combinar atenção local e global de maneira eficiente, produzindo representações hierárquicas adequadas a tarefas densas (TU et al., 2022). Como as abordagens convolucionais e baseadas em atenção possuem forças complementares, detalhe local em um caso, contexto global no outro, surgiram arquiteturas que buscam reuni-las, tema da próxima seção.

2.10 Arquiteturas Híbridas

As arquiteturas híbridas combinam, em um mesmo modelo, componentes convolucionais e mecanismos baseados em atenção, com o objetivo de aproveitar simultaneamente a sensibilidade das convoluções a detalhes locais e a capacidade da autoatenção de capturar relações globais (SHAMSHAD et al., 2023). Essa complementaridade é particularmente atraente em imagens histopatológicas, nas quais a identificação dos núcleos depende tanto de padrões finos de textura e de borda quanto da organização espacial das estruturas no tecido.

Diferentes estratégias de hibridização têm sido propostas na literatura, ilustradas na Figura 10. Em uma vertente, redes convolucionais atuam como extratoras de características locais e um módulo *Transformer* é inserido para modelar dependências de longo alcance entre regiões distantes, antes da reconstrução da máscara (CHEN et al., 2021). Em outra, codificadores hierárquicos baseados em *Transformer* fornecem representações contextuais ricas que alimentam decodificadores voltados à recuperação espacial, em uma combinação entre modelagem global e mecanismos de agregação multiescala (FAN et al., 2020). De modo geral, esses modelos buscam equilibrar a captura de contexto e a preservação de contornos, dois requisitos frequentemente conflitantes na segmentação de estruturas pequenas e irregulares (SHAMSHAD et al., 2023). As arquiteturas híbridas concretamente avaliadas neste trabalho, bem como suas configurações, são descritas no Capítulo de Metodologia.

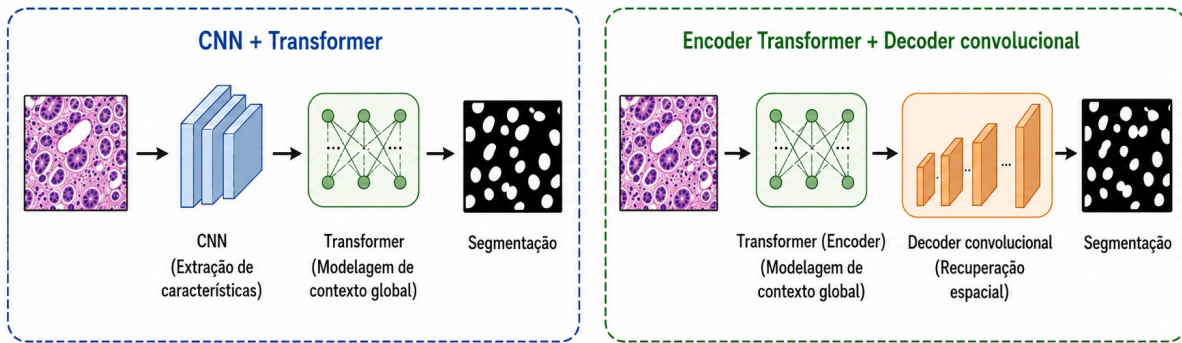


Figura 10 – Duas estratégias de hibridização entre redes convolucionais e *Transformers* para segmentação. À esquerda (*CNN + Transformer*), uma rede convolucional extrai características locais e um módulo *Transformer* modela o contexto global antes da geração da máscara. À direita (*encoder Transformer + decodificador convolucional*), um codificador baseado em *Transformer* fornece representações contextuais que alimentam um decodificador convolucional responsável pela recuperação espacial. Fonte: elaborado pelo autor com base em (CHEN et al., 2021; FAN et al., 2020).

Para além do aprimoramento de um único modelo, seja ele convolucional, baseado em atenção ou híbrido, ganhos adicionais de desempenho e robustez podem ser obtidos pela combinação das predições de múltiplos modelos. Esse princípio fundamenta os comitês, abordados a seguir.

2.11 Comitês de Aprendizagem Profunda

Comitês (do inglês, *ensembles*) consistem na combinação das predições de dois ou mais modelos com o objetivo de obter resultados mais robustos e confiáveis do que os produzidos individualmente (DIETTERICH, 2000; SAGI; ROKACH, 2018). A motivação para essa abordagem fundamenta-se na premissa de que os erros cometidos por um

modelo podem ser compensados pelos acertos dos demais, desde que os modelos apresentem comportamentos suficientemente distintos (DIETTERICH, 2000). Do ponto de vista estatístico, a combinação de modelos tende a reduzir a variância das predições e a atenuar o sobreajuste, sendo especialmente benéfica em cenários com dados ruidosos ou desbalanceados (SAGI; ROKACH, 2018; GANAIE et al., 2022).

A eficácia de um comitê depende criticamente da diversidade entre seus membros, uma vez que modelos muito semelhantes tendem a produzir erros correlacionados, limitando o ganho da combinação (DIETTERICH, 2000). Essa diversidade pode ser introduzida por meio de variações nos dados de treinamento ou da utilização de diferentes arquiteturas e hiperparâmetros (GANAIE et al., 2022). Entre as estratégias clássicas de construção de comitês destacam-se o *bagging*, no qual modelos são treinados de forma independente sobre subconjuntos amostrados dos dados, reduzindo a variância (BREIMAN, 1996), e o *boosting*, em que modelos são treinados sequencialmente para corrigir os erros dos anteriores, reduzindo o viés (FREUND; SCHAPIRE, 1997; FRIEDMAN, 2001). Neste trabalho, a diversidade é obtida fundamentalmente pela heterogeneidade arquitetural, combinando modelos convolucionais, híbridos e baseados em atenção treinados sobre os mesmos dados. A Figura 11 ilustra o funcionamento geral de um comitê aplicado à segmentação de núcleos. A imagem histológica é processada por modelos de arquiteturas distintas, cujas predições individuais são então combinadas por diferentes regras de fusão para produzir a máscara final.

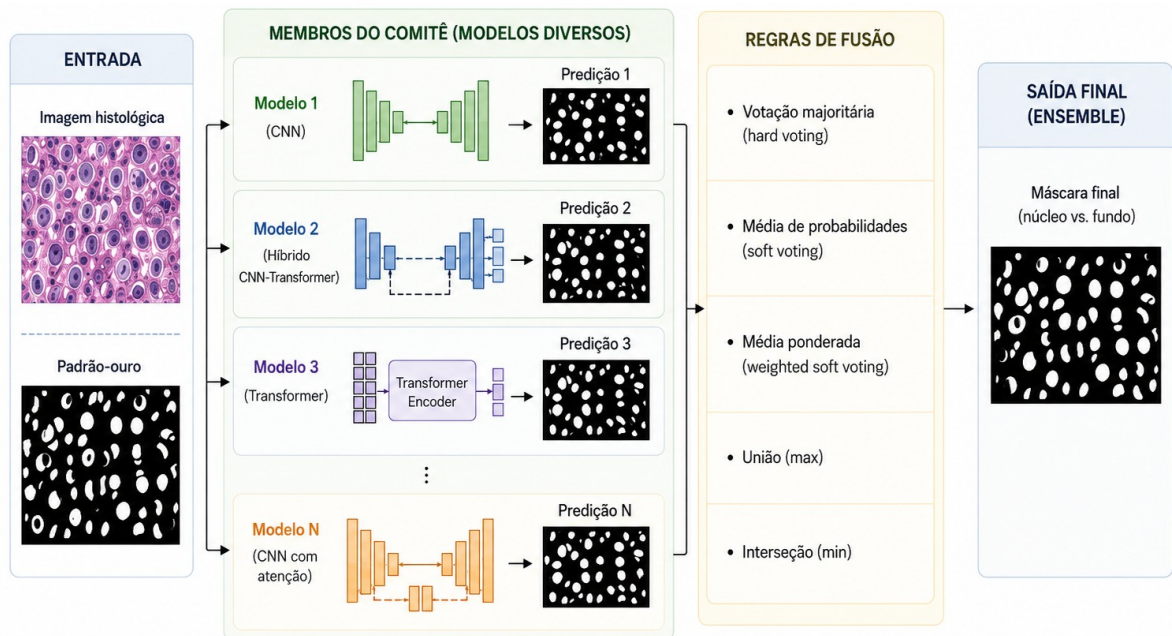


Figura 11 – Representação esquemática de um comitê de modelos de aprendizagem profunda para segmentação de núcleos. A imagem histológica é processada por múltiplos modelos de arquiteturas distintas (convolucionais, híbridas e baseadas em atenção), cujas predições individuais são combinadas por diferentes regras de fusão, gerando a máscara final. Fonte: elaborado pelo autor com base em (DIETTERICH, 2000; KITTLER et al., 1998).

Uma vez definidos os membros do comitê, diferentes regras de fusão podem ser empregadas para combinar suas saídas (KITTLER et al., 1998). Em problemas de classificação e segmentação, as estratégias mais comuns incluem a votação majoritária, na qual prevalece a decisão da maioria dos modelos, e a combinação das saídas probabilísticas, por meio de médias simples ou ponderadas, que preserva a informação de confiança de cada modelo e tende a suavizar regiões de incerteza (KITTLER et al., 1998; SAGI; ROKACH, 2018). As regras de fusão específicas adotadas neste trabalho são formalizadas no Capítulo de Metodologia.

No campo da segmentação de imagens médicas, os comitês têm-se mostrado uma estratégia eficaz para lidar com ruídos, baixa qualidade, sobreposição de estruturas e variabilidade interobservador, melhorando tanto a robustez quanto a generalização das predições (KAMNITSAS et al., 2018).

Estado da Arte

A análise automatizada de imagens histológicas tem sido progressivamente incorporada a fluxos computacionais voltados ao apoio diagnóstico, prognóstico e à quantificação morfológica de tecidos. Nesse contexto, a segmentação de estruturas celulares, em especial de núcleos, constitui uma etapa fundamental, pois permite extrair medidas associadas à forma, ao tamanho, à distribuição espacial e à organização tecidual. Tais informações são particularmente relevantes em lesões orais potencialmente malignas, nas quais alterações morfológicas do epitélio podem estar relacionadas à progressão da doença.

Este capítulo reúne trabalhos relacionados ao problema investigado nesta dissertação, organizados em duas frentes principais. A primeira apresenta estudos que aplicam modelos de aprendizagem profunda à segmentação de estruturas em imagens histológicas de lesões orais, incluindo displasia epitelial oral, desordens potencialmente malignas e carcinoma espinocelular oral. A segunda discute trabalhos que empregam estratégias de *ensemble learning* em tarefas de segmentação histológica, com ênfase na segmentação de núcleos celulares. Considerando que a literatura dedicada especificamente à segmentação de núcleos em leucoplasia bucal ainda é limitada, foram incluídos estudos correlatos em lesões orais e em segmentação nuclear histopatológica, de modo a contextualizar metodologicamente a proposta deste trabalho sem afastar-se de seu escopo central.

3.1 Aprendizagem Profunda para Segmentação em Imagens Histológicas Oraís

A delimitação precisa de núcleos celulares e de estruturas epiteliais em lâminas histológicas é uma etapa importante para a quantificação morfológica e para a análise computacional de lesões orais. Trata-se, entretanto, de uma tarefa desafiadora, uma vez que os núcleos podem apresentar variações significativas de forma, tamanho, densidade, coloração e organização espacial. Além disso, a presença de sobreposição celular, artefatos de preparo, heterogeneidade tecidual e variações de coloração em imagens coradas com

hematoxilina e eosina (H&E) dificulta a definição precisa dos limites entre estruturas celulares e região de fundo.

Nesse cenário, modelos de aprendizagem profunda destacam-se por sua capacidade de aprender representações diretamente a partir dos dados, reduzindo a dependência de descritores morfológicos definidos manualmente. Em imagens histológicas orais, essas abordagens têm sido empregadas tanto para segmentação de núcleos quanto para delimitação de regiões epiteliais, camadas celulares e áreas displásicas. Como ainda são escassos os estudos voltados especificamente à leucoplasia bucal, esta seção apresenta trabalhos relacionados à segmentação em lesões orais e em condições histopatológicas próximas ao domínio investigado nesta dissertação.

No trabalho de Silva et al. (2022), foi proposta uma metodologia computacional para análise e quantificação de Displasia Epitelial Oral (DEO) em imagens histológicas coradas com hematoxilina e eosina (H&E). A abordagem foi organizada em quatro etapas principais: segmentação de núcleos, pós-processamento, extração de características e classificação dos graus da lesão. Na etapa de segmentação, os autores empregaram a arquitetura Mask R-CNN para detectar e segmentar instâncias de núcleos celulares. Em seguida, operações morfológicas foram aplicadas com o objetivo de reduzir regiões falso-positivas e falso-negativas. A partir dos núcleos segmentados, foram extraídas características morfológicas e não morfológicas, incluindo área, orientação, solidez, entropia e índice de Moran, posteriormente fornecidas a um classificador polinomial para distinguir os diferentes graus da lesão. A metodologia foi avaliada em um conjunto composto por 296 regiões de imagens histológicas de língua de camundongos, nas quais 9.155 núcleos celulares foram identificados e analisados. Na etapa de segmentação, o método apresentou acurácias entre 88,92% e 90,35%, enquanto a classificação dos graus da displasia obteve área sob a curva ROC (AUC) entre 0,88 e 0,97. Esse estudo demonstra a relevância da segmentação nuclear como etapa inicial para a extração de descritores quantitativos associados à graduação de lesões orais potencialmente malignas.

Em Azarmehr et al. (2022), foi apresentado um *framework* baseado em *Neural Architecture Search* (NAS) para segmentação do epitélio, dos núcleos individuais e do estroma em imagens histopatológicas de DEO. O objetivo do estudo foi otimizar automaticamente uma arquitetura eficiente para análise de imagens de lâmina completa, considerando que características arquiteturais, citológicas e histológicas da displasia podem ser modeladas a partir da segmentação dessas estruturas. A arquitetura derivada por NAS alcançou F_1 de 93,5% para a segmentação completa do epitélio e 94,5% para a segmentação nuclear, superando modelos do estado da arte avaliados pelos autores. A partir dos núcleos segmentados, foram extraídas características morfométricas utilizadas por um classificador Random Forest para diferenciar lesões de baixo e alto risco. Esse trabalho reforça a importância da segmentação conjunta de estruturas epiteliais e nucleares em DEO, além de evidenciar o potencial de estratégias automatizadas de busca arquitetural para otimizar

modelos de segmentação em histopatologia oral.

O estudo de Zhang et al. (2023) propôs uma abordagem baseada em aprendizagem profunda para segmentar o epitélio oral e quantificar o número de camadas celulares em WSI coradas com H&E. A metodologia combinou dois modelos: uma U-Net, com *backbone* MobileNetV2, responsável por delimitar o tecido epitelial e a camada de células basais, e o modelo HD-Staining, baseado em Mask R-CNN, encarregado da segmentação dos núcleos. Os resultados dos dois modelos foram integrados de modo que a classificação das células fosse proveniente da U-Net, enquanto a segmentação nuclear fosse fornecida pelo HD-Staining. A partir das células segmentadas, os autores introduziram o algoritmo *Onion Peeling*, que utiliza um grafo de Delaunay para identificar as bordas interna e externa do epitélio e contar as camadas celulares de forma iterativa. A abordagem foi avaliada em uma base composta por lâminas de mucosa oral normal e por casos de distúrbios potencialmente malignos orais (OPMD) de 135 pacientes. A U-Net alcançou IoU de 0,70 e acurácia de 92%, enquanto o HD-Staining obteve IoU de 0,76. A contagem automática de camadas apresentou coeficiente de determinação (R^2) de 0,849 em relação à contagem manual, e o número de camadas diferiu significativamente entre os graus de displasia. Esses resultados indicam que a segmentação de estruturas epiteliais e nucleares pode apoiar a quantificação morfológica de alterações associadas à progressão de lesões orais.

No estudo de Shephard et al. (2024), foi proposto um *pipeline* totalmente automatizado para a predição da transformação maligna em DEO, no qual a segmentação atua como uma etapa intermediária da análise. Para isso, foi treinado um modelo HoVer-Net+, capaz de segmentar simultaneamente os núcleos celulares e as camadas intraepiteliais em imagens de lâmina completa, do inglês *whole slide images* (WSI), coradas com H&E. Na avaliação interna da segmentação, o modelo obteve coeficiente Dice (D_C) de 0,69 para os núcleos e F_1 médio de 83% para a segmentação semântica das camadas intraepiteliais. A partir das estruturas segmentadas, foram extraídas características morfológicas e espaciais, posteriormente fornecidas a uma rede rasa responsável por estimar um escore de transformação maligna. O modelo foi desenvolvido a partir de uma coorte interna de 193 casos e validado externamente em duas coortes independentes, totalizando 89 casos. Na tarefa prognóstica, alcançou área sob a curva (AUROC) de 0,75, superando sistemas convencionais. Esse estudo evidencia que a segmentação de núcleos e de camadas epiteliais pode fornecer informações morfológicas relevantes para tarefas de predição clínica em lesões orais potencialmente malignas.

Em Silva et al. (2024), foi introduzida a *OralEpitheliumDB*, uma base pública e anotada de imagens histológicas de DEO, proposta com o objetivo de favorecer a reprodutibilidade de experimentos em um domínio frequentemente limitado pelo uso de dados privados. A base reúne 456 imagens obtidas de 30 línguas de camundongos, categorizadas entre epitélio saudável e DEO leve, moderada e grave. Os núcleos celulares foram manualmente marcados por um especialista treinado e posteriormente validados por um

patologista. Para demonstrar o potencial da base, os autores conduziram experimentos de segmentação semântica, com U-Net e FPN, e de segmentação de instâncias, com Mask R-CNN, HoVer-Net, StarDist e SOLO V2. Também foram avaliados métodos de normalização de cor, como Macenko, Reinhard, Tosta e Vahadane, seguidos de etapas de classificação dos graus da lesão. Na segmentação de núcleos, o melhor resultado individual foi obtido com a U-Net utilizando *backbone* ResNet-50, com F_1 de 83% e acurácia de 92,38%. A classificação dos graus da lesão atingiu acurácia de 94,22% com o algoritmo Random Forest. Os autores observaram ainda que a normalização de cor pode degradar a segmentação, dependendo da imagem de referência empregada, ressaltando a sensibilidade dos modelos às variações cromáticas em histopatologia.

O trabalho de Shephard et al. (2023) propôs um *pipeline* baseado em *Transformers* para a detecção e segmentação de DEO em WSI coradas com H&E. O modelo foi construído sobre a arquitetura TransUNet, de natureza híbrida, na qual uma rede convolucional, baseada em ResNet50, atua como extratora de características, enquanto camadas *Transformer* capturam dependências espaciais de longo alcance. O decodificador, por sua vez, reconstrói o mapa de segmentação por meio de operações de *upsampling* em cascata e conexões de salto. O modelo foi treinado com 260 lâminas de DEO e 105 lâminas de controle, adquiridas em três *scanners* distintos, e validado externamente em 78 casos provenientes de três centros do Reino Unido e do Brasil. Comparado a arquiteturas consolidadas, como U-Net, Swin-UNet, DeepLabV3+ e HoVer-Net+, o método obteve F_1 de 81% e sensibilidade de 84,5% no teste interno, com especificidade de 99,8% nas lâminas de controle. Na validação externa, alcançou F_1 de 71% e sensibilidade de 76,4%, demonstrando capacidade de generalização em dados provenientes de diferentes centros. Esse estudo evidencia o potencial de arquiteturas híbridas e baseadas em atenção para a delimitação de regiões displásicas em lesões orais pré-malignas.

3.2 *Ensembles* de Aprendizagem Profunda para Segmentação de Imagens Histológicas

Modelos de segmentação baseados em aprendizagem profunda podem apresentar comportamentos distintos dependendo de sua arquitetura, do tipo de representação aprendido e do domínio de aplicação. Redes convolucionais tendem a apresentar forte viés indutivo local, favorecendo a descrição de texturas, bordas e padrões espaciais próximos. Em contrapartida, arquiteturas baseadas em atenção foram projetadas para modelar relações de maior alcance e incorporar informações contextuais mais amplas. Modelos híbridos, por sua vez, buscam combinar essas duas propriedades, explorando simultaneamente descritores locais e dependências globais.

Essa diversidade de representações é particularmente relevante em imagens histológicas, nas quais a segmentação de núcleos depende tanto de detalhes finos de borda e

coloração quanto da organização espacial das células no tecido. Nesse contexto, estratégias de *ensemble learning* têm sido empregadas com o objetivo de combinar previsões de múltiplos modelos, explorando a complementaridade entre eles. A premissa central é que modelos com diferentes vieses indutivos ou diferentes condições de treinamento podem cometer erros distintos, de modo que a combinação de suas saídas tende a produzir segmentações mais robustas do que aquelas obtidas por modelos individuais (KUNCHEVA, 2003).

O estudo de Silva et al. (2023) propôs um *ensemble* de modelos convolucionais para a segmentação semântica de núcleos celulares em imagens histológicas de DEO. A abordagem combinou sete modelos de segmentação baseados em CNN por meio de três estratégias de fusão: média simples, média ponderada e votação majoritária. As previsões combinadas foram avaliadas em imagens de DEO e em três bases públicas, atingindo acurácias de 92,14%, 91,21% e 90,67% para as estratégias de média simples, média ponderada e votação majoritária, respectivamente. Os resultados indicaram que a combinação de modelos é capaz de aprimorar o desempenho da segmentação em relação aos modelos individuais, reforçando a aplicabilidade de *ensembles* em imagens histológicas orais.

No trabalho de Prusty et al. (2023), foi proposto um *ensemble* baseado na arquitetura U-Net para a segmentação de núcleos em imagens histopatológicas coradas com H&E. O *ensemble* foi construído a partir de variantes da U-Net com diferentes *backbones* convolucionais como codificadores, especificamente ResNet101, InceptionResNetV2 e DenseNet121. A metodologia também incorporou uma etapa de normalização de cor com preservação de estrutura (SPCN) no pré-processamento e aumento de dados em tempo de teste no pós-processamento. O método foi avaliado em conjuntos de dados de órgão único e de múltiplos órgãos, obtendo F_1 de 84,11%, IoU médio de 0,8167 e acurácia de 92,58% na base multiórgão, além de F_1 de 87,04%, IoU médio de 0,8666 e acurácia de 96,69% na base de órgão único. Os resultados indicaram que a combinação de modelos com diferentes *backbones*, associada a estratégias adequadas de pré-processamento e pós-processamento, pode aprimorar a segmentação de núcleos em imagens histopatológicas.

O estudo de Gudhe et al. (2023) propôs um *ensemble* de natureza bayesiana para a segmentação de instâncias de núcleos em imagens histopatológicas. A abordagem emprega *dropout* de Monte Carlo ativado no momento da inferência, de modo que múltiplas previsões estocásticas de um mesmo modelo são combinadas para produzir a segmentação final e, simultaneamente, quantificar a incerteza epistêmica associada às previsões. O método foi avaliado na base pública PanNuke, composta por 312 campos visuais provenientes de 19 tipos de órgãos, e comparado a uma rede totalmente convolucional, à U-Net, à SegNet e ao HoVer-Net. O *ensemble* bayesiano obteve F_1 médio de 0,893 e IoU de 0,868, superando o HoVer-Net, que alcançou F_1 de 0,871 e IoU de 0,840. Além da segmentação, a abordagem forneceu uma estimativa de confiança das previsões, aspecto relevante em aplicações nas quais a incerteza do modelo pode auxiliar a interpretação dos resultados.

Com foco na capacidade de generalização, o trabalho de Mahbod et al. (2024) propôs um *ensemble* para a segmentação de instâncias de núcleos em imagens histopatológicas, combinando um modelo de aprendizagem profunda do estado da arte com estratégias de normalização de cor não determinística no treinamento e determinística no teste. As predições foram agregadas por meio de *ensembling*. O modelo foi treinado em um único conjunto de treinamento e avaliado em sete bases de teste independentes, simulando cenários com dados não vistos. A estratégia proposta proporcionou ganhos médios de até 0,049 no coeficiente Dice (D_C), 0,054 no índice de Jaccard agregado e 0,059 na qualidade panóptica em relação ao modelo de base. Esses resultados evidenciam que a combinação de modelos, quando associada ao controle da variabilidade de coloração, pode aumentar a robustez da segmentação frente a diferenças entre bases histológicas.

De forma complementar, Silva et al. (2024) investigou um *ensemble* de modelos de segmentação semântica para imagens de DEO, combinando seis arquiteturas convolucionais, incluindo FPN, U-Net e PSPNet, por meio das mesmas três estratégias de fusão. Adicionalmente, foi incorporada uma etapa de aumento de dados baseada em *Generative Adversarial Network* (GAN), com o objetivo de ampliar a variabilidade do conjunto de treinamento e reforçar a robustez do *ensemble*. O método foi desenvolvido sobre a base *OralEpitheliumDB*, composta por 456 imagens histológicas de DEO de língua de camundongos, e avaliado nas bases públicas CryoNuSeg, MoNuSeg e GlaS. A abordagem proposta alcançou acurácia de 93,41% e D_C de 0,88, evidenciando que a combinação entre aumento de dados e *ensemble learning* pode contribuir para a segmentação de núcleos celulares em imagens histológicas.

Mais recentemente, Silva et al. (2025) propuseram um *ensemble* de redes convolucionais para a segmentação de instâncias de núcleos em imagens histológicas de displasia epitelial oral (DEO). O estudo parte da constatação de que a segmentação nuclear nesse domínio é dificultada por variações de coloração, diferenças morfológicas entre os tecidos e presença de artefatos, fatores que podem comprometer o desempenho de modelos individuais. Para mitigar essas limitações, os autores investigaram diferentes modelos de segmentação e estratégias de combinação baseadas em regras de votação e fusão das predições. Entre as abordagens avaliadas, a média ponderada pelo coeficiente Dice (D_C) apresentou o melhor desempenho, alcançando acurácia de 94,09% e D_C de 0,9461. Os resultados indicaram que a combinação de modelos pode superar o desempenho das arquiteturas individuais e reforçaram o potencial dos *ensembles* para aumentar a robustez da segmentação de núcleos em imagens histológicas de DEO.

3.3 Considerações

Os trabalhos apresentados neste capítulo demonstram que a aprendizagem profunda tem se consolidado como uma abordagem relevante para a segmentação de estruturas em

imagens histológicas orais. Estudos recentes têm explorado modelos convolucionais, arquiteturas híbridas e redes baseadas em atenção para segmentar núcleos, epitélio, camadas celulares e regiões displásicas. Esses avanços evidenciam que a segmentação computacional pode fornecer informações morfológicas importantes para tarefas de quantificação, classificação e prognóstico em lesões orais. De modo semelhante, os trabalhos baseados em *ensemble learning* indicam que a combinação de múltiplas predições pode elevar a qualidade das segmentações, reduzir erros individuais e aumentar a robustez dos modelos em diferentes bases histopatológicas.

Apesar desses avanços, observa-se que a literatura ainda apresenta limitações importantes em relação ao problema tratado nesta dissertação. Grande parte dos estudos concentra-se em displasia epitelial oral, desordens potencialmente malignas de forma ampla ou outros domínios histopatológicos, enquanto a segmentação de núcleos celulares em leucoplasia bucal permanece pouco explorada. Essa lacuna é relevante porque a leucoplasia possui características histológicas próprias e constitui uma desordem potencialmente maligna da cavidade oral, exigindo métodos capazes de lidar com variações morfológicas, cromáticas e estruturais presentes nesse domínio específico.

Além da limitação relacionada ao domínio de aplicação, também se observa que muitos trabalhos avaliam arquiteturas isoladas, famílias específicas de modelos ou combinações formadas por redes de natureza semelhante. Embora essas abordagens tenham demonstrado resultados relevantes, ainda são menos frequentes investigações que comparem, sob um mesmo protocolo experimental, arquiteturas convolucionais, híbridas e baseadas em *Transformers*. Essa comparação é importante porque diferentes paradigmas arquiteturais apresentam vieses indutivos distintos e podem capturar aspectos complementares das imagens histológicas. Enquanto redes convolucionais tendem a favorecer padrões locais de textura, borda e coloração, mecanismos de atenção e modelos híbridos podem incorporar relações espaciais mais amplas e dependências contextuais entre regiões da imagem.

No contexto dos *ensembles*, a literatura revisada mostra que a combinação de modelos é uma estratégia promissora para aprimorar a segmentação histológica. Entretanto, a avaliação dessas abordagens costuma concentrar-se principalmente em métricas finais de desempenho, como Dice, IoU, acurácia, sensibilidade e especificidade. Embora essas métricas sejam fundamentais para quantificar a qualidade global da segmentação, elas não esclarecem de forma detalhada como cada modelo contribui para a decisão combinada. Assim, permanece pouco explorado se o ganho obtido pelo *ensemble* decorre da preservação de acertos individuais, da correção de erros específicos, da redução de falsos positivos, da mitigação de falsos negativos ou da suavização das predições probabilísticas.

Diante desse cenário, esta dissertação propõe investigar a segmentação semântica de núcleos celulares em imagens histológicas de leucoplasia bucal por meio de arquiteturas de aprendizagem profunda de diferentes naturezas, incluindo modelos convolucionais, híbridos e baseados em atenção. Além da avaliação individual dessas arquiteturas, são

exploradas estratégias de *ensemble learning* para combinar suas predições, considerando diferentes formas de fusão. A proposta também incorpora uma análise específica da contribuição dos modelos no processo de combinação, permitindo avaliar não apenas o desempenho final das segmentações, mas também o papel desempenhado por cada arquitetura na construção da decisão conjunta. Assim, o trabalho busca contribuir tanto para o avanço da segmentação histológica em leucoplasia bucal quanto para a compreensão do uso de *ensembles* heterogêneos em imagens biomédicas.

Metodologia

A metodologia proposta neste trabalho baseia-se em modelos de aprendizado profundo aplicados à segmentação semântica, como CNNs e *Transformers*, integrados por meio de estratégias de *ensemble learning*. Inicialmente, são utilizadas bases de dados compostas por imagens histológicas e suas respectivas máscaras de referência. Posteriormente, diferentes arquiteturas de segmentação são treinadas e avaliadas individualmente. Nesta etapa, são aplicadas técnicas de aumento de dados com o objetivo de ampliar a variabilidade do conjunto de treinamento. Na etapa seguinte, as predições dos modelos são combinadas utilizando diferentes estratégias de *ensemble*, permitindo explorar a complementaridade entre as arquiteturas. Por fim, os resultados são avaliados por meio de métricas de segmentação e métricas específicas de análise de contribuição dos diversos modelos *ensembles* investigados. Na Figura 12, é apresentada uma visão geral do *pipeline* adotado neste trabalho e essas etapas são detalhadas nas seções subsequentes.

Para o treinamento, todas as imagens e máscaras foram redimensionadas para 256×256 *pixels*. Os codificadores e *backbones* foram inicializados com pesos pré-treinados na ImageNet, e cada arquitetura foi treinada por 100 épocas, com tamanho de lote igual a 4. Utilizou-se o otimizador AdamW, com taxa de aprendizado inicial de 1×10^{-4} , associado ao escalonador *Cosine Annealing*, configurado com $T_{\max} = 100$. A função objetivo foi composta por 60% da perda Lovász, calculada individualmente por imagem, e 40% da perda Focal, configurada com $\alpha = 0,25$. O treinamento foi realizado com precisão mista por meio de *automatic mixed precision* e *gradient scaling*. As estratégias de aumento de dados foram aplicadas exclusivamente ao conjunto de treinamento, enquanto o conjunto de validação recebeu apenas redimensionamento. Ao final de cada época, calculou-se D_C no conjunto de validação após a aplicação da função sigmoide e da binarização das predições com limiar de 0,5, sendo preservado como *checkpoint* final o modelo que apresentou o maior D_C de validação.

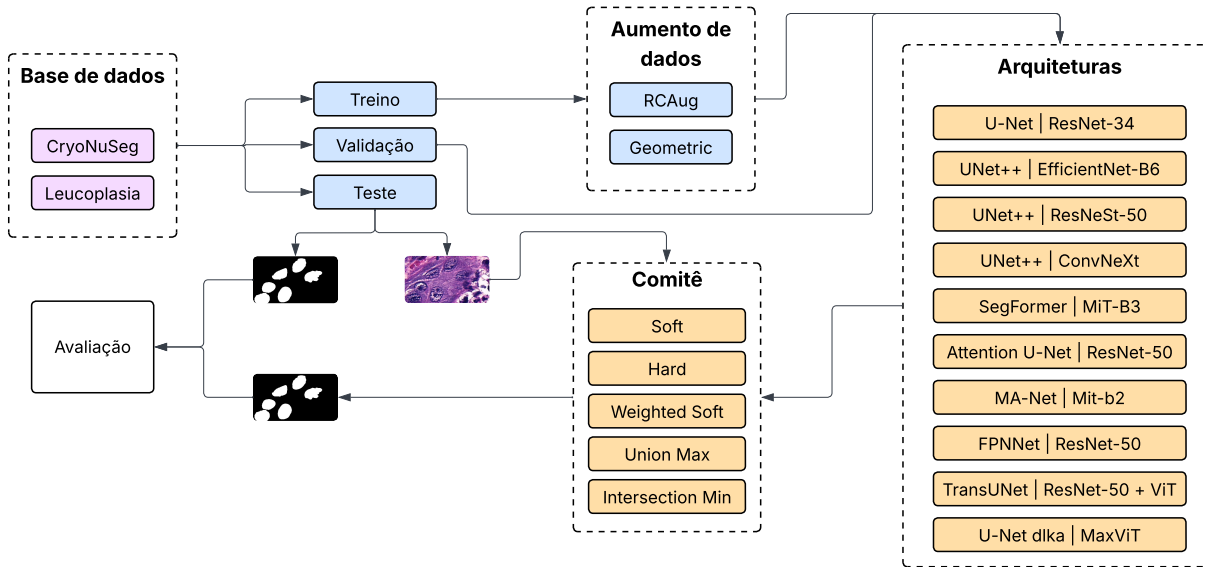


Figura 12 – Fluxograma representando o pipeline da aproximação proposta para segmentação usando modelos *ensembles* com arquiteturas de aprendizagem profunda.

4.1 Bases de Imagens

Neste trabalho, foram utilizados dois conjuntos de dados: um *dataset* específico de lesões de leucoplasia bucal e o *dataset* público CryoNuSeg. Esse segundo *dataset* foi aplicado com o objetivo de avaliar a capacidade de generalização dos modelos propostos.

O primeiro conjunto de dados, denominado Leuco, ilustrado na Figura 13 (A), é composto por 415 imagens histopatológicas de lesões bucais, obtidas a partir de lâminas coradas com hematoxilina e eosina (H&E), adquiridas por microscopia óptica com ampliação de 400×. A obtenção e a utilização dessas imagens foram aprovadas pelo Comitê de Ética em Pesquisa da Universidade Federal de Uberlândia (CAAE: 60858016.1.0000.5152), conforme o parecer consubstanciado nº 2.001.481. As imagens correspondem a casos de Leucoplasia Bucal (LB) e Leucoplasia Verrucosa Proliferativa (LVP), condições classificadas como desordens potencialmente malignas da cavidade oral. As amostras foram previamente processadas por meio da definição de *Regions of Interest* (ROIs), contendo predominantemente núcleos celulares do epitélio. Esse recorte tem como objetivo concentrar a análise nas estruturas mais relevantes para a caracterização morfológica das lesões. As anotações de referência foram realizadas manualmente por um especialista, um estudante de mestrado, e posteriormente validadas por um patologista treinado, assegurando a qualidade do padrão-ouro utilizado no treinamento e na avaliação dos modelos. Embora o *dataset* contenha diferentes classes patológicas, neste estudo ele foi utilizado exclusivamente para a tarefa de segmentação semântica binária, considerando apenas a distinção entre núcleos celulares e região de fundo. Essa simplificação permite focar na capacidade dos modelos de delinear estruturas celulares, independentemente da classificação clínica das lesões.

Adicionalmente, foi utilizado o *dataset* público CryoNuSeg (MAHBOD et al., 2021),

representado na Figura 13 (B), amplamente empregado na literatura para avaliação de métodos de segmentação de núcleos celulares. Esse conjunto de dados é composto por 30 imagens histopatológicas provenientes de seções congeladas de diferentes tecidos humanos, também coradas com H&E. Esse *dataset* disponibiliza máscaras de referência precisas para os núcleos celulares, sendo frequentemente utilizado como *benchmark* em tarefas de segmentação nuclear. Uma característica relevante dessa base é a elevada diversidade morfológica, decorrente da variedade de tecidos e condições de aquisição, o que a torna particularmente desafiadora para modelos de aprendizado profundo.

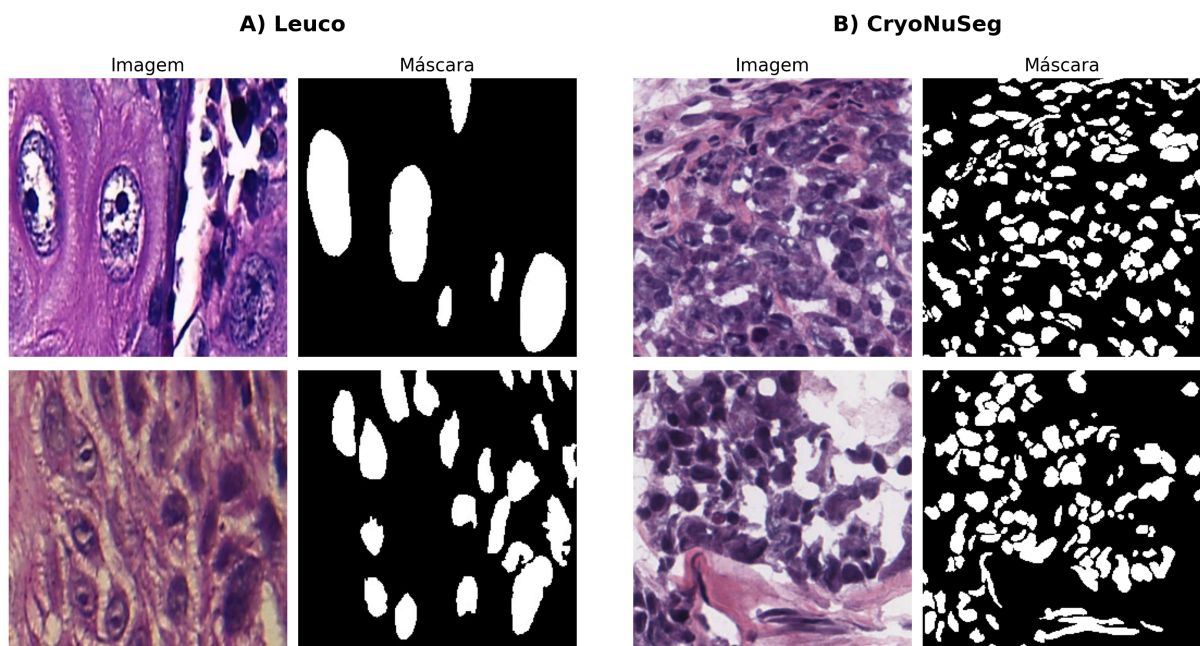


Figura 13 – Representação dos *datasets*: Leuco (A) e CryoNuSeg (B)

4.2 Arquiteturas de redes para segmentação

Como o problema de segmentação das estruturas histológicas envolve variabilidade morfológica, padrões multiescala e dependências espaciais complexas entre estruturas celulares, neste trabalho foram investigadas dez arquiteturas de aprendizagem profunda, que podem ser organizadas em três grandes grupos quanto ao tipo de representação adotado. O primeiro grupo reúne arquiteturas convolucionais e compreende duas variantes de U-Net (a *Baseline* e uma versão com atenção *Spatial and Channel Squeeze-and-Excitation* (scSE)), três variantes de UNet++ (diferenciadas pelos *encoders* EfficientNet-B6, ResNeSt50 e ConvNeXt-Base) e uma *Feature Pyramid Network* (FPN). O segundo grupo contempla abordagens híbridas, nas quais estruturas de segmentação clássicas são combinadas a codificadores baseados em *Transformer*, representadas pela rede MANet e pela TransUNet customizada. O terceiro grupo é formado por arquiteturas mais fortemente orientadas por atenção global, como o SegFormer e um modelo customizado que

integra um *backbone* MaxViT a um decodificador de atenção deformável. Dessa forma, o conjunto abrange diferentes paradigmas arquiteturais, desde abordagens clássicas do tipo *encoder-decoder* até modelos baseados em pirâmides multiescala, em fusão por atenção no decodificador e em atenção global.

A definição desse conjunto de modelos foi orientada por critérios estabelecidos previamente, considerando que, em uma segunda etapa, as predições individuais seriam combinadas por meio de estratégias de *ensemble learning*. Nesse sentido, a seleção das arquiteturas buscou: (i) maximizar a diversidade de representação, contemplando modelos convolucionais, híbridos e baseados em atenção, uma vez que vieses indutivos distintos são fatores determinantes para o ganho proporcionado por *ensembles*; (ii) cobrir diferentes mecanismos de fusão de características no decodificador, como conexões de atalho diretas, caminhos densamente aninhados, fusão piramidal e refinamento por atenção; (iii) incluir arquiteturas amplamente validadas na literatura de segmentação biomédica, garantindo desempenho individual competitivo; (iv) privilegiar modelos com codificadores pré-treinados disponíveis, favorecendo a transferência de aprendizado em um cenário de dados limitados. Dessa forma, cada arquitetura foi escolhida tanto por seu mérito individual na tarefa de segmentação de núcleos quanto por sua potencial contribuição, enquanto componente diverso, para as estratégias de *ensembles* investigadas posteriormente.

A Tabela 1 sintetiza as arquiteturas investigadas, destacando os *encoders* utilizados e a categorização de cada modelo quanto ao tipo de representação. Essa síntese permite investigar, em um mesmo protocolo experimental, como diferentes mecanismos de modelagem local e global influenciam a segmentação de núcleos em tecidos histológicos.

Tabela 1 – Síntese das arquiteturas de segmentação avaliadas, incluindo tipo de modelo, *encoder* utilizado e estratégia de modelagem.

Modelo	Arquitetura	Encoder	Categoria
Baseline U-Net	U-Net	ResNet34	CNN
U-Net + scSE	U-Net	ResNet34	CNN
UNet++ EfficientNet	UNet++	EfficientNet-B6	CNN
UNet++ ResNeSt	UNet++	ResNeSt50d	CNN
UNet++ ConvNeXt	UNet++	ConvNeXt-Base	CNN
FPN ResNet50	FPN	ResNet50	CNN
MAnet MiT-B2	MAnet	MiT-B2 (Transformer)	Baseada em Transformer
TransUNet	Encoder-Decoder custom	ResNet50 + Transformer	Híbrida
SegFormer	SegFormer	MiT-B3 (Transformer)	Baseada em Transformer
MaxViT + DLKA	Custom (MaxViT + DLKA)	MaxViT-Tiny	Híbrida

As subseções a seguir descrevem em detalhe cada uma dessas arquiteturas, organizadas pelas famílias correspondentes.

4.2.1 Arquiteturas baseadas em U-Net

Nesta seção, são descritas as variantes baseadas na arquitetura U-Net investigadas: a *Baseline U-Net* e uma versão acrescida de atenção scSE no decodificador. A U-Net

(RONNEBERGER; FISCHER; BROX, 2015) é uma das arquiteturas mais consolidadas na segmentação de imagens biomédicas, sendo composta por um caminho de contração (*encoder*) e um caminho de expansão (*decoder*), interligados por conexões de atalho (*skip connections*), que pode ser visualizada na Figura 14. No *encoder*, a resolução espacial da imagem é reduzida progressivamente, enquanto a profundidade dos mapas de características aumenta, permitindo a extração de descritores semânticos cada vez mais abstratos. No *decoder*, os mapas são sucessivamente ampliados e combinados com as características preservadas nas conexões laterais, o que favorece a recuperação de detalhes espaciais finos durante a reconstrução da máscara segmentada (RONNEBERGER; FISCHER; BROX, 2015). Essa estrutura é especialmente adequada para segmentação histológica, pois combina contexto global com preservação de contornos locais (RONNEBERGER; FISCHER; BROX, 2015; GRAHAM et al., 2019).

Neste estudo, a primeira arquitetura investigada foi a *Baseline U-Net*, implementada com *encoder* ResNet34 (HE et al., 2016) pré-treinado em ImageNet. Nessa configuração, a ResNet34 substitui o *encoder* convolucional original da U-Net e passa a atuar como extratora de características em múltiplos níveis hierárquicos. O decodificador mantém a lógica típica da U-Net, realizando *upsampling* progressivo e fusão com os mapas intermediários oriundos do *encoder*. Essa arquitetura foi adotada como referência experimental por representar uma solução robusta, amplamente utilizada e de menor complexidade relativa em comparação às variantes mais recentes (RONNEBERGER; FISCHER; BROX, 2015; HE et al., 2016).

A segunda configuração avaliada correspondeu à arquitetura U-Net com *encoder* ResNet34, porém acrescida de atenção do tipo scSE (ROY; NAVAB; WACHINGER, 2018) no decodificador. Esse mecanismo atua recalibrando as ativações do decodificador sob duas perspectivas complementares: ao longo dos canais, enfatizando filtros mais relevantes, e ao longo do espaço, reforçando regiões mais informativas da imagem. Essa variação busca aumentar a seletividade do decodificador na recuperação das regiões nucleares, reduzindo ativações espúrias em áreas de fundo com textura semelhante ao tecido de interesse.

4.2.2 Arquiteturas UNet++

Nesta seção, são apresentadas as três variantes baseadas em UNet++ investigadas neste trabalho, diferenciadas pelo *encoder* adotado e pela eventual incorporação de mecanismos de atenção no decodificador. A arquitetura UNet++ (ZHOU et al., 2018) pode ser entendida como uma generalização da U-Net clássica, na qual as conexões entre *encoder* e *decoder* deixam de ser simples atalhos diretos e passam a ser mediadas por caminhos densamente conectados, ilustrados na Figura 15. Essa reformulação reduz a lacuna semântica entre os mapas de características de baixa e alta resolução, favorecendo a fusão de informações de diferentes escalas de forma mais gradual e refinada. Em contextos histopatológicos, essa característica é particularmente útil, pois estruturas celulares podem

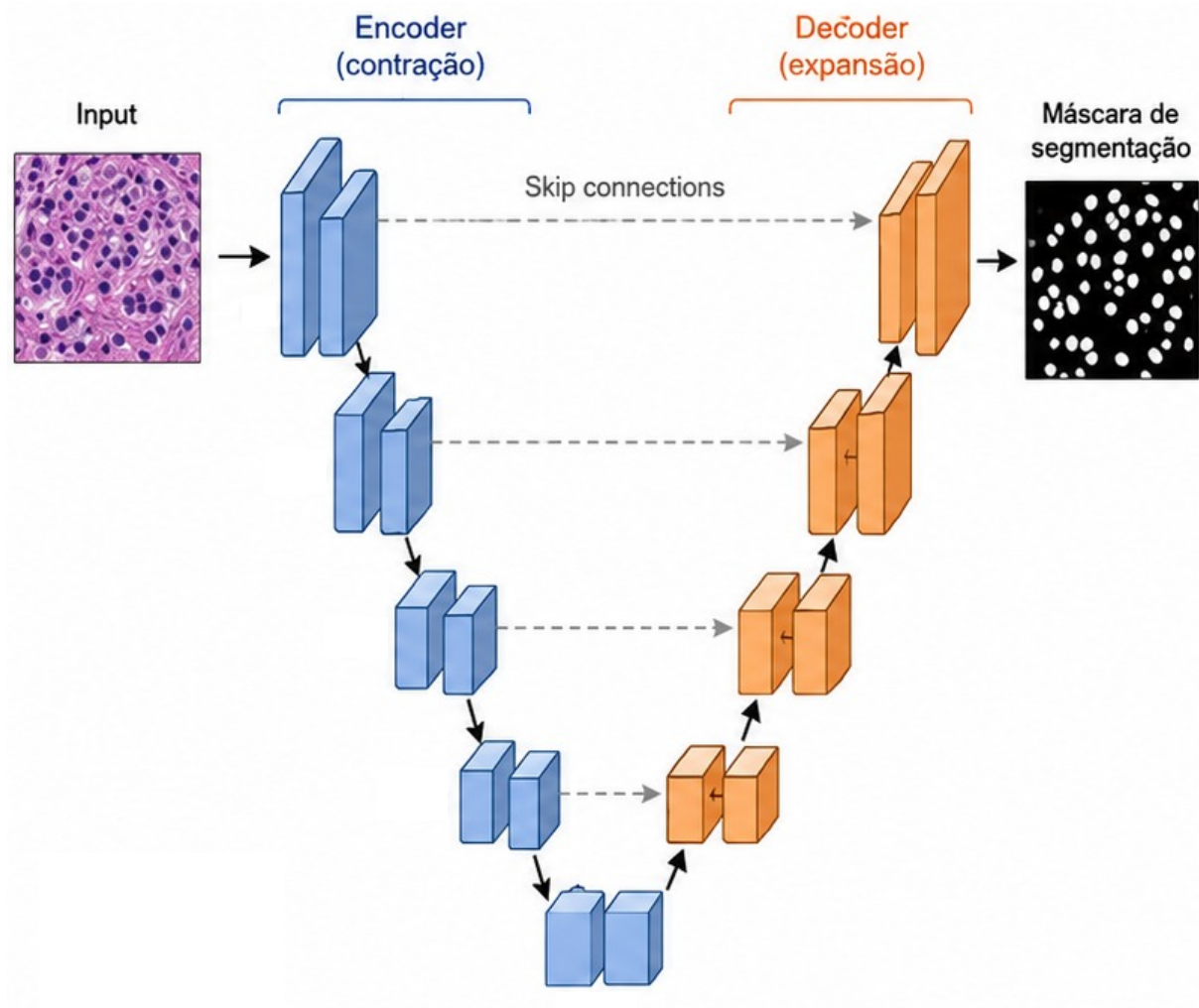


Figura 14 – Representação da arquitetura U-Net, composta por um caminho de contração (*encoder*) e um caminho de expansão (*decoder*), interligados por conexões de atalho que permitem combinar informações semânticas profundas com detalhes espaciais de maior resolução. Fonte: elaborado pelo autor com base em (RONNEBERGER; FISCHER; BROX, 2015).

apresentar grande variabilidade morfológica e exigir integração multiescala mais sofisticada do que a oferecida pela U-Net convencional (ZHOU et al., 2018; GRAHAM et al., 2019).

A primeira variante de UNet++ empregada neste estudo utilizou *encoder* EfficientNet-B6 (TAN; LE, 2019) pré-treinado em ImageNet. O EfficientNet-B6 foi incluído por representar uma família de redes convolucionais escalonadas de forma composta, equilibrando profundidade, largura e resolução. Ao ser integrado ao UNet++, ele fornece um *encoder* com elevada capacidade de extração de características e boa eficiência paramétrica, enquanto o decodificador aninhado do UNet++ promove fusão refinada entre níveis de resolução distintos.

A segunda variante de UNet++ utilizou o *encoder* ResNeSt (ZHANG et al., 2020), também com pesos ImageNet, em conjunto com atenção scSE (ROY; NAVAB; WACHIN-

GER, 2018) no decodificador. O ResNeSt incorpora mecanismos de *split-attention*, permitindo que o *encoder* modele relações mais ricas entre grupos de canais ainda na etapa de extração de características. Quando combinado com a topologia aninhada do UNet++ e com a recalibração espacial e por canal do scSE, esse modelo passa a explorar atenção tanto no *encoder* quanto no *decoder*, tornando-se uma alternativa mais expressiva para representação de padrões morfológicos complexos.

A terceira configuração dessa família foi a UNet++ com *encoder* ConvNext (LIU et al., 2022), pré-treinado em ImageNet. Essa arquitetura foi inspirada em vários princípios populares em *Transformers*, mas preserva a natureza convolucional do processamento. Quando acoplada ao decodificador do UNet++, ela forma um modelo que combina robustez convolucional, melhor capacidade representacional e fusão multiescala progressiva.

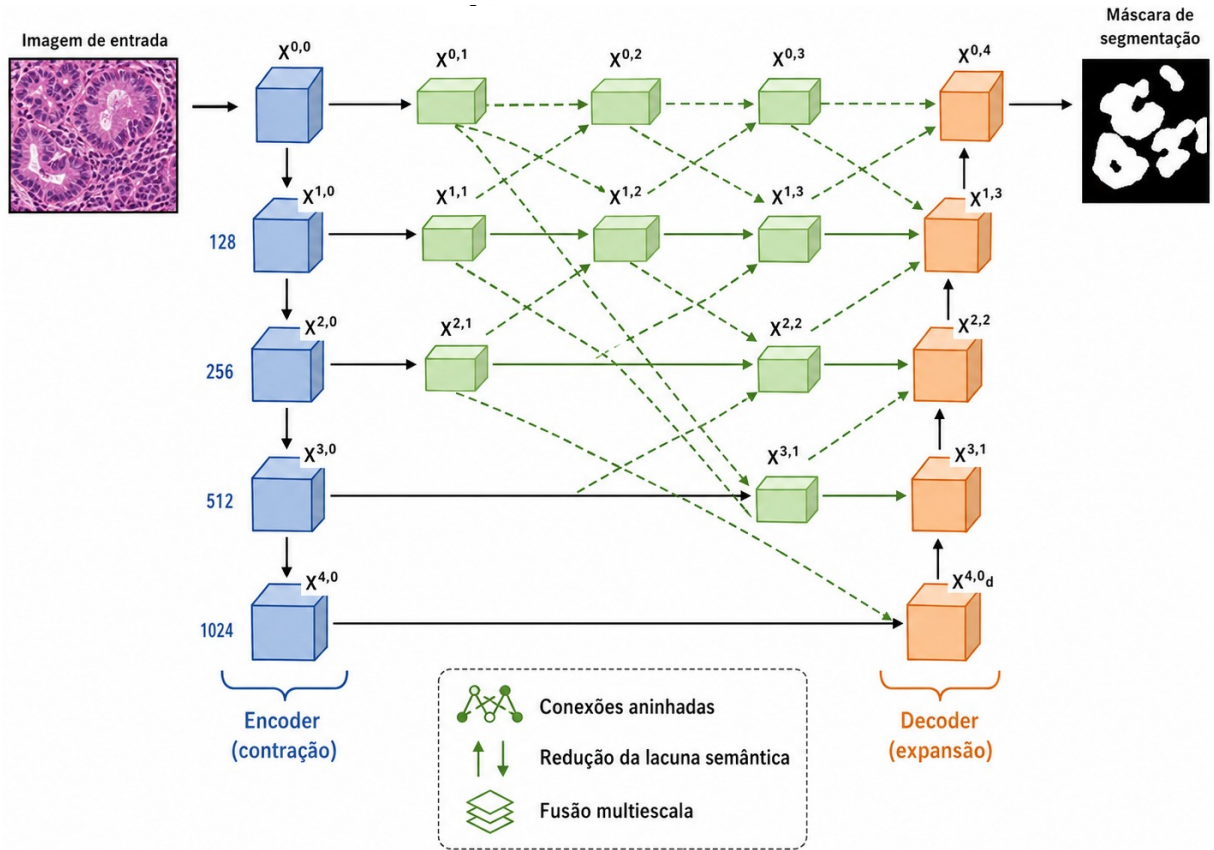


Figura 15 – Representação da arquitetura UNet++, caracterizada por conexões de atalho densamente aninhadas entre o *encoder* e o *decoder*, permitindo uma fusão gradual de características em diferentes níveis de resolução e reduzindo a lacuna semântica entre os mapas extraídos e reconstruídos. Fonte: elaborado pelo autor com base em (ZHOU et al., 2018).

4.2.3 Feature Pyramid Network

A *Feature Pyramid Network* (FPN) (LIN et al., 2017) constitui uma estratégia de segmentação baseada na construção explícita de uma pirâmide de características. Em

vez de depender exclusivamente de um único fluxo de expansão, a FPN utiliza um caminho *bottom-up*, derivado do *encoder*, e um caminho *top-down*, no qual características profundas e semanticamente ricas são projetadas novamente para níveis de maior resolução. Conexões laterais integram os dois fluxos, produzindo mapas multiescala úteis para detectar estruturas de diferentes tamanhos. No presente estudo, a FPN foi implementada com *encoder* ResNet50 (HE et al., 2016) pré-treinado em ImageNet. Essa configuração foi incluída para investigar o efeito de uma fusão piramidal explícita na segmentação de núcleos, em contraste com os modelos do tipo U-Net. Na Figura 16, é ilustrado o funcionamento desta arquitetura.

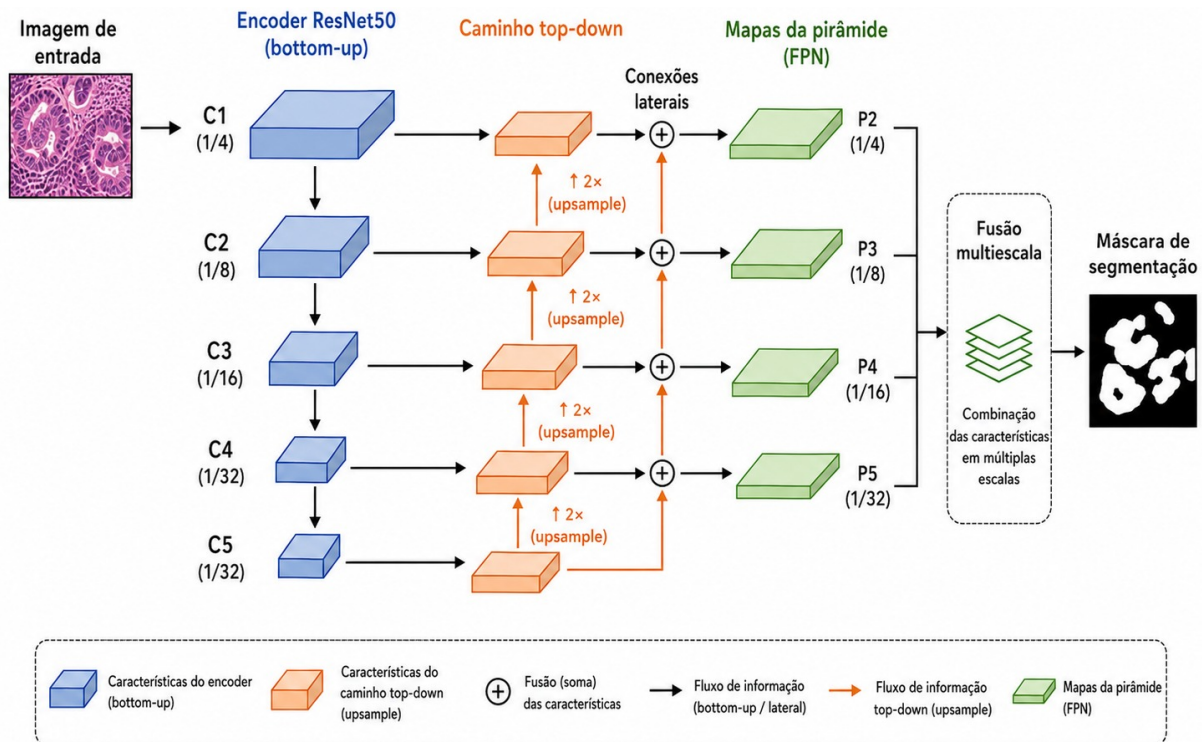


Figura 16 – Representação da arquitetura *Feature Pyramid Network* (FPN), baseada na combinação de um caminho *bottom-up*, um caminho *top-down* e conexões laterais para construção de mapas de características multiescala, úteis para segmentar estruturas de diferentes tamanhos. Fonte: elaborado pelo autor com base em (LIN et al., 2017).

4.2.4 MAnet

Foi avaliada a arquitetura MAnet (FAN et al., 2020) com *encoder* *Mix Transformer* (MiT)-B2, ilustrada na Figura 17. O MAnet combina a lógica de recuperação espacial típica de arquiteturas *encoder-decoder* com mecanismos de atenção voltados à agregação multiescala. O *encoder* escolhido para esse modelo foi o MiT-B2, isto é, um *Mix Transformer* (XIE et al., 2021) pré-treinado em ImageNet. Diferentemente de *encoders* puramente convolucionais, o MiT-B2 opera por estágios hierárquicos e tende a capturar dependências contextuais mais amplas. Assim, a combinação entre MAnet e MiT-B2 constitui

uma abordagem híbrida, em que um decodificador voltado à segmentação é alimentado por representações hierárquicas baseadas em *Transformer*.

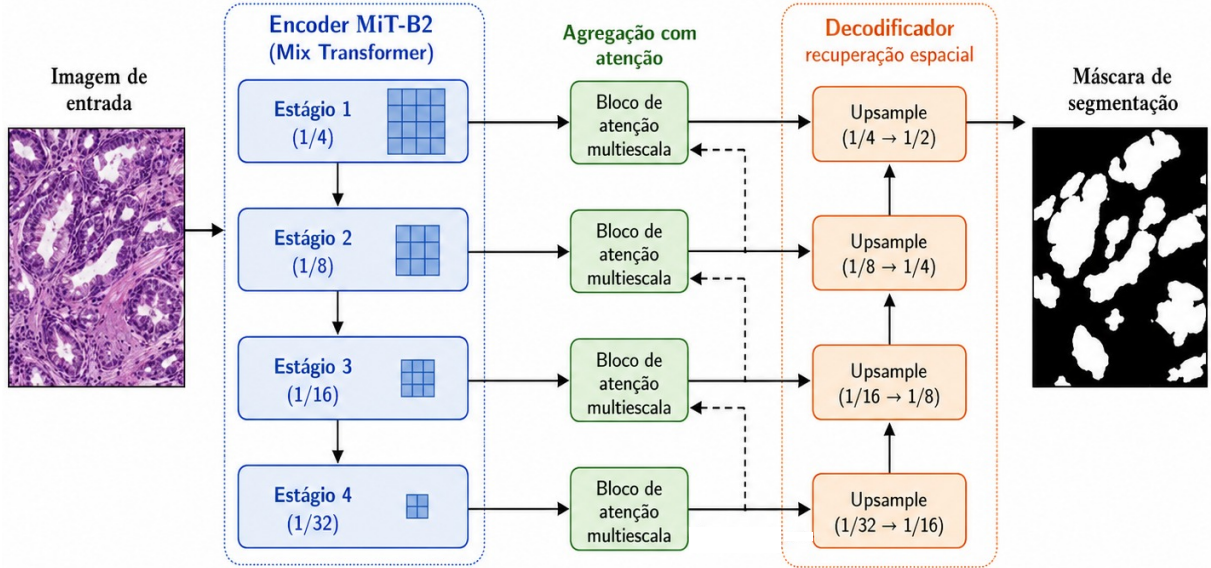


Figura 17 – Representação da arquitetura MANet, que combina uma estrutura do tipo *encoder-decoder* com mecanismos de atenção voltados à agregação multiescala de características. Neste trabalho, a arquitetura foi empregada com *encoder* MiT-B2, baseado em *Transformer*. Fonte: elaborado pelo autor com base em (FAN et al., 2020; XIE et al., 2021).

4.2.5 SegFormer

O SegFormer (XIE et al., 2021) representa uma arquitetura de segmentação orientada por *Transformer*, projetada para combinar representação hierárquica eficiente e decodificação simples, podendo ser visualizada em Figura 18. Diferentemente das arquiteturas baseadas em U-Net, o SegFormer não depende de um decodificador simétrico com longas conexões de atalho em todos os níveis. Em vez disso, ele utiliza um *encoder* hierárquico do tipo MiT para gerar descritores em múltiplas escalas e um esquema de agregação mais enxuto para produzir a segmentação final.

Essa arquitetura foi incluída por representar uma abordagem mais fortemente baseada em atenção global. Em imagens histopatológicas, esse tipo de modelagem pode ser vantajoso quando a segmentação depende não apenas de textura local, mas também do contexto morfológico mais amplo em que os núcleos estão inseridos. Por outro lado, como o decodificador é mais simples do que em arquiteturas do tipo U-Net++, o comportamento do SegFormer também permite analisar o equilíbrio entre sofisticação no *encoder* e simplicidade na reconstrução espacial (XIE et al., 2021).

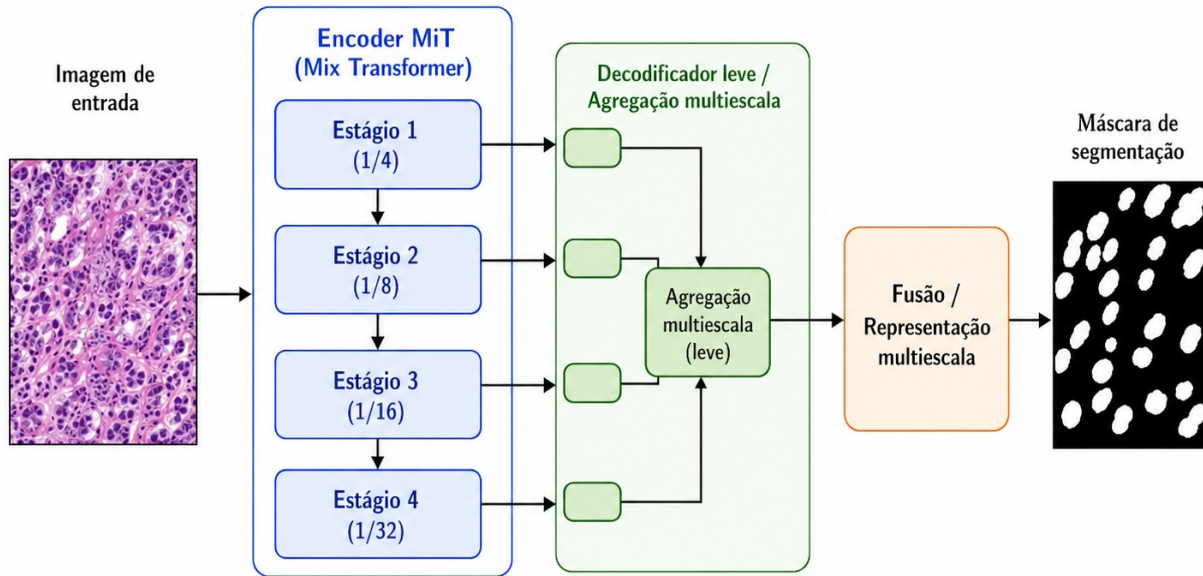


Figura 18 – Representação da arquitetura SegFormer, composta por um *encoder* hierárquico baseado em *Mix Transformer* (MiT) e um decodificador leve para agregação de características em múltiplas escalas e geração da máscara de segmentação. Fonte: elaborado pelo autor com base em (XIE et al., 2021).

4.2.6 TransUNet híbrida

A TransUNet (CHEN et al., 2021) foi incluída como representante das arquiteturas híbridas, que combinam codificação convolucional e atenção do tipo *Transformer* em um único modelo. Sua organização preserva o formato geral da U-Net, um caminho de codificação, um caminho de decodificação e conexões de atalho, mas insere um estágio baseado em autoatenção entre o codificador e o decodificador, ilustrada na Figura 19. Um codificador convolucional extrai mapas de características hierárquicos, preservando informação local de textura e bordas. A representação mais profunda é então convertida em uma sequência de *tokens*, enriquecida com codificação posicional e processada por blocos *Transformer*, capazes de modelar relações de longo alcance entre regiões distantes da imagem. Por fim, a saída retorna ao formato espacial e é refinada por um decodificador que, por meio das conexões de atalho, reaproveita os detalhes finos capturados nas camadas convolucionais iniciais.

A motivação para incluir essa arquitetura está em sua natureza híbrida. Modelos puramente convolucionais possuem campo receptivo predominantemente local e capturam bem contornos e texturas, mas têm dificuldade em representar dependências espaciais de longo alcance. Em contrapartida, abordagens baseadas exclusivamente em atenção modelam o contexto global, porém podem ser menos eficientes na recuperação de detalhes finos. Ao integrar os dois paradigmas, a TransUNet trata simultaneamente o detalhe local e o arranjo global das estruturas. Em histopatologia, essa combinação é particularmente relevante, pois a delimitação dos núcleos depende tanto de padrões locais de textura e coloração quanto da organização espacial das células no tecido. Adicionalmente, por

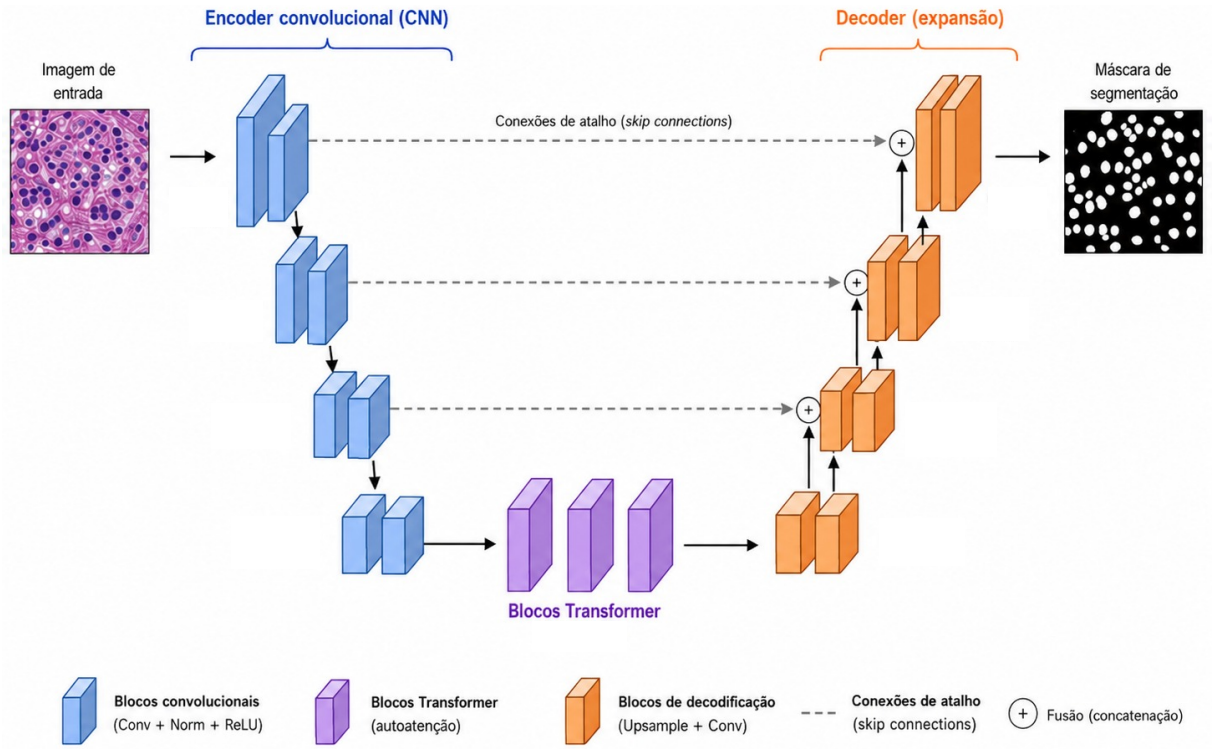


Figura 19 – Representação da arquitetura TransUNet, que combina um codificador convolucional, responsável pela extração de características locais, com blocos *Transformer*, responsáveis pela modelagem de dependências de longo alcance, seguidos por um decodificador responsável pela reconstrução espacial da máscara. Fonte: elaborado pelo autor com base em (CHEN et al., 2021).

adotar um viés indutivo distinto dos demais modelos, a TransUNet amplia a diversidade do conjunto, característica desejável para a posterior combinação por *ensemble*.

4.2.7 MaxViT com decodificador Deformable LKA

A décima arquitetura investigada foi um modelo que associa um *backbone* MaxViT (TU et al., 2022) a um decodificador dotado de atenção *Deformable Large Kernel Attention* (DLKA) (AZAD et al., 2024), representado na Figura 20. O *backbone* MaxViT atua como codificador e gera representações em múltiplas escalas, combinando convoluções e mecanismos de atenção que capturam de forma eficiente tanto o contexto local quanto o global da imagem. Essas representações alimentam um decodificador no estilo U-Net, em que cada estágio amplia progressivamente a resolução, integra as características provenientes do codificador por meio de conexões de atalho e, em seguida, refina o resultado com um módulo de atenção.

Após a fusão entre as características, camadas profundas e rasas, cada bloco aplica o módulo Deformable LKA (AZAD et al., 2024), responsável por um refinamento adaptativo orientado por atenção espacial. Conceitualmente, esse módulo combina duas ideias complementares. A primeira é o uso de convoluções de grande *kernel*, que ampliam o campo receptivo e incorporam contexto amplo a baixo custo computacional. A segunda

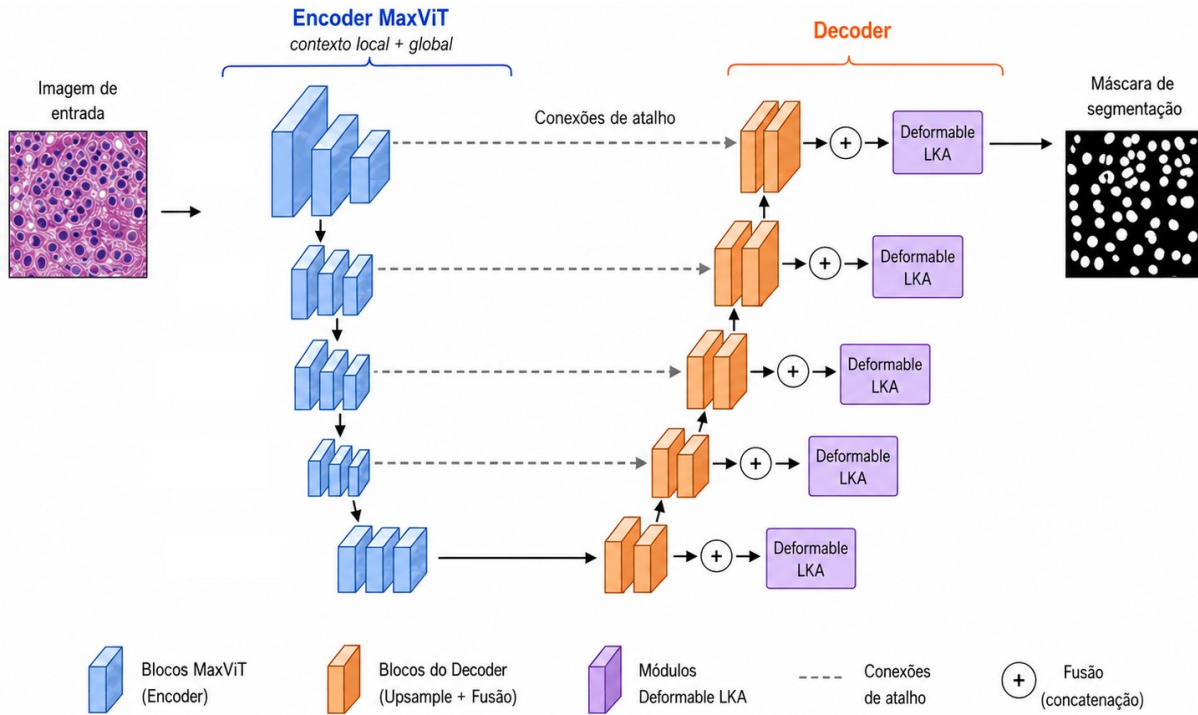


Figura 20 – Representação da arquitetura MaxViT com decodificador Deformable LKA, na qual o *backbone* MaxViT atua como codificador para extração de características locais e globais, enquanto o decodificador com atenção *Deformable Large Kernel Attention* refina as representações durante a reconstrução da máscara segmentada. Fonte: elaborado pelo autor com base em (TU et al., 2022; AZAD et al., 2024).

é a amostragem deformável, que permite ao campo receptivo ajustar-se dinamicamente ao conteúdo da imagem, deslocando as posições de amostragem conforme a forma das estruturas. Dessa maneira, o modelo aproxima o comportamento da autoatenção com maior eficiência do que mecanismos de atenção densa.

A motivação para incluir essa arquitetura foi avaliar uma estratégia em que o tratamento do contexto global não fica a cargo apenas do *backbone*, mas também do próprio decodificador. Em imagens histopatológicas, isso pode ser vantajoso porque a definição precisa da máscara nuclear depende, ao mesmo tempo, de contornos finos e da capacidade de adaptar o foco do modelo a padrões irregulares, sobreposição celular e variações locais de textura e coloração. Soma-se a isso a sua contribuição para a diversidade do conjunto, já que o refinamento por atenção deformável difere substancialmente dos mecanismos empregados pelas demais arquiteturas avaliadas.

4.3 Aumento de Dados

O aumento de dados é uma estratégia amplamente empregada em tarefas de visão computacional com o objetivo de ampliar artificialmente a variabilidade dos dados de treinamento, contribuindo para a redução de *overfitting* e melhoria da capacidade de

generalização dos modelos. Em imagens histopatológicas, essa etapa torna-se particularmente relevante devido à elevada variabilidade morfológica dos tecidos, bem como às variações introduzidas durante o processo de aquisição, como diferenças de coloração, orientação e preparo das lâminas.

Neste trabalho, foram investigadas duas estratégias de aumento de dados: transformações geométricas clássicas (denominadas *Geometric*) e o método *RCAug* (SANTOS et al., 2020). Essas abordagens foram selecionadas por apresentarem baixo custo computacional e por sua capacidade de simular variações realistas frequentemente observadas em imagens histológicas.

A estratégia *Geometric* consiste na aplicação de transformações espaciais que não alteram o conteúdo semântico da imagem, mas introduzem variações na sua representação. Foram utilizadas operações como rotações aleatórias, espelhamentos horizontais e verticais, translações e escalonamento. Essas transformações simulam mudanças de orientação e posicionamento das amostras, comuns durante o preparo e digitalização das lâminas histológicas, contribuindo para tornar os modelos mais invariantes a esse tipo de variação.

Por sua vez, a estratégia *Random Composition Augmentation* (RCAug) (SANTOS et al., 2023) produz diferentes versões de uma imagem por meio da combinação estocástica de transformações espaciais, distorções, alterações de aparência e oclusões. Inicialmente, podem ser aplicados espelhamentos horizontal e vertical e rotações em múltiplos de 90° , cada um com probabilidade de 0,5. Em seguida, com probabilidade de 0,7, é aplicada uma transformação conjunta de translação, escala e rotação, permitindo deslocamentos de até 10% das dimensões da imagem, fatores de escala entre 0,8 e 1,2 e rotações no intervalo de -180° a 180° . Com probabilidade de 0,4, uma operação é selecionada aleatoriamente entre distorção de grade, deformação elástica e distorção óptica, introduzindo alterações não rígidas na disposição das estruturas. Adicionalmente, com probabilidade de 0,5, é escolhida uma transformação de aparência entre variação de matiz, saturação e intensidade, modificação aleatória de brilho e contraste ou equalização adaptativa de histograma por CLAHE. Por fim, o *coarse dropout* é aplicado com probabilidade de 0,3, ocultando até oito regiões de no máximo 20×20 pixels. Como essas operações são amostradas durante o carregamento das imagens, uma mesma amostra pode receber diferentes combinações ao longo das épocas. Em comparação ao cenário *Geometric*, o RCAug explora não apenas mudanças de orientação, posição e escala, mas também deformações locais, variações cromáticas e oclusões, ampliando a diversidade estrutural e visual apresentada aos modelos durante o treinamento.

Em ambos os casos, foi adotado um fator de aumento fixo de $3\times$ por imagem original, resultando em um conjunto de treinamento expandido sem alteração do conteúdo semântico das imagens. Essa configuração foi escolhida com o objetivo de equilibrar o ganho de variabilidade com o custo computacional do treinamento (SHORTEN; KHOSHGOF-

TAAR, 2019; SANTOS et al., 2023).

Adicionalmente, os experimentos foram conduzidos considerando diferentes cenários: (i) sem aplicação de aumento de dados, (ii) utilizando apenas transformações geométricas e (iii) utilizando o método *RCAug*. Essa configuração permite avaliar de forma sistemática o impacto de cada estratégia no desempenho dos modelos, bem como sua contribuição para a robustez das segmentações.

4.4 Estratégias de *Ensemble Learning*

Após análise individual dos modelos, este trabalho investigou o uso de técnicas de *ensemble learning* com o objetivo de explorar a complementaridade entre as arquiteturas. Em problemas de segmentação semântica, diferentes arquiteturas tendem a apresentar comportamentos complementares, alguns modelos são mais sensíveis à identificação das regiões de interesse, enquanto outros são mais conservadores e produzem menos falsos positivos. Nesse contexto, a combinação das saídas de múltiplos modelos pode resultar em segmentações mais robustas do que aquelas produzidas isoladamente.

Neste estudo, cada modelo gera, para uma mesma imagem de entrada, um mapa de probabilidade $P_i \in [0, 1]^{H \times W}$, em que cada posição indica a probabilidade de um *pixel* pertencer à classe de interesse. Como a tarefa considerada é binária, essa classe corresponde à região nuclear, enquanto os demais *pixels* são tratados como fundo. A partir desse conjunto de predições, foram avaliadas cinco estratégias de combinação: *Soft Voting*, *Hard Voting*, *Weighted Soft Voting*, *Union Max* e *Intersection Min*. Em todos os casos, após a etapa de fusão das predições, os mapas de probabilidade resultantes foram convertidos em máscaras binárias por meio de uma regra de decisão *pixel a pixel* com limiar fixo de 0,5. Essa escolha é consistente com a prática usual em tarefas de segmentação binária, nas quais a saída probabilística do modelo é binarizada para distinguir a classe de interesse do fundo (SHIROKIKH et al., 2021; SILVA; NARAYANAN; HARDIE, 2022). Além disso, em abordagens de *ensemble* baseadas na combinação de escores ou mapas de probabilidade, trabalhos prévios também empregam a média das predições seguida da aplicação de um limiar de 0,5 para obtenção da máscara final (SILVA; NARAYANAN; HARDIE, 2022). Neste estudo, o mesmo limiar foi mantido para todas as estratégias de fusão, sem ajuste posterior, a fim de garantir uma comparação padronizada entre os métodos avaliados.

Sejam n modelos participantes do *ensemble* e $\{P_1, P_2, \dots, P_n\}$ seus respectivos mapas de probabilidade. A seguir, são formalmente descritas as estratégias de fusão utilizadas para combinar os mapas de probabilidade gerados pelos modelos.

4.4.1 *Soft Voting*

Na estratégia de *Soft Voting*, as saídas probabilísticas dos modelos são combinadas por meio da média aritmética *pixel a pixel*:

$$P_{soft} = \frac{1}{n} \sum_{i=1}^n P_i. \quad (1)$$

Após essa etapa, a máscara final é obtida pela limiarização do mapa médio:

$$\hat{P}_{soft} = \mathbb{1}(P_{soft} > 0,5), \quad (2)$$

em que $\mathbb{1}(\cdot)$ representa a função indicadora, que assume valor 1 quando a condição é satisfeita e 0 caso contrário.

A principal característica dessa abordagem é a preservação da informação probabilística de cada modelo. Em vez de considerar apenas decisões binárias, o *Soft Voting* considera o grau de confiança atribuído pelos modelos a cada *pixel*. Dessa forma, predições com alta concordância probabilística tendem a ser reforçadas, enquanto regiões de incerteza são suavizadas. Em problemas de segmentação histológica, essa característica é particularmente relevante, pois os contornos nucleares podem ser pouco definidos ou apresentar ambiguidades locais de textura e coloração (ZHAO et al., 2020; SHI et al., 2022).

4.4.2 *Hard Voting*

Na estratégia de *Hard Voting*, cada mapa de probabilidade é inicialmente convertido em uma máscara binária individual:

$$\hat{P}_i = \mathbb{1}(P_i > 0,5). \quad (3)$$

Em seguida, a decisão final é tomada por votação majoritária:

$$\hat{P}_{hard} = \mathbb{1}\left(\sum_{i=1}^n \hat{P}_i > \frac{n}{2}\right). \quad (4)$$

Nesse caso, um *pixel* é classificado como positivo apenas se a maioria dos modelos o tiver classificado como pertencente à classe de interesse. Diferentemente do *Soft Voting*, essa técnica desconsidera os valores probabilísticos contínuos e opera apenas com decisões discretas. Como consequência, o *Hard Voting* tende a ser menos sensível a pequenas variações de confiança entre modelos, mas também pode perder informação útil em situações em que uma minoria de modelos apresenta alta convicção correta sobre determinada região.

O *Hard Voting* é uma estratégia de consenso explícito. Sua vantagem está na simplicidade e na robustez frente a flutuações probabilísticas pequenas; sua limitação, por outro

lado, está em tratar igualmente decisões muito confiantes e decisões marginais logo acima do limiar.

4.4.3 *Weighted Soft Voting*

A estratégia de *Weighted Soft Voting* é uma extensão do *Soft Voting*, na qual a contribuição de cada modelo não é uniforme. Em vez de realizar uma média simples, utiliza-se uma média ponderada pelas métricas de desempenho individual dos modelos. No presente estudo, os pesos foram definidos a partir do coeficiente Dice obtido por cada modelo individual.

Formalmente, a combinação é dada por:

$$P_{soft} = \frac{\sum_{i=1}^n w_i P_i}{\sum_{i=1}^n w_i}, \quad (5)$$

em que w_i representa o peso associado ao modelo i . A máscara final é então obtida por:

$$\hat{P}_{soft} = \mathbb{1}(P_{soft} > 0,5). \quad (6)$$

Essa estratégia parte da hipótese de que modelos mais precisos individualmente devem exercer maior influência sobre a decisão final do *ensemble*. Assim, ao utilizar o Dice como peso, busca-se valorizar modelos que apresentam maior sobreposição média com o padrão-ouro. Essa abordagem procura equilibrar dois fatores: a preservação da informação probabilística e a priorização de modelos mais confiáveis. Em cenários nos quais há modelos com desempenho muito desigual, o *Weighted Soft Voting* tende a ser uma alternativa mais informativa do que a média simples.

4.4.4 *Union Max*

A estratégia *Union Max* combina os modelos selecionando, para cada *pixel*, o maior valor de probabilidade dentre todos os mapas,

$$P_{union} = \max_i(P_i), \quad (7)$$

e a máscara binária final é dada por:

$$\hat{P}_{union} = \mathbb{1}(P_{union} > 0,5). \quad (8)$$

Essa abordagem pode ser interpretada como uma operação de união entre as predições dos modelos. Basta que ao menos um deles atribua alta probabilidade a determinado *pixel* para que esse valor seja preservado no mapa combinado. Por esse motivo, o *Union Max* tende a favorecer a sensibilidade, isto é, aumenta a chance de recuperar *pixels* verdadeiramente pertencentes à classe positiva. Em contrapartida, também pode aumentar o

número de falsos positivos, já que predições indevidas de apenas um modelo são suficientes para influenciar o resultado final.

No contexto da segmentação nuclear, essa estratégia é útil quando se deseja priorizar a detecção de regiões potencialmente relevantes, mesmo ao custo de uma segmentação mais permissiva (KAMNITSAS et al., 2017).

4.4.5 *Intersection Min*

De forma complementar, a estratégia *Intersection Min* seleciona, para cada *pixel*, a menor probabilidade entre os modelos:

$$P_{inter} = \min_i(P_i). \quad (9)$$

A máscara final é então obtida por:

$$\hat{P}_{inter} = \mathbb{1}(P_{inter} > 0,5). \quad (10)$$

Essa técnica pode ser interpretada como uma operação de interseção entre as predições. Um *pixel* será mantido como positivo apenas se todos os modelos atribuírem a ele probabilidade suficientemente alta. Por isso, trata-se de uma abordagem mais restritiva, que tende a reduzir falsos positivos e, conseqüentemente, favorecer a especificidade. Em contrapartida, regiões nucleares mais difíceis ou ambíguas podem ser descartadas caso haja discordância entre os modelos.

Em aplicações práticas, o *Intersection Min* é particularmente interessante quando se deseja obter segmentações mais conservadoras, privilegiando a confiabilidade dos *pixels* positivos em vez da cobertura completa das estruturas.

4.5 Métricas de Avaliação

A avaliação dos resultados foi realizada em dois níveis complementares. No primeiro, foram utilizadas métricas clássicas de segmentação semântica, com o objetivo de quantificar a qualidade global das máscaras produzidas pelos modelos individuais e pelos ensembles. No segundo, foram definidas métricas específicas para analisar a contribuição de cada modelo dentro do ensemble, permitindo compreender como a combinação preserva acertos individuais, descarta decisões corretas ou corrige erros previamente cometidos.

As métricas tradicionais de segmentação foram calculadas a partir da comparação *pixel a pixel* entre a máscara predita e o padrão-ouro. Para isso, consideram-se quatro quantidades fundamentais:

- *True Positive* (TP): número de *pixels* corretamente classificados como pertencentes à classe positiva, isto é, *pixels* nucleares corretamente segmentados;

- *True Negative* (TN): número de *pixels* corretamente classificados como fundo;
- *False Positive* (FP): número de *pixels* incorretamente classificados como positivos, ou seja, regiões de fundo segmentadas indevidamente como núcleo;
- *False Negative* (FN): número de *pixels* da classe positiva que não foram identificados pelo modelo, isto é, regiões nucleares classificadas como fundo.

Esses quatro termos formam a base das métricas descritas a seguir. Em segmentação histopatológica, sua interpretação é direta: *TP* e *TN* representam acertos do modelo, enquanto *FP* e *FN* representam dois tipos distintos de erro, associados respectivamente ao excesso e à omissão de segmentação.

4.5.1 Coeficiente *Dice*

O coeficiente Dice, também conhecido como *Dice Similarity Coefficient* (DSC), mede o grau de sobreposição entre a máscara predita e a máscara de referência:

$$Dice = \frac{2TP}{2TP + FP + FN}. \quad (11)$$

Essa métrica varia entre 0 e 1, sendo que valores mais próximos de 1 indicam maior concordância entre predição e padrão-ouro. O Dice é particularmente relevante em tarefas de segmentação porque penaliza simultaneamente falsos positivos e falsos negativos, sendo bastante sensível à qualidade da sobreposição espacial. Em estruturas pequenas e irregulares, como núcleos celulares, essa característica torna a métrica especialmente informativa.

4.5.2 *Intersection over Union*

A métrica *Intersection over Union* (IoU), também chamada de índice de Jaccard, quantifica a razão entre a interseção e a união das regiões positivas da predição e da referência:

$$IoU = \frac{TP}{TP + FP + FN}. \quad (12)$$

Assim como o Dice, o IoU mede a qualidade da segmentação em termos de sobreposição. No entanto, ele é mais rigoroso, pois não possui o fator 2 no numerador. Como consequência, para um mesmo resultado de segmentação, os valores de IoU costumam ser numericamente menores que os de Dice. O uso conjunto dessas duas métricas permite uma visão mais robusta da qualidade espacial da máscara segmentada.

4.5.3 Acurácia

A acurácia mede a proporção total de *pixels* corretamente classificados:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \quad (13)$$

Essa métrica fornece uma visão global do desempenho do modelo, considerando simultaneamente as classes positiva e negativa. Contudo, em problemas com desbalanceamento entre classes, como ocorre frequentemente em segmentação histológica, a acurácia isoladamente pode ser enganosa. Isso ocorre porque uma imagem com grande predominância de fundo pode apresentar alta acurácia mesmo quando a segmentação dos núcleos é insatisfatória. Por esse motivo, nesta dissertação a acurácia é interpretada sempre em conjunto com Dice, IoU, sensibilidade e especificidade.

4.5.4 Sensibilidade

A sensibilidade, também chamada de *recall* ou taxa de verdadeiros positivos, mede a capacidade do modelo de identificar corretamente os pixels pertencentes à classe positiva:

$$Sens = \frac{TP}{TP + FN}. \quad (14)$$

Valores elevados de sensibilidade indicam que poucas regiões nucleares deixaram de ser detectadas. Em outras palavras, essa métrica quantifica o quanto o modelo consegue evitar falsos negativos. Em aplicações biomédicas, a sensibilidade é particularmente importante quando a omissão de estruturas relevantes é considerada um erro crítico.

4.5.5 Especificidade

A especificidade mede a capacidade do modelo de classificar corretamente os pixels de fundo:

$$Spec = \frac{TN}{TN + FP}. \quad (15)$$

Valores elevados de especificidade indicam que o modelo evita segmentar indevidamente regiões que não pertencem à classe de interesse. Em termos práticos, essa métrica quantifica o controle sobre falsos positivos. Em imagens histopatológicas, a especificidade é importante porque artefatos, variações de coloração ou texturas do tecido podem induzir o modelo a supersegmentar áreas não nucleares.

4.5.6 Métricas de contribuição individual no *ensemble*

Além das métricas tradicionais de segmentação, esta dissertação propõe e formaliza três métricas complementares para analisar a contribuição individual dos modelos no

ensemble: *kept rate*, *discarded rate* e *error mitigation*. Essas métricas são calculadas por meio da comparação entre a saída binária final do *ensemble*, as saídas binárias dos modelos individuais e o padrão-ouro, permitindo quantificar, respectivamente, a preservação dos acertos individuais, a perda desses acertos após a combinação e a correção, pelo *ensemble*, dos erros cometidos por cada modelo. Sua utilização complementa as métricas globais de segmentação ao tornar possível compreender como o ganho de desempenho emerge da interação entre as arquiteturas e verificar se a complementaridade de seus padrões de erro está sendo efetivamente aproveitada.

Considere um modelo individual M_i , sua predição binária $\hat{P}_i$, a predição binária do *ensemble* $\hat{P}_{ens}$ e o padrão-ouro Y . A partir dessas informações, métricas são apresentadas para essa análise.

4.5.7 Kept Rate

A métrica *Kept Rate* representa a proporção dos *pixels* corretamente classificados pelo modelo individual que permanecem corretos após a combinação no *ensemble*. Seja C_i o conjunto de *pixels* em que o modelo M_i está correto, x o pixel pertencente ao conjunto de avaliação. Então:

$$KeptRate = \frac{|\{x \in C_i : \hat{P}_{ens}(x) = Y(x)\}|}{|C_i|}. \quad (16)$$

Essa métrica quantifica o grau de preservação dos acertos individuais. Valores altos indicam que o *ensemble* mantém a maior parte das decisões corretas já realizadas pelo modelo analisado quando aplicado a operação de combinação no *ensemble*.

4.5.8 Discarded Rate

A métrica *Discarded Rate* mede a proporção dos *pixels* corretamente classificados por um modelo individual que deixaram de estar corretos após a combinação no *ensemble*. Seja novamente C_i o conjunto de acertos do modelo M_i :

$$DiscardedRate = \frac{|\{x \in C_i : \hat{P}_{ens}(x) \neq Y(x)\}|}{|C_i|}. \quad (17)$$

Essa métrica corresponde ao complemento da preservação dos acertos, isto é, ela evidencia o chamado “potencial perdido” pelo processo de combinação. Valores elevados de *Discarded Rate* indicam que o *ensemble* está revertendo um número relevante de decisões corretas do modelo individual. Isso significa que *pixels* que o modelo individual havia rotulado corretamente, seja como núcleo ou seja como fundo, passaram a ser rotulados de forma incorreta após a fusão, caracterizando a perda da informação útil que aquele modelo já possuía isoladamente.

4.5.9 *Error Mitigation*

A métrica *Error Mitigation* quantifica a capacidade do *ensemble* de corrigir erros cometidos por um modelo individual. Seja E_i o conjunto de *pixels* em que o modelo M_i está incorreto. Então:

$$ErrorMitigation = \frac{|\{x \in E_i : \hat{P}_{ens}(x) = Y(x)\}|}{|E_i|}. \quad (18)$$

Valores elevados de *Error Mitigation* indicam que a combinação foi capaz de compensar limitações individuais do modelo analisado, transformando erros em acertos. Essa métrica é particularmente útil para verificar se a diversidade entre modelos está sendo explorada de forma efetiva pelo *ensemble*.

Experimentos e Análise dos Resultados

Os experimentos realizados neste trabalho têm como objetivo avaliar o desempenho de diferentes arquiteturas de segmentação semântica, bem como investigar o impacto de estratégias de aumento de dados e técnicas de *ensemble learning* na qualidade das segmentações obtidas.

As avaliações foram conduzidas utilizando dois conjuntos de dados, o *dataset* de leucoplasia bucal (Leuco), empregado como domínio principal de estudo, e o *dataset* público CryoNuSeg, utilizado para análise de generalização dos modelos. Para cada arquitetura, foram considerados três cenários experimentais: (i) sem aumento de dados, (ii) com aumento baseado em transformações geométricas e (iii) com a estratégia RCAug.

Inicialmente, os modelos foram avaliados individualmente, sendo posteriormente combinados por meio de diferentes abordagens de *ensemble*. O desempenho foi mensurado por métricas tradicionais de segmentação, complementadas por métricas específicas que analisam a contribuição de cada modelo no processo de combinação.

Nas subseções a seguir, são apresentados e discutidos os resultados obtidos em cada etapa do *pipeline* experimental.

5.1 Resultados dos modelos individuais

Nesta seção, são apresentados os resultados obtidos pelas arquiteturas avaliadas de forma individual, considerando diferentes estratégias de aumento de dados. O objetivo desta análise é investigar o desempenho de cada modelo, bem como sua capacidade de generalização em um cenário com maior variabilidade morfológica.

Inicialmente, são discutidos os resultados no *dataset* de leucoplasia, seguido pela avaliação no *dataset* CryoNuSeg. Em seguida, é apresentada uma análise do impacto das estratégias de aumento de dados no desempenho dos modelos.

5.1.1 Desempenho no *dataset* de leucoplasia

A Tabela 2 apresenta os resultados obtidos pelas arquiteturas avaliadas no *dataset* de leucoplasia, considerando os três cenários experimentais.

Tabela 2 – Desempenho dos modelos no *dataset* de Leucoplasia, considerando diferentes estratégias de aumento de dados.

Modelo	D_C	A_C (%)	IoU	S_E (%)	E_S (%)
Baseline U-Net ResNet-50	0,8477	93,42	0,7356	85,46	95,59
Baseline U-Net ResNet-50 [†]	0,8721	94,46	0,7732	88,21	96,16
Baseline U-Net ResNet-50 [△]	0,8710	94,41	0,7715	88,14	96,12
Attention U-Net ResNet-50	0,8489	93,51	0,7375	85,14	95,79
Attention U-Net ResNet-50 [†]	0,8703	94,40	0,7704	87,80	96,20
Attention U-Net ResNet-50 [△]	0,8731	94,47	0,7747	88,73	96,04
Efficient-Unet++ EfficientNet-B6	0,8578	93,88	0,7510	86,19	95,97
Efficient-Unet++ EfficientNet-B6 [†]	0,8763	94,62	0,7799	89,04	96,14
Efficient-Unet++ EfficientNet-B6 [△]	0,8754	94,56	0,7784	89,21	96,02
ResNeSt-Unet++ ResNeSt-50d	0,8674	94,31	0,7658	86,85	96,35
ResNeSt-Unet++ ResNeSt-50d [†]	0,8839	95,04	0,7920	88,20	96,90
ResNeSt-Unet++ ResNeSt-50d [△]	0,8853	95,06	0,7942	88,97	96,72
MA-Net MiT-b2	0,8527	93,65	0,7433	85,80	95,79
MA-Net MiT-b2 [†]	0,8802	94,84	0,7861	88,62	96,53
MA-Net MiT-b2 [△]	0,8782	94,70	0,7828	89,26	96,18
ConvNeXt U-Net ConvNeXt Base	0,8676	94,28	0,7661	87,43	96,15
ConvNeXt U-Net ConvNeXt Base [†]	0,8791	94,75	0,7843	89,22	96,25
ConvNeXt U-Net ConvNeXt Base [△]	0,8823	94,87	0,7893	89,68	96,29
FPN ResNet-50	0,8304	92,73	0,7100	83,09	95,36
FPN ResNet-50 [†]	0,8630	94,07	0,7590	87,23	95,93
FPN ResNet-50 [△]	0,8589	93,84	0,7527	87,48	95,58
TransUNet ResNet-50 + ViT	0,8505	93,58	0,7399	85,27	95,84
TransUNet ResNet-50 + ViT [†]	0,8678	94,22	0,7665	88,53	95,78
TransUNet ResNet-50 + ViT [△]	0,8693	94,31	0,7688	88,31	95,95
SegFormer MiT-b3	0,8631	94,09	0,7592	87,00	96,02
SegFormer MiT-b3 [†]	0,8765	94,64	0,7802	88,80	96,23
SegFormer MiT-b3 [△]	0,8756	94,59	0,7787	88,92	96,13
Unet dlka MaxVit	0,8657	94,35	0,7663	88,58	96,01
Unet dlka MaxVit [†]	0,8715	94,62	0,7748	88,26	96,18
Unet dlka MaxVit [△]	0,8744	94,71	0,7793	89,08	96,17

[†] Modelos treinados com a estratégia RCAug.

[△] Modelos treinados com aumento geométrico.

As métricas D_C e IoU são apresentadas no intervalo [0,1], enquanto A_C , S_E e E_S são apresentadas em porcentagem. Os valores em negrito representam o melhor resultado dentro de cada arquitetura avaliada; as células destacadas em cinza indicam o melhor desempenho global em cada métrica.

De forma geral, as arquiteturas da família UNet++ figuraram entre as de melhor desempenho, com destaque para a ResNeSt-UNet++ e a ConvNeXt U-Net. No cenário sem aumento de dados, a ConvNeXt U-Net ($D_C = 0,8679$) e a ResNeSt-UNet++ ($D_C = 0,8674$) obtiveram os melhores resultados entre os modelos avaliados.

A aplicação de aumento de dados resultou em melhorias consistentes em todas as arquiteturas avaliadas. O D_C aumentou em todos os modelos, com ganhos que variaram de cerca de 0,009 (Unet dlka) a aproximadamente 0,033 (FPN) em relação ao cenário sem aumento. A título de exemplo, a Baseline U-Net apresentou ganho de cerca de 0,023 com aumento geométrico, enquanto a ResNeSt-UNet++ melhorou aproximadamente 0,018. Entre as estratégias, o aumento geométrico produziu o melhor desempenho global (ResNeSt-UNet++, $D_C = 0,8853$), seguido de perto pela estratégia RCAug (ResNeSt-UNet++, $D_C = 0,8839$). Esses resultados indicam que a combinação de arquiteturas profundas com estratégias de aumento de dados contribui de forma consistente para o desempenho dos modelos.

Além disso, observa-se que arquiteturas mais simples, como a FPN, apresentaram desempenho inferior em todos os cenários, sugerindo limitações na modelagem de estruturas complexas presentes nas imagens histopatológicas. Por outro lado, modelos baseados em diferentes paradigmas, como o SegFormer e o TransUNet, apresentaram desempenho próximo, embora ligeiramente inferior às melhores arquiteturas convolucionais.

Outro ponto relevante é o aumento, em geral consistente, da S_E com o uso de *data augmentation*, indicando maior capacidade dos modelos em identificar corretamente os núcleos celulares. Em parte das arquiteturas, observou-se também aumento da E_S . Esse padrão admite uma leitura complementar. Nos modelos de menor desempenho inicial, o aumento de dados tende a melhorar a segmentação de modo mais geral, enquanto, nos modelos mais robustos, o ganho concentra-se na detecção dos núcleos, que se reflete em contornos mais precisos. De modo geral, esse comportamento reforça o papel do aumento de dados na redução de falsos negativos, aspecto que pode ser relevante a depender do contexto clínico e do tipo de erro envolvido.

5.1.2 Desempenho no *dataset* CryoNuSeg

A Tabela 3 apresenta os resultados obtidos no *dataset* CryoNuSeg, utilizado para avaliar a capacidade de generalização dos modelos. De maneira geral, observa-se uma redução no desempenho em relação ao *dataset* de leucoplasia, o que pode ser atribuído à maior variabilidade morfológica presente no CryoNuSeg. No cenário sem aumento de dados, a Efficient-UNet++ apresentou o melhor desempenho ($D_C = 0,8343$), seguida pela ResNeSt-UNet++ ($D_C = 0,8312$), indicando maior robustez dessas arquiteturas frente a mudanças de domínio.

Diferente do observado no *dataset* de leucoplasia, o impacto do aumento de dados foi menos consistente. A estratégia RCAug apresentou ganhos moderados em algumas

Tabela 3 – Desempenho dos modelos no *dataset* CryoNuSeg, considerando diferentes estratégias de aumento de dados.

Modelo	D_C	A_C (%)	IoU	S_E (%)	E_S (%)
Baseline U-Net ResNet-50	0,8200	87,27	0,6949	81,08	90,71
Baseline U-Net ResNet-50 [†]	0,8090	86,79	0,6792	78,27	91,53
Baseline U-Net ResNet-50 [△]	0,8303	87,39	0,7098	86,27	88,01
Attention U-Net ResNet-50	0,7893	85,18	0,6520	77,64	89,37
Attention U-Net ResNet-50 [†]	0,8236	87,45	0,7001	81,96	90,50
Attention U-Net ResNet-50 [△]	0,8203	87,27	0,6953	81,26	90,62
Efficient-Unet++ EfficientNet-B6	0,8343	88,14	0,7157	83,53	90,70
Efficient-Unet++ EfficientNet-B6 [†]	0,8209	87,46	0,6961	80,32	91,44
Efficient-Unet++ EfficientNet-B6 [△]	0,8307	88,00	0,7104	82,31	91,17
ResNeSt-Unet++ ResNeSt-50d	0,8312	87,56	0,7112	85,66	88,63
ResNeSt-Unet++ ResNeSt-50d [†]	0,8245	87,62	0,7014	81,34	91,12
ResNeSt-Unet++ ResNeSt-50d [△]	0,8380	88,15	0,7212	85,69	89,53
MA-Net MiT-b2	0,7901	85,55	0,6530	76,05	90,84
MA-Net MiT-b2 [†]	0,7861	84,79	0,6476	78,20	88,45
MA-Net MiT-b2 [△]	0,8176	86,90	0,6914	82,07	89,60
ConvNeXt U-Net ConvNeXt Base	0,8026	86,18	0,6703	78,59	90,40
ConvNeXt U-Net ConvNeXt Base [†]	0,8169	86,97	0,6904	81,29	90,12
ConvNeXt U-Net ConvNeXt Base [△]	0,8201	86,94	0,6950	83,23	89,00
FPN ResNet-50	0,7391	81,45	0,5862	73,50	85,88
FPN ResNet-50 [†]	0,7603	81,26	0,6133	83,13	80,21
FPN ResNet-50 [△]	0,7511	81,73	0,6013	77,09	84,31
TransUNet ResNet-50 + ViT	0,7854	84,49	0,6467	79,41	87,31
TransUNet ResNet-50 + ViT [†]	0,7927	85,11	0,6566	79,66	88,14
TransUNet ResNet-50 + ViT [△]	0,7806	85,01	0,6401	74,59	90,80
SegFormer MiT-b3	0,7913	85,36	0,6547	77,63	89,66
SegFormer MiT-b3 [†]	0,8002	85,40	0,6669	81,76	87,42
SegFormer MiT-b3 [△]	0,8045	85,26	0,6730	84,82	85,51
Unet dlka MaxVit	0,7920	85,35	0,6568	81,16	87,09
Unet dlka MaxVit [†]	0,8136	86,94	0,6869	82,85	87,15
Unet dlka MaxVit [△]	0,8158	86,97	0,6898	83,15	87,92

[†] Modelos treinados com a estratégia RCAug.

[△] Modelos treinados com aumento geométrico.

As métricas D_C e IoU são apresentadas no intervalo [0,1], enquanto A_C , S_E e E_S são apresentadas em porcentagem. Os valores em negrito representam o melhor resultado dentro de cada arquitetura avaliada; as células destacadas em cinza indicam o melhor desempenho global em cada métrica.

arquiteturas, por exemplo, a Attention U-Net passou de $D_C = 0,7893$ (sem aumento) para 0,8236 com RCAug, ganho de cerca de 0.035, enquanto em outros casos houve estabilidade ou pequenas variações no desempenho. Já as transformações geométricas

mostraram resultados mais estáveis, com destaque para a ResNeSt-Unet++ ($D_C = 0,838$), que novamente apresentou o melhor desempenho geral.

Observa-se também que arquiteturas mais complexas nem sempre resultaram em melhor generalização. Modelos como ConvNeXt U-Net e Unet DLKA apresentaram desempenho competitivo, porém sem superar consistentemente as melhores variantes da U-Net++. Esse comportamento sugere que a capacidade de generalização depende não apenas da complexidade do modelo, mas também de sua adequação às características do domínio alvo.

Adicionalmente, nota-se maior variabilidade nas métricas de S_E e E_S , indicando que o equilíbrio entre falsos positivos e falsos negativos torna-se mais desafiador em cenários com maior diversidade de padrões morfológicos. É possível, ainda, que o tamanho reduzido do CryoNuSeg (30 imagens) contribua para essa variabilidade, tornando as estimativas de desempenho mais sensíveis a diferenças entre amostras.

A análise conjunta dos dois conjuntos de dados evidencia alguns padrões. As arquiteturas da família UNet++, em especial a ResNeSt-Unet++, apresentaram desempenho consistente e entre os mais altos em ambos os *datasets*, destacando-se como uma escolha robusta para a segmentação de núcleos celulares. Em contrapartida, arquiteturas mais simples, como a FPN, situaram-se sistematicamente abaixo das demais.

O uso de aumento de dados mostrou-se determinante no domínio de treinamento (leucoplasia), produzindo melhorias consistentes em todas as métricas e em todas as arquiteturas. Entre as estratégias, o aumento geométrico apresentou os melhores resultados em termos de D_C , seguido de perto pelo RCAug. Esse comportamento é coerente com o aumento da variabilidade do conjunto de treinamento, que tende a reduzir o *overfitting* e a favorecer a detecção das estruturas celulares.

No CryoNuSeg, utilizado para avaliar a generalização, o efeito do aumento de dados foi menos consistente, em parte das arquiteturas houve ganhos moderados, enquanto em outras o desempenho permaneceu praticamente estável. Esse resultado sugere que, embora o aumento de dados seja eficaz para ajustar os modelos ao domínio de origem, sua contribuição para a generalização inter-domínio é mais limitada, não garantindo, por si só, robustez diante de maior variabilidade morfológica.

Por fim, os modelos baseados em diferentes paradigmas, convolucionais, híbridos e baseados em *Transformers*, exibiram padrões de erro distintos. Como detalhado na análise de contribuição (Seção 5.2.5), arquiteturas de naturezas diferentes corrigem erros umas das outras, comportamento complementar que está reconhecidamente associado ao ganho proporcionado por *ensembles* (KUNCHEVA, 2003) e que motiva a aplicação das técnicas de combinação exploradas na próxima seção.

5.2 Resultados com *Ensemble Learning*

Nesta seção, são apresentados os resultados obtidos a partir da combinação de múltiplas arquiteturas de segmentação por meio de diferentes estratégias de *ensemble learning*. O objetivo é avaliar o impacto dessas técnicas na qualidade das segmentações, bem como analisar a contribuição individual dos modelos no processo de combinação.

5.2.1 Comparação entre Técnicas de *Ensemble*

A Tabela 4 apresenta o desempenho das diferentes técnicas de *ensemble*, considerando, para cada abordagem, a melhor combinação de modelos obtida em cada estratégia de aumento de dados.

Tabela 4 – Comparação entre técnicas de *ensemble* no *dataset* de leucoplasia considerando diferentes estratégias de aumento de dados.

Técnica	Cenário	D_C	A_C (%)	IoU	S_E (%)	E_S (%)
Hard	Sem aumento	0,8809	94,87	0,7871	88,59	96,58
	RCAug	0,8907	95,28	0,8030	89,86	96,76
	Geométrico	0,8912	95,29	0,8038	90,20	96,67
Soft	Sem aumento	0,8814	94,90	0,7880	88,52	96,64
	RCAug	0,8913	95,31	0,8040	89,86	96,79
	Geométrico	0,8924	95,34	0,8057	90,22	96,74
Weighted Soft	Sem aumento	0,8815	94,90	0,7881	88,52	96,64
	RCAug	0,8913	95,31	0,8040	89,86	96,79
	Geométrico	0,8924	95,34	0,8057	90,22	96,74
Union Max	Sem aumento	0,8714	94,16	0,7720	92,39	94,64
	RCAug	0,8856	94,90	0,7948	92,17	95,65
	Geométrico	0,8845	94,80	0,7929	93,00	95,29
Intersection Min	Sem aumento	0,8660	94,49	0,7636	83,15	97,58
	RCAug	0,8782	94,97	0,7829	84,65	97,79
	Geométrico	0,8830	95,14	0,7905	85,64	97,73

De forma geral, observa-se que as estratégias baseadas em média probabilística, *soft voting* e *weighted soft voting*, apresentaram os melhores resultados em todos os cenários avaliados. No caso sem aumento de dados, ambas as abordagens alcançaram desempenho similar ($D_C = 0,8815$), enquanto no cenário com RCAug atingiram $D_C = 0,8913$ e, com aumento geométrico, o melhor resultado global ($D_C = 0,8924$).

Esse comportamento indica que a combinação de probabilidades permite uma agregação mais refinada das predições individuais, preservando informações de incerteza dos modelos e resultando em maior equilíbrio entre S_E e E_S . Além disso, observa-se que a versão ponderada não apresentou ganhos expressivos em relação ao *soft voting* simples, sugerindo que os modelos já possuem desempenhos relativamente homogêneos.

A técnica de *hard voting*, embora apresente desempenho competitivo, obteve resultados consistentemente inferiores às abordagens probabilísticas. Isso pode ser explicado pela

perda de informação ao reduzir as predições a valores binários antes da combinação, limitando a capacidade de capturar nuances nas regiões de fronteira.

As abordagens baseadas em operações de conjunto sobre as máscaras binárias apresentaram comportamentos distintos. O método *union max* resultou em maior S_E (por exemplo, 93,00% no cenário geométrico), indicando maior capacidade de detectar regiões de interesse. No entanto, esse ganho ocorre à custa de uma redução na E_S , evidenciando o aumento de falsos positivos. Em contraste, o *intersection min* apresentou E_S elevada (até 97,79%), refletindo um comportamento mais conservador, porém com redução significativa na S_E .

Comparando os diferentes cenários de aumento de dados, observa-se que o desempenho das técnicas de *ensemble* acompanha a melhoria dos modelos individuais, com ganhos progressivos do cenário sem aumento para RCAug e, posteriormente, para aumento geométrico. Esse resultado reforça que o *ensemble* atua como um mecanismo complementar, potencializando a qualidade das predições, mas dependente da robustez dos modelos base.

De maneira geral, os resultados indicam que o *soft voting* representa a estratégia mais eficaz para a tarefa de segmentação considerada, oferecendo o melhor compromisso entre S_E e E_S , além de apresentar maior estabilidade entre diferentes configurações experimentais.

5.2.2 Impacto do Número de Modelos no *Ensemble*

Dado que a estratégia de *soft voting* apresentou o melhor desempenho global entre as técnicas avaliadas, foi conduzida uma análise adicional investigando o impacto do número de modelos no *ensemble*. Para isso, foram avaliadas combinações contendo de dois a cinco modelos, considerando os diferentes cenários de aumento de dados. Para cada quantidade de modelos, reporta-se a combinação que apresentou o melhor desempenho em termos de D_C entre as combinações avaliadas.

A Tabela 5 apresenta os resultados obtidos para cada configuração. De forma geral, observa-se um comportamento consistente entre os cenários, com aumento progressivo no desempenho à medida que novos modelos são incorporados ao *ensemble*.

No cenário sem aumento de dados, o D_C evolui de 0,8674 com dois modelos para 0,8721 com cinco modelos. Tendência semelhante é observada nos cenários com aumento geométrico e RCAug, nos quais os ganhos são mais expressivos nas primeiras combinações e tendem a se estabilizar à medida que mais modelos são adicionados.

Esse comportamento indica que a incorporação de novos modelos contribui para a melhoria do desempenho até certo ponto, a partir do qual os ganhos tornam-se marginais. Por exemplo, no cenário geométrico, o desempenho com quatro e cinco modelos é praticamente idêntico em termos de D_C (0,8814), com diferenças mínimas em S_E e E_S (de 90,50% e 96,37% com quatro modelos para 90,47% e 96,38% com cinco), sugerindo um ponto de saturação.

Tabela 5 – Impacto do número de modelos no desempenho do *ensemble* utilizando *soft voting*.

Cenário	Qte. Modelos	D_C	A_C (%)	IoU	S_E (%)	E_S (%)
Sem aumento	2	0,8674	94,65	0,7658	89,11	96,01
	3	0,8702	94,77	0,7702	89,35	96,09
	4	0,8718	94,83	0,7727	89,68	96,08
	5	0,8721	94,84	0,7733	89,72	96,09
RCAug	2	0,8788	95,16	0,7838	89,41	96,57
	3	0,8808	95,22	0,7869	89,96	96,51
	4	0,8814	95,25	0,7880	90,02	96,53
	5	0,8813	95,24	0,7879	90,12	96,49
Geométrico	2	0,8790	95,14	0,7841	89,86	96,43
	3	0,8805	95,18	0,7864	90,48	96,33
	4	0,8814	95,22	0,7879	90,50	96,37
	5	0,8814	95,22	0,7879	90,47	96,38

As métricas D_C e IoU são apresentadas no intervalo $[0,1]$, enquanto A_C , S_E e E_S são apresentadas em porcentagem.

Além disso, observa-se que os maiores ganhos ocorrem na transição de dois para três modelos, indicando que a diversidade inicial entre arquiteturas desempenha um papel fundamental na melhoria do *ensemble*. A partir desse ponto, modelos adicionais tendem a contribuir com informações mais redundantes do que complementares.

Outro aspecto relevante é o comportamento da S_E e da E_S . Em geral, a S_E tende a aumentar com o número de modelos, enquanto a E_S se mantém estável ou apresenta pequenas variações. Esse resultado sugere que o *ensemble* se torna progressivamente mais capaz de identificar regiões de interesse, sem comprometer significativamente a taxa de falsos positivos.

De maneira geral, os resultados indicam que *ensembles* com três ou quatro modelos já são capazes de capturar a maior parte dos ganhos proporcionados pela combinação, enquanto a inclusão de modelos adicionais apresenta benefícios limitados. Esse achado reforça que a diversidade entre os modelos é mais determinante do que a quantidade absoluta de modelos no *ensemble*.

5.2.3 Análise Qualitativa das Segmentações

A Figura 21 apresenta uma comparação qualitativa das segmentações para um exemplo representativo do *dataset* de leucoplasia, no cenário de aumento geométrico. A visualização codifica, por cor, os acertos e os tipos de erro de cada modelo, permitindo observar diretamente o comportamento das predições. De modo geral, os erros concentram-se nas bordas dos núcleos (finas faixas em vermelho e azul), refletindo discordâncias nas regiões de fronteira, ao passo que o interior das estruturas é majoritariamente recuperado

de forma correta. Entre os modelos individuais observam-se padrões distintos: alguns produzem regiões espúrias (falsos positivos, em vermelho), como a FPN e a TransUNet na porção inferior da imagem, enquanto outros deixam de detectar núcleos inteiros (falsos negativos, em azul). O resultado do *ensemble* por *soft voting* (destacado no topo) apresenta menor incidência desses erros extremos, equilibrando falsos positivos e falsos negativos e gerando contornos mais próximos do *ground truth*. Esse comportamento é coerente com a análise quantitativa de contribuição (Seção 5.2.5). Ao combinar modelos com erros complementares, o *ensemble* preserva os acertos individuais e corrige parte dos erros isolados.

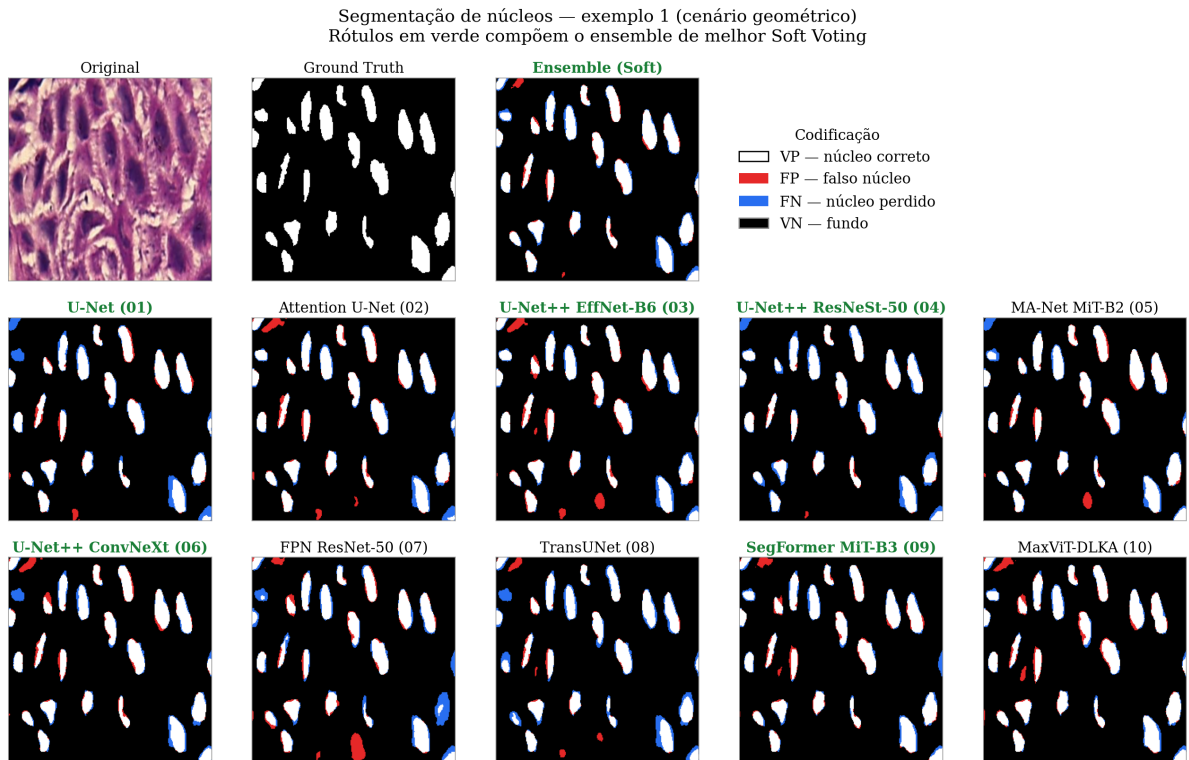


Figura 21 — Comparação qualitativa das segmentações no *dataset* de leucoplasia (cenário geométrico), para um exemplo representativo: imagem original, *ground truth*, resultado do *ensemble* por *soft voting* e predições dos dez modelos individuais. As cores indicam verdadeiros positivos (branco), falsos positivos (vermelho), falsos negativos (azul) e verdadeiros negativos (fundo, em preto). Os rótulos em verde identificam os modelos que compõem o *ensemble* selecionado.

5.2.4 Análise da Influência do Limiar de Binarização

Nesta etapa, foi avaliada a influência do limiar de binarização aplicado aos mapas probabilísticos gerados pelo *ensemble*. O objetivo foi investigar como a variação do limiar influencia o desempenho da segmentação, considerando o equilíbrio entre a detecção de regiões de interesse e a supressão de falsos positivos.

Para isso, o limiar foi variado no intervalo de 0 a 1, com incrementos de 0,1, sendo avaliadas as métricas D_C , A_C , IoU, S_E e E_S no *dataset* de leucoplasia. A análise foi conduzida para três cenários: sem aumento de dados (Tabela 6), com transformações geométricas (Tabela 7) e com a estratégia RCAug (Tabela 8).

Tabela 6 – Análise de variação do limiar no *dataset* de leucoplasia (sem aumento de dados).

Limiar	D_C	A_C (%)	IoU	S_E (%)	E_S (%)
0.0	0,3528	21,42	0,2142	100,00	0,00
0.1	0,5607	66,63	0,3896	99,41	57,70
0.2	0,8490	92,65	0,7377	96,49	91,61
0.3	0,8713	94,05	0,7720	94,00	94,07
0.4	0,8801	94,66	0,7858	91,41	95,55
0.5	0,8814	94,90	0,7880	88,52	96,64
0.6	0,8770	94,89	0,7810	85,02	97,58
0.7	0,8616	94,48	0,7568	80,19	98,37
0.8	0,8216	93,30	0,6972	72,06	99,08
0.9	0,4484	84,73	0,2890	28,99	99,92
1.0	0,0000	78,58	0,0000	0,00	100,00

As métricas D_C e IoU são apresentadas no intervalo $[0,1]$, enquanto A_C , S_E e E_S são apresentadas em porcentagem.

Tabela 7 – Análise de variação do limiar com aumento geométrico.

Limiar	D_C	A_C (%)	IoU	S_E (%)	E_S (%)
0.0	0,3528	21,42	0,2142	100,00	0,00
0.1	0,4201	40,88	0,2659	100,00	24,77
0.2	0,7418	85,17	0,5895	99,45	81,28
0.3	0,8640	93,48	0,7605	96,67	92,61
0.4	0,8862	94,85	0,7957	93,66	95,17
0.5	0,8924	95,34	0,8057	90,22	96,74
0.6	0,8832	95,16	0,7909	85,36	97,84
0.7	0,8506	94,20	0,7399	77,07	98,87
0.8	0,4996	85,68	0,3330	33,38	99,93
0.9	0,0026	78,61	0,0013	0,13	100,00
1.0	0,0000	78,58	0,0000	0,00	100,00

As métricas D_C e IoU são apresentadas no intervalo $[0,1]$, enquanto A_C , S_E e E_S são apresentadas em porcentagem.

Os resultados evidenciam um comportamento consistente entre os cenários. Para valores baixos de limiar (próximos de 0), observa-se S_E elevada (próxima de 1), porém com E_S nula, indicando que praticamente todos os pixels são classificados como pertencentes à região de interesse. Por outro lado, para limiares elevados (próximos de 1), ocorre o comportamento oposto, com E_S elevada e S_E próxima de zero.

Tabela 8 – Análise de variação do limiar com estratégia RCAug.

Limiar	D_C	A_C (%)	IoU	S_E (%)	E_S (%)
0.0	0,3528	21,42	0,2142	100,00	0,00
0.1	0,4197	40,78	0,2656	100,00	24,64
0.2	0,7296	84,19	0,5743	99,56	80,00
0.3	0,8623	93,39	0,7579	96,55	92,53
0.4	0,8869	94,90	0,7969	93,44	95,29
0.5	0,8913	95,31	0,8040	89,86	96,79
0.6	0,8830	95,16	0,7905	85,19	97,88
0.7	0,8487	94,14	0,7371	76,68	98,90
0.8	0,4626	85,00	0,3009	30,14	99,95
0.9	0,0004	78,59	0,0002	0,02	100,00
1.0	0,0000	78,58	0,0000	0,00	100,00

As métricas D_C e IoU são apresentadas no intervalo $[0,1]$, enquanto A_C , S_E e E_S são apresentadas em porcentagem.

No cenário sem aumento de dados (Tabela 6), o melhor desempenho em termos de D_C foi obtido para o limiar igual a 0,5 ($D_C = 0,8814$), com queda progressiva para valores mais baixos ou mais altos. No cenário com aumento geométrico (Tabela 7), o melhor resultado também ocorreu em 0,5 ($D_C = 0,8924$), representando o melhor desempenho global entre os cenários avaliados. De forma semelhante, com a estratégia RCAug (Tabela 8), observa-se ganho consistente em todas as métricas, com melhor resultado igualmente em 0,5 ($D_C = 0,8913$).

Esses resultados indicam que o limiar de 0,5 constitui o ponto operacional mais adequado quando o critério adotado é a qualidade global da sobreposição, uma vez que esse valor maximizou D_C , IoU e A_C nos três cenários avaliados. Entretanto, o limiar de 0,5 não corresponde exatamente ao ponto de maior equilíbrio entre S_E e E_S . No cenário sem aumento de dados, essas métricas apresentam valores praticamente iguais no limiar 0,3, com $S_E = 94,00\%$ e $E_S = 94,07\%$. Nos cenários com aumento geométrico e RCAug, o cruzamento das curvas ocorre entre os limiares 0,3 e 0,4, ficando mais próximo de 0,4. A utilização do limiar 0,5 produz segmentações ligeiramente mais conservadoras, reduzindo a ocorrência de falsos positivos e elevando E_S , embora isso resulte em uma redução moderada de S_E . Dessa forma, enquanto um limiar próximo de 0,4 poderia ser considerado em aplicações nas quais a redução de falsos negativos fosse prioritária, a adoção de 0,5 permanece justificada neste trabalho por maximizar D_C , definido como a principal métrica de avaliação da qualidade das segmentações. Além disso, sob as estratégias de aumento de dados, o desempenho mostrou-se relativamente estável entre os limiares vizinhos de 0,4 e 0,5, indicando que a escolha desse ponto operacional não depende de uma variação abrupta no comportamento dos modelos.

A Figura 22 ilustra o comportamento das métricas em função do limiar para os três cenários avaliados. Nota-se a relação inversa entre S_E e E_S , bem como a existência de

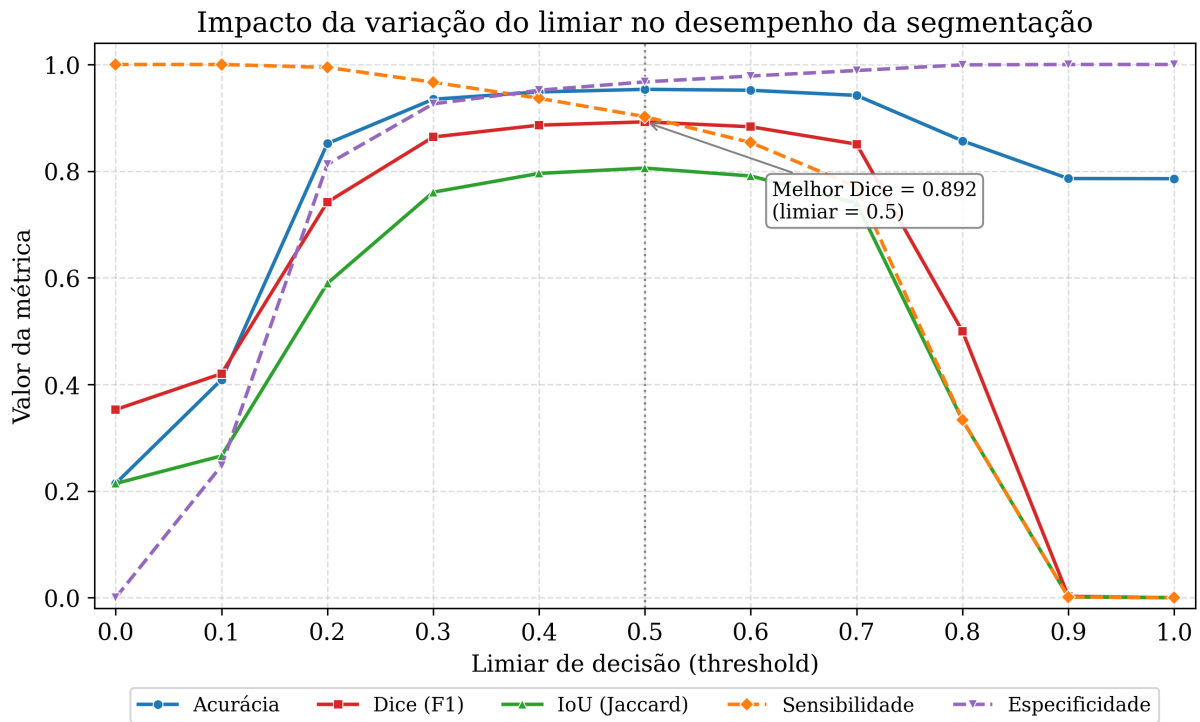


Figura 22 – Resultado da variação do limiar

uma região próxima ao limiar de 0,5 na qual o D_C é maximizado.

5.2.5 Análise de Contribuição dos Modelos no *Ensemble*

Como os melhores desempenhos individuais foram obtidos com o aumento geométrico e o *soft voting* mostrou-se a técnica de combinação mais eficaz entre as avaliadas, a análise de contribuição foi restrita a essa configuração (aumento geométrico com *soft voting*), permitindo avaliar o potencial máximo de combinação entre os modelos. Para essa configuração, foi realizada uma análise detalhada da contribuição individual de cada modelo no *ensemble*, por meio de três métricas: *kept rate*, *discarded rate* e *error mitigation*.

A métrica de *kept rate* representa a proporção de predições corretas de um modelo individual que são preservadas pelo *ensemble*. Por sua vez a *discarded rate* indica a fração dessas predições corretas que são perdidas após a combinação. A métrica de *error mitigation* quantifica a capacidade do *ensemble* de corrigir predições incorretas do modelo individual.

A Tabela 9 apresenta os resultados obtidos para cada arquitetura participante do *ensemble*.

Os resultados mostram taxas de retenção entre 0,9900 e 0,9924, indicando que o *ensemble* preservou mais de 99% dos *pixels* corretamente classificados por cada modelo individual. Consequentemente, as taxas de descarte permaneceram entre 0,0076 e 0,0100, demonstrando que o ganho produzido pela combinação foi acompanhado por uma perda inferior ou igual a 1% das decisões previamente corretas.

Tabela 9 – Análise de contribuição dos modelos no *ensemble* utilizando *soft voting* com aumento geométrico.

Modelo	<i>Kept rate</i>	<i>Discarded rate</i>	<i>Error mitigation</i>
Baseline U-Net	0,9915	0,0085	0,3107
Efficient-Unet++	0,9924	0,0076	0,2757
ResNeSt-Unet++	0,9900	0,0100	0,2489
ConvNeXt U-Net	0,9919	0,0081	0,2402
SegFormer	0,9913	0,0087	0,2916

A mitigação de erros variou entre 0,2402 e 0,3107. Isso significa que o *ensemble* corrigiu entre 24,02% e 31,07% dos *pixels* classificados incorretamente pelos modelos individuais. Os maiores valores foram observados para a Baseline U-Net (0,3107) e o SegFormer (0,2916), indicando que a decisão conjunta compensou uma parcela maior das limitações individuais dessas arquiteturas. Em contrapartida, a ConvNeXt U-Net (0,2402) e a ResNeSt-Unet++ (0,2489) apresentaram valores menores, sugerindo que os erros residuais desses modelos eram mais frequentemente compartilhados pelos demais membros ou estavam associados a regiões de maior dificuldade.

Em conjunto, as três métricas propostas evidenciam um comportamento favorável da fusão: o comitê preserva quase integralmente a informação correta já produzida pelos modelos e, simultaneamente, corrige uma parcela relevante de seus erros. As métricas não representam uma estimativa de contribuição marginal, que exigiria a remoção individual de membros do comitê, mas quantificam de forma direta o efeito da decisão conjunta sobre os acertos e erros de cada arquitetura.

5.3 Análise Estatística dos Resultados

Esta seção apresenta a avaliação estatística dos resultados obtidos, com o objetivo de verificar a significância das diferenças entre os modelos individuais, o método de *ensemble*, e o impacto das estratégias de aumento de dados.

Inicialmente, foi realizada a verificação da normalidade dos dados por meio do teste de Shapiro-Wilk para todos os cenários avaliados: modelos individuais, *ensemble* e estratégias de aumento de dados.

Para os cenários de aumento de dados, os resultados são apresentados na Tabela 11.

Observa-se que, em todos os casos, os valores de p são superiores a 0.05, não havendo evidências para rejeição da hipótese de normalidade. No entanto, devido ao reduzido número de amostras ($n=3$), optou-se pela utilização de testes não paramétricos como abordagem complementar, garantindo maior robustez na análise.

Para avaliar diferenças globais entre os modelos, foi aplicado o teste de Friedman, apropriado para dados pareados e não paramétricos. O resultado indicou diferenças estatisticamente significativas entre os modelos avaliados:

Tabela 10 – Teste de normalidade de Shapiro-Wilk para os modelos avaliados.

Modelo	W	p-valor
Baseline U-Net	0.923	0.463
Attention U-Net	0.863	0.277
Efficient-Unet++	0.999	0.956
ResNeSt-Unet++	0.954	0.589
MA-Net	0.922	0.458
ConvNeXt U-Net	0.924	0.468
FPN	0.995	0.866
TransUNet	0.962	0.627
SegFormer	0.966	0.644
Unet dlka	0.975	0.697
Ensemble	0.976	0.700

Tabela 11 – Teste de normalidade de Shapiro-Wilk para estratégias de aumento de dados.

Cenário	W	p-valor
Sem aumento	0.976	0.700
RCAug	0.890	0.354
Geométrico	0.898	0.380

$$\chi^2(10) = 29.1, \quad p = 0.001 \quad (19)$$

Este resultado demonstra que pelo menos um dos modelos apresenta desempenho significativamente distinto dos demais.

Para investigar especificamente o desempenho do *ensemble*, foram realizados testes t pareados entre o *ensemble* e cada modelo individual, conforme apresentado na Tabela 12.

Tabela 12 – Teste t pareado entre o *ensemble* e os modelos individuais.

Comparação	t	p-valor
Ensemble vs Baseline U-Net	11.43	0.008
Ensemble vs Attention U-Net	24.94	0.002
Ensemble vs Efficient-Unet++	26.67	0.001
Ensemble vs ResNeSt-Unet++	53.27	< 0.001
Ensemble vs MA-Net	12.55	0.006
Ensemble vs ConvNeXt U-Net	11.29	0.008
Ensemble vs FPN	43.44	< 0.001
Ensemble vs TransUNet	109.32	< 0.001
Ensemble vs SegFormer	15.39	0.004
Ensemble vs Unet dlka	9.91	0.010

Observa-se que o *ensemble* apresentou desempenho estatisticamente superior a todos os modelos individuais ($p < 0.05$ em todas as comparações). Além disso, destaca-se que o *ensemble* obteve o melhor desempenho em todas as sementes avaliadas, evidenciando consistência e robustez.

A análise do impacto das estratégias de aumento de dados foi realizada considerando três cenários: sem aumento, RCAug e geométrico.

Inicialmente, foi aplicado o teste de Friedman:

$$\chi^2(2) = 5.64, \quad p = 0.060 \quad (20)$$

Embora o resultado não tenha atingido significância estatística ao nível de 5%, observa-se uma tendência de diferença entre os cenários.

Para uma análise mais detalhada, foram realizados testes t pareados entre os cenários, conforme apresentado na Tabela 13.

Tabela 13 – Teste t pareado entre estratégias de aumento de dados.

Comparação	t	p-valor
Geométrico vs Sem aumento	13.52	0.003
RCAug vs Sem aumento	9.88	0.005
Geométrico vs RCAug	-1.90	0.901

Os resultados indicam que ambas as estratégias de aumento de dados apresentam melhora estatisticamente significativa em relação ao cenário sem aumento ($p < 0.01$). Por outro lado, não foi observada diferença estatisticamente significativa entre as estratégias RCAug e Geométrico.

Os resultados obtidos evidenciam que o uso de técnicas de *ensemble* proporciona ganhos estatisticamente significativos em relação aos modelos individuais, sendo consistente em todas as execuções analisadas.

Adicionalmente, verificou-se que o uso de aumento de dados é um fator determinante para a melhoria do desempenho dos modelos. Embora não tenha sido possível identificar diferença significativa entre as estratégias de aumento avaliadas, ambas apresentaram ganhos consistentes em relação ao cenário sem aumento.

Por fim, destaca-se que, apesar do número reduzido de amostras ($n=3$), a consistência dos resultados ao longo de diferentes sementes reforça a robustez das conclusões obtidas.

5.4 Comparação com o Estado da Arte

Para analisar a abordagem proposta em relação aos métodos existentes na literatura, os resultados mais promissores obtidos pelos comitês foram comparados com trabalhos recentes avaliados nos dois domínios considerados neste estudo: o conjunto de leucoplasia bucal e o *benchmark* público CryoNuSeg. Para essa comparação, foi utilizado o coeficiente Dice (D_C), por ser uma das métricas mais frequentemente reportadas em estudos de segmentação nuclear. Os resultados estão apresentados na Tabela 14, na qual o *ensemble* por *soft voting*, estratégia que apresentou o melhor desempenho global nas análises anteriores, é comparado aos métodos da literatura. Os valores associados à proposta são reportados nos três cenários experimentais: sem aumento de dados, com RCAug e com aumento geométrico.

Tabela 14 – Comparação da abordagem proposta com trabalhos recentes da literatura, considerando o coeficiente Dice (D_C) nos *datasets* de leucoplasia bucal e CryoNuSeg.

Método	Leucoplasia	CryoNuSeg
Oliveira et al. (2020)	0,8369	—
Swain, Cheoi e Ko (2024)	—	0,8942
Abdel-Nasser, Singh e Mohamed (2022)	—	0,8653
Atlan e Hançer (2025)	—	0,8424
Mahbod et al. (2021)	—	0,7910
Abordagem proposta – sem aumento	0,8814	0,8450
Abordagem proposta – RCAug	0,8913	0,8577
Abordagem proposta – geométrico	0,8924	0,8569

Os valores de D_C são apresentados no intervalo $[0,1]$. Os valores em negrito indicam o melhor resultado em cada *dataset*.

Na base de leucoplasia bucal, a abordagem proposta apresentou desempenho superior ao método de Oliveira et al. (2020), avaliado sobre o mesmo conjunto de imagens. O melhor resultado foi obtido pelo comitê com aumento geométrico, que alcançou D_C de 0,8924, enquanto o trabalho de Oliveira et al. (2020) reportou D_C de 0,8369. Esse ganho indica a relevância da combinação de arquiteturas recentes de segmentação por meio de *ensemble learning* para o domínio investigado. Considerando a escassez de estudos voltados especificamente à segmentação de núcleos em imagens histopatológicas de leucoplasia bucal, esse resultado evidencia o potencial da abordagem proposta como referência inicial para trabalhos futuros nesse contexto.

Na base CryoNuSeg, utilizada para avaliar o comportamento da proposta em um domínio histológico distinto, os comitês apresentaram desempenho competitivo em relação aos métodos da literatura. A abordagem com RCAug obteve o melhor resultado nesse conjunto, com D_C de 0,8577, valor superior aos resultados reportados por Atlan e Hançer (2025) e Mahbod et al. (2021), e próximo ao desempenho de Abdel-Nasser, Singh e Mohamed (2022). Por outro lado, o método de Swain, Cheoi e Ko (2024) apresentou o maior valor de Dice entre os trabalhos avaliados. Ainda assim, os resultados obtidos mostram que a metodologia proposta mantém desempenho consistente mesmo quando aplicada a uma base com maior variabilidade morfológica e composta por tecidos distintos daqueles utilizados no domínio principal deste estudo.

Estes resultados mostram que a metodologia proposta apresenta desempenho competitivo em relação aos trabalhos da literatura e é capaz de melhorar a segmentação semântica de núcleos em imagens histopatológicas de leucoplasia bucal. Embora nem sempre apresente os maiores valores no conjunto CryoNuSeg, os comitês desenvolvidos mantiveram resultados sólidos em um domínio distinto, evidenciando a robustez da abordagem em diferentes cenários de segmentação nuclear. Esses achados reforçam que a combinação

de arquiteturas heterogêneas, especialmente por meio de estratégias probabilísticas como o *soft voting*, constitui uma alternativa promissora para explorar a complementaridade entre modelos e aprimorar a qualidade das segmentações.

Conclusões

Esta dissertação investigou o uso de comitês de modelos heterogêneos de aprendizagem profunda para a segmentação semântica de núcleos celulares em imagens histopatológicas de leucoplasia bucal. A proposta foi motivada pela importância da segmentação nuclear como etapa fundamental para análises morfométricas em lesões orais potencialmente malignas e pela escassez de estudos voltados especificamente a esse domínio. Além disso, buscou-se explorar a complementaridade entre arquiteturas de diferentes naturezas, considerando modelos convolucionais, híbridos e baseados em mecanismos de atenção.

Foram avaliadas dez arquiteturas de segmentação, submetidas a diferentes estratégias de aumento de dados e posteriormente combinadas por técnicas de *ensemble learning*. Os resultados mostraram que modelos da família UNet++, especialmente a ResNeSt-Unet++ e a ConvNeXt U-Net, apresentaram desempenho individual consistente. No conjunto de leucoplasia, o melhor resultado individual foi obtido pela ResNeSt-Unet++ com aumento geométrico, alcançando coeficiente Dice (D_C) de 0.8853. As estratégias de aumento de dados contribuíram para a melhoria do desempenho no domínio de treinamento, embora seu efeito na generalização para o CryoNuSeg tenha sido menos uniforme.

A combinação dos modelos por meio de comitês demonstrou ser uma estratégia eficaz para aprimorar a segmentação. As abordagens baseadas em média probabilística, especialmente o *soft voting*, apresentaram os melhores resultados, atingindo D_C de 0.8924 no cenário com aumento geométrico e superando o melhor modelo individual. Esse resultado indica que a preservação da informação probabilística das predições favorece uma fusão mais equilibrada do que estratégias baseadas apenas em decisões binárias ou operações extremas sobre máscaras.

A análise da quantidade de modelos no comitê indicou que os maiores ganhos ocorreram nas primeiras combinações, com tendência de saturação a partir de três ou quatro arquiteturas. Esse comportamento sugere que a complementaridade entre os modelos é mais relevante do que a simples inclusão de um maior número de membros. Como contribuição metodológica desta dissertação, foram propostas e formalizadas as métricas *kept rate*, *discarded rate* e *error mitigation*, destinadas a quantificar, respectivamente, a pre-

servação dos acertos individuais, a perda desses acertos após a combinação e a correção dos erros de cada modelo pelo comitê. Na configuração analisada, composta pelo *soft voting* com aumento geométrico, a aplicação dessas métricas mostrou que o comitê preservou entre 99,00% e 99,24% dos acertos de cada modelo, descartou no máximo 1,00% deles e corrigiu entre 24,02% e 31,07% dos erros individuais. Dessa forma, foi possível compreender não apenas que o comitê melhora o desempenho, mas também como esse ganho emerge da interação entre as arquiteturas.

A validação estatística reforçou os achados experimentais, indicando diferenças significativas entre os modelos avaliados e desempenho superior do comitê em relação às arquiteturas individuais nas comparações realizadas. Assim, as hipóteses propostas foram majoritariamente corroboradas nos cenários avaliados: os comitês apresentaram desempenho superior aos modelos individuais, as estratégias probabilísticas de fusão foram mais eficazes, o aumento de dados contribuiu sobretudo no domínio de treinamento e a complementaridade entre arquiteturas mostrou-se um fator importante para o ganho da combinação.

Apesar dos resultados obtidos, algumas limitações devem ser consideradas. A abordagem adotada realiza segmentação semântica binária, não separando instâncias individuais de núcleos, o que limita análises morfométricas por célula em regiões de sobreposição. Além disso, os conjuntos de dados utilizados possuem tamanho limitado, especialmente o CryoNuSeg, o que pode influenciar as estimativas de generalização. Por fim, embora os testes estatísticos tenham indicado resultados consistentes, o número reduzido de execuções recomenda cautela na extrapolação dos achados.

Em síntese, esta dissertação demonstrou que comitês heterogêneos de aprendizagem profunda constituem uma alternativa eficaz para a segmentação semântica de núcleos em imagens histopatológicas de leucoplasia bucal. Ao combinar arquiteturas com diferentes vieses indutivos, a abordagem proposta obteve desempenho superior aos modelos individuais e, com o apoio das métricas de contribuição propostas nesta dissertação, forneceu uma análise mais transparente sobre o papel de cada modelo na decisão conjunta. Dessa forma, o trabalho contribui tanto para o avanço da segmentação histológica em leucoplasia bucal quanto para a compreensão do uso de comitês heterogêneos em imagens biomédicas.

6.1 Trabalhos Futuros

Como continuidade deste trabalho, uma direção relevante consiste na ampliação das bases de leucoplasia bucal, incluindo imagens de diferentes centros, equipamentos de digitalização e protocolos de coloração, a fim de avaliar com maior rigor a capacidade de generalização dos modelos.

Outra possibilidade é estender a abordagem para segmentação de instâncias, permitindo a separação individual dos núcleos e a extração de atributos morfométricos por

célula, especialmente em regiões de sobreposição e aglomeração celular.

Também se mostra promissora a investigação de novas estratégias de composição dos comitês, incluindo métodos aprendidos de fusão, seleção automática de membros, modelos fundacionais adaptados ao domínio histopatológico e critérios baseados na diversidade ou correlação de erros entre modelos.

Por fim, as segmentações obtidas podem ser integradas a etapas posteriores de análise histopatológica, como extração de características morfométricas, graduação da displasia epitelial e predição de transformação maligna, aproximando a abordagem proposta de um *pipeline* computacional mais completo de apoio ao diagnóstico.

6.2 Contribuições desta Pesquisa

Ao longo do desenvolvimento desta dissertação de mestrado, foram elaboradas as seguintes publicações relacionadas à pesquisa:

1. L. N. F. B. SAAR, A. B. Silva, D. L. L. de Oliveira, P. Oliveira, L. A. Neves, A. S. Martins, T. A. Tosta, P. R. de Faria, M. Z. do Nascimento. Enhancing Oral Leukoplakia Segmentation in Histopathological Images Through Heterogeneous Deep Learning Ensemble with Soft Voting. **BRACIS 36th Brazilian Conference on Intelligent Systems**, 2026.

Além das publicações listadas anteriormente, um outro trabalho relacionado a esta pesquisa foi submetido para avaliação em evento científico, encontrando-se, até o momento, em processo de revisão.

1. L. N. F. B. SAAR, A. B. Silva, D. L. L. de Oliveira, P. Oliveira, L. A. Neves, A. S. Martins, T. A. Tosta, P. R. de Faria, M. Z. do Nascimento. Deep Learning Ensembles for Nuclear Segmentation of Histopathology Images. **39th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**, 2026.

Adicionalmente, com o objetivo de promover a reprodutibilidade e o reaproveitamento dos resultados obtidos nesta pesquisa, as máscaras de segmentação de núcleos celulares geradas pelos modelos e comitês desenvolvidos neste trabalho foram organizadas e disponibilizadas publicamente no repositório Zenodo:

1. Saar, L. N., Zanchetta do Nascimento, M., Silva, A., Lucas Latorre de Oliveira, D., Neves, L. A., Ávila Oliveira, P. A., Martins, A. S., Tosta, T. & de Faria, P. R. (2026). Semantic Segmentation Masks of Cell Nuclei in Oral Leukoplakia Histopathological Images [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.21206251>

Referências

- ABDEL-NASSER, M.; SINGH, V. K.; MOHAMED, E. M. Efficient staining-invariant nuclei segmentation approach using self-supervised deep contrastive network. **Diagnostics**, MDPI, v. 12, n. 12, p. 3024, 2022.
- AGUIRRE-URIZAR, J. M.; MENDOZA, I. Lafuente-Ibáñez de; WARNAKULASURIYA, S. Malignant transformation of oral leukoplakia: Systematic review and meta-analysis of the last five years. **Oral Diseases**, v. 27, n. 8, p. 1881–1895, 2021.
- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. **2017 International Conference on Engineering and Technology (ICET)**. 2017. p. 1–6. Disponível em: <<https://doi.org/10.1109/ICEngT echnol.2017.8308186>>.
- ATLAN, F.; HANÇER, E. A methodology for nuclei segmentation in h&e images using efficientnetb7 guided u-net and multiple data augmentation strategies. **Turkish Journal of Electrical Engineering and Computer Sciences**, The Scientific and Technological Research Council of Türkiye (TÜBİTAK), v. 33, n. 3, p. 372–391, 2025.
- AZAD, R. et al. Beyond self-attention: Deformable large kernel attention for medical image segmentation. In: **Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)**. [S.l.: s.n.], 2024. p. 1287–1297. ArXiv:2309.00121.
- AZARMEHR, N. et al. A neural architecture search based framework for segmentation of epithelium, nuclei and oral epithelial dysplasia grading. In: SPRINGER. **Annual Conference on Medical Image Understanding and Analysis**. [S.l.], 2022. p. 357–370.
- BAGAN, J. et al. Proliferative verrucous leukoplakia: a concise update. **Oral Diseases**, v. 16, n. 4, p. 328–332, 2010.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. [S.l.]: Springer, 2006. (Information Science and Statistics).
- BRAY, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, v. 74, n. 3, p. 229–263, 2024.
- BREIMAN, L. Bagging predictors. **Machine Learning**, v. 24, n. 2, p. 123–140, 1996.

- CHEN, J. et al. Transunet: Transformers make strong encoders for medical image segmentation. **arXiv preprint arXiv:2102.04306**, 2021.
- DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2009. p. 248–255.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: **Multiple Classifier Systems (MCS)**. [S.l.]: Springer, 2000, (Lecture Notes in Computer Science, v. 1857). p. 1–15.
- DOI, K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. **Computerized Medical Imaging and Graphics**, v. 31, n. 4–5, p. 198–211, 2007.
- DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: **International Conference on Learning Representations (ICLR)**. [S.l.: s.n.], 2021.
- FAN, T. et al. Ma-net: A multi-scale attention network for liver and tumor segmentation. **IEEE Access**, v. 8, p. 179656–179665, 2020.
- FISCHER, A. H. et al. Hematoxylin and eosin staining of tissue and cell sections. **Cold Spring Harbor Protocols**, v. 2008, n. 5, p. pdb.prot4986, 2008.
- FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. **Journal of Computer and System Sciences**, v. 55, n. 1, p. 119–139, 1997.
- FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **Annals of Statistics**, v. 29, n. 5, p. 1189–1232, 2001.
- GANAIÉ, M. A. et al. Ensemble deep learning: A review. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 115, p. 105151, 2022.
- GARTNER, L. P. **Tratado de Histologia Em Cores**. [S.l.]: Elsevier Brasil, 2007.
- GONZALEZ, R. C.; WOODS, R. **Digital Image Processing**. [S.l.]: Pearson, 2018.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>.
- GRAHAM, S. et al. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. **Medical Image Analysis**, v. 58, p. 101563, 2019.
- GUDHE, N. R. et al. Nuclei instance segmentation from histopathology images using bayesian dropout based deep learning. **BMC Medical Imaging**, Springer Science and Business Media LLC, v. 23, n. 1, p. 162, out. 2023. ISSN 1471-2342. Disponível em: <<http://dx.doi.org/10.1186/s12880-023-01121-3>>.
- HAYKIN, S. S. et al. **Neural networks and learning machines**. [S.l.]: Pearson education Upper Saddle River, 2009. v. 3.
- HE, K. et al. Mask r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2961–2969.

_____. Deep residual learning for image recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2018. p. 7132–7141.

IRSHAD, H. et al. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. **IEEE reviews in biomedical engineering**, IEEE, v. 7, p. 97–114, 2014.

JOHNSON, K. **Histology and Cell Biology**. [S.l.]: Harwal Publishing Company, 1991. (A Williams&Wilkins medical publication). ISBN 9780683062106.

JUNQUEIRA, L. C. U.; CARNEIRO, J. **Histologia básica: texto e atlas**. [S.l.]: Guanabara Koogan, 2013.

KADASKAR, M.; PATIL, N. Image analysis of nuclei histopathology using deep learning: A review of segmentation, detection, and classification. **SN Computer Science**, v. 4, n. 5, 2023.

KADEMANI, D. **Improving Outcomes in Oral Cancer: A Clinical and Translational Update**. Springer Nature, 2019. Disponível em: <<https://doi.org/10.1007/978-3-030-30094-4>>.

KAMNITSAS, K. et al. Ensembles of multiple models and architectures for robust brain tumour segmentation. In: **Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes), MICCAI Workshop**. [S.l.]: Springer, 2017. p. 450–462.

_____. Ensembles of multiple models and architectures for robust brain tumour segmentation. In: **Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes), MICCAI Workshop**. [S.l.: s.n.], 2018. p. 450–462.

KIRILLOV, A. et al. Panoptic segmentation. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2019. p. 9404–9413.

KITTLER, J. et al. On combining classifiers. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 3, p. 226–239, 1998.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, p. 1097–1105, 2012.

KUMAR, N. et al. A multi-organ nucleus segmentation challenge. **IEEE transactions on medical imaging**, IEEE, v. 39, n. 5, p. 1380–1391, 2019.

KUMAR, V.; ABBAS, A. K.; ASTER, J. C. **Robbins patologia básica**. 9. ed. [S.l.]: Elsevier Brasil, 2013.

- KUMAR, V.; ASTER, J. C.; ABBAS, A. **Robbins and Cotran Patologia-Bases Patológicas das Doenças**. [S.l.]: Elsevier Brasil, 2010.
- KUNCHEVA, L. I. That elusive diversity in classifier ensembles. In: SPRINGER. **Iberian conference on pattern recognition and image analysis**. [S.l.], 2003. p. 1126–1138.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2324, 1998.
- LIN, T.-Y. et al. Feature pyramid networks for object detection. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2017. p. 2117–2125.
- LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical Image Analysis**, v. 42, p. 60–88, 2017.
- LIU, Z. et al. A convnet for the 2020s. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2022. p. 11976–11986.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 3431–3440.
- MACENKO, M. et al. A method for normalizing histology slides for quantitative analysis. In: **IEEE International Symposium on Biomedical Imaging (ISBI)**. [S.l.: s.n.], 2009. p. 1107–1110.
- MADABHUSHI, A.; LEE, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. **Medical Image Analysis**, v. 33, p. 170–175, 2016.
- MAHBOD, A. et al. Improving generalization capability of deep learning-based nuclei instance segmentation by non-deterministic train time and deterministic test time stain normalization. **Computational and Structural Biotechnology Journal**, Elsevier BV, v. 23, p. 669–678, jan. 2024. ISSN 2001-0370. Disponível em: <<http://dx.doi.org/10.1016/j.csbj.2023.12.042>>.
- _____. Cryonuseg: A dataset for nuclei instance segmentation of cryosectioned h&e-stained histological images. **Computers in biology and medicine**, Elsevier, v. 132, p. 104349, 2021.
- MINAEE, S. et al. Image segmentation using deep learning: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, 2021.
- MITCHELL, T. M. **Machine Learning**. [S.l.]: McGraw-Hill, 1997. (McGraw-Hill Series in Computer Science).
- MÜLLER, S. Oral epithelial dysplasia, atypical verrucous lesions and oral potentially malignant disorders: focus on histopathology. **Oral surgery, oral medicine, oral pathology and oral radiology**, Elsevier, v. 125, n. 6, p. 591–602, 2018. Disponível em: <<https://doi.org/10.1016/j.oooo.2018.02.012>>.

- NIAZI, M. K. K.; PARWANI, A. V.; GURCAN, M. N. Digital pathology and artificial intelligence. **The Lancet Oncology**, v. 20, n. 5, p. e253–e261, 2019.
- OLIVEIRA, P. A. d. Á. et al. Uso da inteligência artificial na caracterização nuclear e classificação entre leucoplasia bucal e leucoplasia verrucosa proliferativa. Universidade Federal de Uberlândia, 2020.
- OTSU, N. A threshold selection method from gray-level histograms. **IEEE transactions on systems, man, and cybernetics**, IEEE, v. 9, n. 1, p. 62–66, 1979.
- PAN, S. J.; YANG, Q. A survey on transfer learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 22, n. 10, p. 1345–1359, 2010.
- PRUSTY, M. R. et al. Nuclei segmentation in histopathology images using structure-preserving color normalization based ensemble deep learning frameworks. **Computers, Materials & Continua**, Tech Science Press, v. 77, n. 3, p. 3077–3094, 2023. ISSN 1546-2226. Disponível em: <<http://dx.doi.org/10.32604/cmc.2023.042718>>.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: **Medical Image Computing and Computer-Assisted Intervention (MICCAI)**. [S.l.]: Springer, 2015. p. 234–241.
- ROY, A. G.; NAVAB, N.; WACHINGER, C. Concurrent spatial and channel “squeeze & excitation” in fully convolutional networks. In: **Medical Image Computing and Computer-Assisted Intervention (MICCAI)**. [S.l.]: Springer, 2018. p. 421–429.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986.
- SAGI, O.; ROKACH, L. Ensemble learning: A survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 4, p. e1249, 2018. Disponível em: <<https://doi.org/10.1002/widm.1249>>.
- SANTOS, D. dos et al. Impacts of color space transformations on dysplastic nuclei segmentation using cnn. In: SBC. **Anais do XVI Workshop de Visão Computacional**. 2020. p. 6–11. Disponível em: <<https://doi.org/10.5753/wvc.2020.13475>>.
- SANTOS, D. F. D. et al. Influence of data augmentation strategies on the segmentation of oral histological images using fully convolutional neural networks. **Journal of Imaging Informatics in Medicine**, 2023.
- SANTOS, M. d. O. Estimativa de incidência de câncer no brasil, 2023-2025. **Revista Brasileira de Cancerologia**, v. 69, n. 1, 2023. Disponível em: <<https://rbc.inca.gov.br/index.php/revista/article/view/3700>>.
- SHAFI, S.; PARWANI, A. V. Artificial intelligence in diagnostic pathology. **Diagnostic Pathology**, v. 18, p. 109, 2023.
- SHAMSHAD, F. et al. Transformers in medical imaging: A survey. **Medical Image Analysis**, v. 88, p. 102802, 2023.

- SHEPHARD, A. J. et al. A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. **npj Precision Oncology**, Springer Science and Business Media LLC, v. 8, n. 1, jun. 2024. ISSN 2397-768X. Disponível em: <<http://dx.doi.org/10.1038/s41698-024-00624-8>>.
- _____. **Transformer-based Model for Oral Epithelial Dysplasia Segmentation**. 2023. Disponível em: <<https://arxiv.org/abs/2311.05452>>.
- SHI, P. et al. Nuclei segmentation of he stained histopathological images based on feature global delivery connection network. **PLOS ONE**, Public Library of Science, v. 17, n. 9, p. e0273682, 2022.
- SHIROKIKH, B. et al. Accelerating 3d medical image segmentation by adaptive small-scale target localization. **Journal of Imaging**, MDPI, v. 7, n. 2, p. 35, 2021.
- SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, v. 6, n. 1, p. 1–48, 2019.
- SILVA, A. B. et al. Cnn ensembles for nuclei instance segmentation in oed histological images. In: IEEE. **2025 IEEE 38th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2025. p. 369–374.
- _____. Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. **Expert Systems with Applications**, Elsevier, v. 193, p. 116456, 2022. Disponível em: <<https://doi.org/10.1016/j.eswa.2021.116456>>.
- _____. Oralepitheliumdb: A dataset for oral epithelial dysplasia image segmentation and classification. **Journal of Imaging Informatics in Medicine**, Springer Science and Business Media LLC, v. 37, n. 4, p. 1691–1710, fev. 2024. ISSN 2948-2933. Disponível em: <<http://dx.doi.org/10.1007/s10278-024-01041-w>>.
- _____. Cnn ensembles for nuclei segmentation on histological images of oed. In: **2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)**. IEEE, 2023. p. 601–604. Disponível em: <<http://dx.doi.org/10.1109/CBMS58004.2023.00286>>.
- _____. Ensemble of semantic segmentation models for oral epithelial dysplasia images. In: **2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. IEEE, 2024. p. 1–6. Disponível em: <<http://dx.doi.org/10.1109/sibgrapi62404.2024.10716304>>.
- SILVA, M. S. D.; NARAYANAN, B. N.; HARDIE, R. C. A patient-specific algorithm for lung segmentation in chest radiographs. **AI**, MDPI, v. 3, n. 4, p. 931–947, 2022.
- SMITH, J. et al. Biomarkers in dysplasia of the oral cavity: a systematic review. **Oral oncology**, Elsevier, v. 45, n. 8, p. 647–653, 2009.
- SONG, A. H. et al. Artificial intelligence for digital and computational pathology. **Nature Reviews Bioengineering**, v. 1, p. 930–949, 2023.
- SRINIDHI, C. L.; CIGA, O.; MARTEL, A. L. Deep neural network models for computational histopathology: A survey. **Medical Image Analysis**, v. 67, p. 101813, 2021.

- SWAIN, B. R.; CHEOI, K. J.; KO, J. Sam guided task-specific enhanced nuclei segmentation in digital pathology. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. [S.l.], 2024. p. 542–551.
- TAJBAKHSH, N. et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? **IEEE Transactions on Medical Imaging**, v. 35, n. 5, p. 1299–1312, 2016.
- TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **International Conference on Machine Learning (ICML)**. [S.l.: s.n.], 2019. p. 6105–6114.
- TELLEZ, D. et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. **Medical Image Analysis**, v. 58, p. 101544, 2019.
- TU, Z. et al. Maxvit: Multi-axis vision transformer. In: **European Conference on Computer Vision (ECCV)**. [S.l.]: Springer, 2022. p. 459–479.
- VASWANI, A. et al. Attention is all you need. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [S.l.: s.n.], 2017. v. 30, p. 5998–6008.
- WAAL, I. van der. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. **Oral Oncology**, v. 45, n. 4–5, p. 317–323, 2009.
- WARNAKULASURIYA, S.; JOHNSON, N. W.; WAAL, I. van der. Nomenclature and classification of potentially malignant disorders of the oral mucosa. **Journal of Oral Pathology & Medicine**, v. 36, n. 10, p. 575–580, 2007.
- WARNAKULASURIYA, S. et al. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. **Journal of Oral Pathology & Medicine**, Wiley Online Library, v. 37, n. 3, p. 127–133, 2008.
- XIE, E. et al. Segformer: Simple and efficient design for semantic segmentation with transformers. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [S.l.: s.n.], 2021. v. 34, p. 12077–12090.
- YOSINSKI, J. et al. How transferable are features in deep neural networks? In: **Advances in Neural Information Processing Systems (NeurIPS)**. [S.l.: s.n.], 2014. v. 27, p. 3320–3328.
- ZHANG, H. et al. Resnest: Split-attention networks. **arXiv preprint arXiv:2004.08955**, 2020.
- ZHANG, X. et al. A deep learning onion peeling approach to measure oral epithelium layer number. **Cancers**, MDPI AG, v. 15, n. 15, p. 3891, jul. 2023. ISSN 2072-6694. Disponível em: <<http://dx.doi.org/10.3390/cancers15153891>>.
- ZHAO, B. et al. Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. **Medical Image Analysis**, Elsevier, v. 65, p. 101786, 2020.

ZHOU, Z. et al. Unet++: A nested u-net architecture for medical image segmentation. In: **Deep Learning in Medical Image Analysis (DLMIA) Workshop, MICCAI**. [S.l.]: Springer, 2018. p. 3–11.