
**Comitês de Redes Neurais Convolucionais para
a Segmentação de Núcleos em Imagens
Histológicas de Displasias Epiteliais Orais**

Adriano Barbosa Silva



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

Uberlândia
2025

Adriano Barbosa Silva

**Comitês de Redes Neurais Convolucionais para
a Segmentação de Núcleos em Imagens
Histológicas de Displasias Epiteliais Orais**

Tese de doutorado 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 Doutor em Ciência da Computação.

Área de concentração: Ciência de Dados

Orientador: Marcelo Zanchetta do Nascimento

Coorientador: Paulo Rogério de Faria

Uberlândia

2025

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

S586 2025	Silva, Adriano Barbosa, 1991- Comitês de Redes Neurais Convolucionais para a Segmentação de Núcleos em Imagens Histológicas de Displasias Epiteliais Orais [recurso eletrônico] / Adriano Barbosa Silva. - 2025. Orientador: Marcelo Zanchetta do Nascimento. Coorientador: Paulo Rogério de Faria. Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação. Modo de acesso: Internet. Disponível em: http://doi.org/10.14393/ufu.te.2025.180 Inclui bibliografia. Inclui ilustrações. 1. Computação. I. Nascimento, Marcelo Zanchetta do, 1976-, (Orient.). II. Faria, Paulo Rogério de, 1976-, (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:	Tese, 01/2025, PPGCO				
Data:	10 de Março de 2025	Hora de início:	08:58	Hora de encerramento:	12:35
Matrícula do Discente:	12013CCP001				
Nome do Discente:	Adriano Barbosa Silva				
Título do Trabalho:	Comitês de Redes Neurais Convolucionais para a Segmentação de Núcleos em Imagens Histológicas de Displasias Epiteliais Orais				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciência de Dados				
Projeto de Pesquisa de vinculação:	APQ-00727-24 Ensemble Learning e Geometria Fractal Empregados para Classificação de Tumores Odontogênicos Bucais em Imagens Histológicas HE311404/2021-9 - Detecção de Lesões da Cavidade Bucal por Abordagens em Multiresolução e Aprendizagem Profunda 307318/2022-2 - Identificação de alterações nucleares de células de desordens potencialmente malignas e de carcinoma de células escamosas de cavidade bucal através de técnica de multiescala Curvelet e Geometria Fractal				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Paulo Rogério de Faria - FO/UFU (Coorientador), André Ricardo Backes - DC/UFSCar, Thiago Pirola Ribeiro - FACOM/UFU, Cleverton Roberto de Andrade - FOAr/Unesp, Fátima de Lourdes dos Santos Nunes Marques - EACH/USP Leste e Marcelo Zanchetta do Nascimento - FACOM/UFU, orientador do candidato.

Os examinadores participaram desde as seguintes localidades: André Ricardo Backes - São Carlos/SP, Cleverton Roberto de Andrade - Araraquara/SP, Fátima de Lourdes dos Santos Nunes Marques - São Paulo/SP. Os outros membros da banca e o aluno participaram da cidade de Uberlândia.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Marcelo Zanchetta do Nascimento, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

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

Aprovado

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

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 **Thiago Pirola Ribeiro, Professor(a) do Magistério Superior**, em 11/03/2025, às 14:34, 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 **Marcelo Zanchetta do Nascimento, Professor(a) do Magistério Superior**, em 11/03/2025, às 17:53, 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 **André Ricardo Backes, Usuário Externo**, em 12/03/2025, às 09:05, 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 **Paulo Rogerio de Faria, Professor(a) do Magistério Superior**, em 12/03/2025, às 09:22, 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 **Fátima de Lourdes dos Santos Nunes Marques, Usuário Externo**, em 12/03/2025, às 09:25, 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 **Clevertton Roberto de Andrade, Usuário Externo**, em 21/03/2025, às 09:34, 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 **6097907** e o código CRC **13BADE7F**.

Este trabalho é dedicado à Camila, pois, sem o apoio dela, eu não teria chegado até aqui.

Agradecimentos

Agradeço ao meu orientador, Marcelo Zanchetta do Nascimento, pela orientação, pela confiança depositada em meu trabalho e pelo incentivo constante ao longo desta jornada. Meus agradecimentos estendem-se também ao meu coorientador, Paulo Rogério de Faria, pelo conhecimento compartilhado em diversos momentos e pelo suporte durante o desenvolvimento desta pesquisa.

Expresso minha gratidão aos colegas do Laboratório Interdisciplinar de Processamento e Análise de Imagens (LIPAI), pelo companheirismo, ajuda, troca de conhecimento e incentivo que recebi de vocês durante esta jornada. Um agradecimento especial ao Alessandro, ao Leandro e à Thaína, pelo suporte prestado com os experimentos e análises. Levarei para sempre as amizades e os aprendizados que aqui adquiri.

À minha família, pelo apoio incondicional em todas as fases da minha vida. Agradeço aos meus pais, Luci e José, e ao meu irmão, Leonardo, cuja motivação foi muito importante para a realização deste trabalho. Um agradecimento especial à minha esposa, Camila, por seu companheirismo, paciência e por estar sempre ao meu lado, oferecendo apoio nesta etapa tão significativa da minha trajetória.

Aos amigos do Laboratório de Inovações em Tecnologias Digitais, Wanderson, Heidie, Guilherme, Guilherme Brilhante, Michel e Adilmar, pela colaboração, discussões e pelos momentos de descontração que tornaram esta caminhada mais leve. Também sou grato aos amigos que fiz em Uberlândia, que contribuíram para que minha estadia fosse mais acolhedora e agradável.

Agradeço à Faculdade de Computação pelo suporte institucional e pela infraestrutura que viabilizaram este trabalho. Expresso minha gratidão aos professores da FACOM, que contribuíram para minha formação acadêmica e científica, ampliando meus horizontes de conhecimento e auxiliando no desenvolvimento deste projeto. Finalmente, agradeço à Universidade Federal de Uberlândia pelo acesso aos laboratórios, aos espaços físicos e pelo suporte acadêmico necessário à execução desta pesquisa.

Por fim, expresso meu agradecimento pelo apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (Código de Financiamento

001), do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Projetos: 311404/2021-9, 307318/2022-2), da Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG (Projeto: APQ-00727-24) e da Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (Projeto: 2022/03020-1).

Resumo

O diagnóstico precoce de distúrbios potencialmente malignos é a maneira mais confiável para reduzir a incidência de carcinoma na cavidade oral. Condições como leucoplasia e eritroplasia possuem risco de transformação maligna e frequentemente se manifestam em conjunto com Displasias Epiteliais Orais (DEOs), que são alterações no tamanho e formato das estruturas celulares. A segmentação de núcleos celulares é importante para o diagnóstico, possibilitando a extração de informações dessas estruturas sem interferências da região do epitélio. Nos últimos anos, Redes Neurais Convolucionais (CNN) têm recebido destaque devido ao seu desempenho em tarefas de segmentação de núcleos. Entretanto, alcançar alta precisão na segmentação permanece um desafio. Este trabalho propõe uma metodologia baseada em sistemas de comitês de CNN para aprimorar a segmentação de núcleos celulares em imagens de DEOs. Modelos de segmentação foram combinados usando regras como a soma ponderada, buscando minimizar erros individuais dos modelos de CNN e aprimorar o desempenho global da segmentação. Três tipos de comitês foram propostos: um composto por segmentadores semânticos, outro por segmentadores de instâncias e um comitê híbrido, combinando ambas as abordagens. A abordagem proposta foi avaliada em uma nova base de imagens criada para este estudo, contendo casos de DEOs, e em bases públicas. Além disso, foram investigadas técnicas de aumento de dados, RCAug e HistoGAN, para analisar seu impacto na segmentação. A metodologia proposta alcançou resultados expressivos, com acurácia de 95,07% e coeficiente Dice de 0,91 na base de DEOs. Nas bases públicas, o método apresentou coeficiente Dice de 0,89, mostrando-se competitivo em relação aos demais métodos da literatura. Esses resultados evidenciam a robustez da abordagem proposta e seu potencial para ser utilizada como ferramenta de apoio ao diagnóstico em casos de DEO, contribuindo para a identificação precoce de distúrbios celulares.

Palavras-chave: Displasia Epitelial Oral. Processamento Digital de Imagens. Segmentação de Núcleos. Redes Neurais Convolucionais. Comitês de Segmentadores.

Abstract

Early diagnosis of potentially malignant disorders is the most reliable way to reduce the incidence of carcinoma in the oral cavity. Conditions such as leukoplakia and erythroplakia carry a risk of malignant transformation and often manifest alongside oral epithelial dysplasias (OEDs), which are alterations in the size and shape of cellular structures. The segmentation of cell nuclei is essential for diagnosis, allowing the extraction of information from these structures without interference from the epithelial region. In recent years, convolutional neural networks (CNN) have gained prominence due to their performance in nucleus segmentation tasks. However, achieving high segmentation accuracy remains a challenge. This work proposes a methodology based on ensemble neural network systems to improve the segmentation of cell nuclei in OED images. Segmentation models were combined using rules such as weighted sum, aiming to minimize individual errors of CNN models and enhance overall segmentation performance. Three types of ensembles were proposed: one composed of semantic segmenters, another using instance segmenters, and a hybrid ensemble combining both approaches. The proposed approach was evaluated on a newly created dataset for this study, containing OED cases, as well as on public datasets. Additionally, data augmentation techniques, RCAug and HistoGAN, were investigated to analyze their impact on segmentation. The methodology achieved remarkable results, with an accuracy of 95.07% and a Dice coefficient of 0.91 on the OED dataset. On public datasets, the method obtained a Dice coefficient of 0.89, demonstrating competitiveness compared to other methods in the literature. These results highlight the robustness of the proposed approach and its potential as a diagnostic support tool for OED cases, contributing to the early identification of cellular disorders.

Keywords: Oral Epithelial Dysplasia. Digital Image Processing. Nuclei Segmentation. Convolutional Neural Networks. Segmentation Ensemble.

Lista de ilustrações

Figura 1	– Exemplos de tecidos histológicos da cavidade bucal: (a) tecido saudável, (b) DEO leve, (c) DEO moderada e (d) DEO severa. As setas vermelhas indicam a proporção do epitélio que apresenta núcleos com alterações morfológicas.	32
Figura 2	– Etapas comumente empregadas em tarefas de análise computacional de imagens histológicas.	34
Figura 3	– Exemplos dos diferentes tipos de segmentação: (a) imagem original, (b) segmentação semântica (rotulação por classes de <i>pixels</i>), (c) segmentação de instâncias (rotulação por objetos individuais) e (d) segmentação panóptica (rotulação por classes de <i>pixels</i> e por instâncias de objetos). Imagem adaptada de (KIRILLOV et al., 2019)).	37
Figura 4	– Exemplos de neurônios: (a) neurônio biológico (fonte: (SILVA, 2019)) e (b) neurônio artificial (adaptada de (HAYKIN et al., 2009)).	38
Figura 5	– Exemplo de uma operação de convolução. Os elementos da imagem <i>a</i> são multiplicados pelos elementos do filtro <i>b</i> e os resultados são somados, gerando o mapa resultante.	39
Figura 6	– Exemplos de operações de <i>pooling</i> : (a) <i>average pooling</i> e (b) <i>max pooling</i> (imagens retiradas de: (SILVA, 2019)).	40
Figura 7	– Exemplo de funcionamento de um comitê com três modelos de DL. Os pesos e limiares de decisão dos modelos são agregados através de uma operação de somatório, gerando um novo limiar para o comitê. Imagem adaptada de (POLIKAR, 2012).	41
Figura 8	– Exemplos de diferentes composições para comitês de modelos de DL. Imagem adaptada de Mohammed e Kora (2023).	43
Figura 9	– Fluxograma com as principais etapas para a segmentação de núcleos celulares em tecidos histológicos de DEO.	53
Figura 10	– Processo de criação do padrão-ouro de segmentação: (a) anotação manual, (b) revisões apontadas pelo patologista e (c) anotações revisadas.	54

Figura 11 – Exemplos de tecidos histológicos utilizados na construção da base de dados: (a) tecido saudável, (b) DEO leve, (c) DEO moderada, (d) DEO severa, (e-h) máscaras do padrão ouro para cada classe das lâminas digitalizadas.	55
Figura 12 – Exemplos de tecidos histológicos da literatura utilizados para avaliação do método proposto: (a) MoNuSeg 2018, (b) CryoNuSEg, (c) GlaS, (d-f) máscaras do padrão ouro.	55
Figura 13 – Arquitetura da CNN <i>ResNet50</i> empregada como <i>backbone</i> para a segmentação de núcleos. Fonte: (SILVA, 2019).	58
Figura 14 – Arquitetura da CNN MobileNetV3 empregada como <i>backbone</i> para a segmentação de núcleos. Imagem adaptada de (HOWARD et al., 2019).	58
Figura 15 – Arquitetura do modelo U-Net usado para a segmentação de núcleos. Imagem adaptada de (RONNEBERGER; FISCHER; BROX, 2015).	59
Figura 16 – Arquitetura do modelo FPN usado para a segmentação de núcleos. Imagem adaptada de (SILVA, 2019).	60
Figura 17 – Arquitetura do modelo PSPNet usado para a segmentação de núcleos. Imagem adaptada de (ZHAO et al., 2017).	62
Figura 18 – Arquitetura da rede Mask R-CNN usada na segmentação de núcleos. Imagem adaptada de (SILVA, 2019).	63
Figura 19 – Exemplos de polígonos gerados pelo modelo StarDist para determinar a área dos núcleos celulares. Imagem elaborada pelo autor.	64
Figura 20 – Estrutura da arquitetura HoVer-Net usado para segmentação via instâncias. Imagem adaptada de (GRAHAM et al., 2019).	65
Figura 21 – Arquitetura do modelo SOLOv2 usado para a segmentação de núcleos. Imagem adaptada de (WANG et al., 2020).	66
Figura 22 – Etapas empregadas para o pós processamento: (a) máscara inicial, (b) máscara após a operação de dilatação, (c) máscara após a operação de preenchimento de regiões e (d) máscara resultante após as operações de erosão e remoção de artefatos. As setas em vermelho indicam as alterações resultantes das operações.	68
Figura 23 – Curvas de treinamento e validação para as métricas de acurácia e taxa de perda em diferentes configurações de dados: (a) e (b) <i>baseline</i> , (c) e (d) HistoGAN, e (e) e (f) RCAug. A coluna da esquerda apresenta os resultados para o modelo de segmentação FPN-ResNet50 e a coluna da direita exhibe as curvas correspondentes ao modelo Mask R-CNN.	77

Figura 24 – Comparação visual entre o padrão ouro e o modelo FPN-ResNet50 para um tecido com DEO leve: (a) imagem original, (b) padrão ouro, (c) resultado da segmentação, (d) máscara gerada pelo modelo. Regiões na cor vermelha representam FN e regiões em verde indicam FP. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.	78
Figura 25 – Análise visual entre o modelo Mask R-CNN para um tecido com DEO leve: (a) imagem original, (b) padrão ouro, (c) resultado da segmentação, (d) máscara gerada pelo modelo. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.	79
Figura 26 – Comparação visual entre o padrão ouro e as abordagens de segmentação semântica em um tecido com DEO leve: (a) imagem original, (b) padrão ouro marcado pelo especialista, (c-e) regras de votação. Regiões na cor vermelha representam FNs e regiões em verde indicam FPs. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.	83
Figura 27 – Comparação visual entre o padrão ouro e as abordagens de comitês para segmentação de instâncias em um tecido com DEO leve: (a) imagem original, (b) padrão ouro marcado pelo especialista, (c-e) regras de votação. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.	85
Figura 28 – Comparação visual entre o padrão ouro e as abordagens de comitês híbridos em um tecido com DEO leve: (a) padrão ouro semântico, (b) comitê semântico com soma ponderada pelo D_C , (c) Comitê híbrido semântico, (d) padrão ouro de instâncias, (e) comitê via instâncias com soma ponderada pelo D_C , (f) comitê híbrido via instâncias. Regiões na cor vermelha representam FN e regiões em verde indicam FP. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.	88

Lista de tabelas

Tabela 1 –	Combinações de palavras-chave utilizadas na busca bibliográfica. . . .	48
Tabela 2 –	Configurações de treinamento dos modelos analisados.	66
Tabela 3 –	Avaliação dos modelos de CNN para segmentação semântica de núcleos em tecidos histológicos.	73
Tabela 4 –	Avaliação dos modelos de CNN para segmentação via instâncias de núcleos celulares.	75
Tabela 5 –	Avaliação dos resultados com o modelo FPN-ResNet50 e a técnica de validação cruzada.	80
Tabela 6 –	Avaliação do modelo Mask R-CNN com a técnica de validação cruzada.	80
Tabela 7 –	Avaliação dos sistemas de votação de segmentadores semânticos. . . .	82
Tabela 8 –	Avaliação dos sistemas de votação de segmentadores via instâncias. . .	84
Tabela 9 –	Desempenho das estratégias de segmentação em diferentes classes de OED.	86
Tabela 10 –	Avaliação dos sistemas de comitês híbridos.	87
Tabela 11 –	Comparação entre os sistemas propostos e métodos de segmentação na literatura (D_C).	89

Lista de siglas

A_{CC} *Acurácia*

CAD *Computer Aided Diagnosis*

CNN *Convolutional Neural Network*

DEO *Displasia Epitelial Oral*

DL *Deep Learning*

D_C *Coeficiente Dice*

E_S *Especificidade*

GAN *Generative Adversarial Network*

ROI *Region of Interest*

RNA *Redes Neurais Artificiais*

RPN *Region Proposal Network*

S_E *Sensibilidad*

Sumário

1	INTRODUÇÃO	23
1.1	Motivação	25
1.2	Objetivos e Desafios da Pesquisa	26
1.3	Hipóteses	27
1.4	Contribuições	27
1.5	Organização desta Tese	28
2	FUNDAMENTAÇÃO TEÓRICA	29
2.1	Câncer	29
2.2	Displasias Epiteliais Orais	30
2.3	Preparação e Digitalização de Tecidos Histológicos	32
2.4	Sistemas Computacionais para Apoio ao Diagnóstico	33
2.5	Segmentação	35
2.6	Redes Neurais Artificiais	37
2.7	Comitês de Aprendizagem Profunda	40
2.8	Métricas de Avaliação para Segmentação	44
3	ESTADO DA ARTE	47
3.1	Aprendizagem Profunda para Segmentação de Núcleos em Te- cidos de DEO	47
3.2	Comitês de Aprendizagem Profunda Aplicados em Imagens Médicas	49
3.3	Considerações	51
4	METODOLOGIA	53
4.1	Bancos de imagens	53
4.2	Aumento de dados	55
4.2.1	HistoGAN	56

4.2.2	RCAug	56
4.3	Modelos de redes para segmentação de núcleos celulares	57
4.3.1	Arquiteturas de redes para extração de características	57
4.3.2	Arquiteturas de rede para segmentação semântica	59
4.3.3	Arquiteturas de rede para segmentação via instâncias	62
4.3.4	Treinamento dos modelos de segmentação	66
4.4	Pós-processamento em modelos de rede para segmentação . . .	67
4.5	Sistemas de comitês de segmentadores	69
5	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	73
5.1	Análise da segmentação com os modelos baseados em aprendi- zagem profunda	73
5.2	Estudo de desempenho entre modelos de comitês	80
5.3	Comparação entre os modelos propostos e métodos da literatura	88
6	CONCLUSÕES	91
6.1	Trabalhos Futuros	92
6.2	Contribuições em Produção Bibliográfica	93
	REFERÊNCIAS	97

Introdução

Segundo o Instituto Nacional do Câncer (INCA), o câncer é a segunda causa mais comum de óbito no mundo, com uma crescente taxa de incidência. Em estimativas realizadas em 2023, projetou-se que até 2025 ocorreriam cerca de 20 milhões de novos casos de câncer globalmente (SANTOS, 2023). No Brasil, para o triênio de 2023 a 2025, estimou-se aproximadamente 704 mil novos diagnósticos de câncer, sendo 15.100 casos de câncer da cavidade oral. Esse tipo de câncer ocupa o oitavo lugar em termos de mortalidade global (SANTOS, 2023). Esses números ressaltam a importância da realização de diagnósticos precoces como ferramenta para melhorar as chances de cura e o prognóstico dos pacientes acometidos.

Uma das estratégias para reduzir a incidência de carcinoma na cavidade bucal é identificar lesões com potencial de transformação maligna. Dentre essas lesões, se destacam a leucoplasia e a eritroplasia, que frequentemente exibem Displasia Epitelial Oral (DEO) em diferentes graus. A eritroplasia, por exemplo, é frequentemente associada a DEOs graves (TILAKARATNE et al., 2019). Microscopicamente, as DEOs são caracterizadas por alterações no tamanho e formato das estruturas celulares, se manifestando em conjunto com outras lesões e apresentando potencial para transformação maligna (KADEMANI, 2019). Estudos apontam que leucoplasias com DEOs severas estão associadas a uma maior probabilidade de transformação maligna e uma menor taxa de sobrevida (SHEARSTON et al., 2019; JAYASOORIYA et al., 2020). Nesse contexto, a histopatologia possibilita a identificação e gradação dessas alterações, possibilitando uma melhor avaliação do risco de malignização (KADEMANI, 2019).

No processo de diagnóstico de DEOs, a análise histopatológica de lâminas de tecido obtidas de lesões suspeitas é fundamental (KUMAR; ASTER; ABBAS, 2010). A caracterização dos núcleos celulares é importante, pois as alterações nucleares estão relacionadas ao grau de diferenciação celular e à severidade da DEO (MÜLLER, 2018). No entanto, a subjetividade da análise microscópica pode introduzir variabilidade no diagnóstico, influenciada por fatores como a carga de trabalho e experiência (KADEMANI, 2019). Isso ocorre porque o diagnóstico se baseia no tamanho da lesão e no nível de severidade das

alterações nucleares. Como essas características não são relacionadas, é possível que um especialista atribua o mesmo grau de DEO para lesões com tamanhos diferentes (KUMAR; ABBAS; ASTER, 2013; KADEMANI, 2019). Estratégias como a dupla leitura têm se mostrado eficazes para minimizar essas discrepâncias (BRENNAN et al., 2019; ZELST et al., 2020), mas seu uso em larga escala pode elevar significativamente os custos operacionais, tornando-se inviável em muitas rotinas diagnósticas.

Com o avanço tecnológico, os Sistemas Computacionais de Apoio ao Diagnóstico (do inglês, *Computer Aided Diagnosis* (CAD)) têm se consolidado como ferramentas para melhorar a eficiência e a precisão das análises histopatológicas. Esses sistemas vêm sendo amplamente adotados como apoio ao especialista, especialmente em estratégias de dupla leitura, contribuindo para reduzir a subjetividade associada à avaliação manual (CHAN; HADJIISKI; SAMALA, 2020). Algoritmos de processamento de imagens e aprendizado profundo (do inglês, *Deep Learning* (DL)) são amplamente empregados nesses sistemas, permitindo a identificação de padrões complexos e a quantificação de elementos morfológicos (KUMAR et al., 2019). Um CAD típico integra etapas como pré-processamento, segmentação, extração de características e classificação, sendo que cada uma delas desempenha um papel importante para a análise e a interpretação das imagens (YANASE; TRIANTAPHYLLOU, 2019).

O processo de segmentação consiste em dividir a imagem em regiões, seguindo um conjunto de critérios. Essa etapa permite separar os objetos de interesse, como núcleos celulares, de outros objetos e regiões, possibilitando que a análise manual feita pelo patologista ou outras etapas de um sistema CAD possam ser realizadas sem a interferência das demais regiões (YANG et al., 2017). Com a imagem segmentada, informações numéricas podem ser extraídas para representar as regiões, transformando os dados em informações importantes para os algoritmos de classificação (HUMEAU-HEURTIER, 2019). Essas características variam de informações morfológicas à propriedades de textura (WEI; GAN; JI, 2017). Dessa forma, a precisão da segmentação influencia na obtenção dessas características (WEI; GAN; JI, 2017). No entanto, imagens histológicas apresentam desafios significativos para a segmentação, como variações na aparência celular, bordas mal definidas e células sobrepostas ou aglomeradas (XIE et al., 2018). Em lesões de DEO, essas dificuldades se agravam devido ao aspecto visual irregular dos núcleos celulares, que pode variar mesmo dentro de um mesmo grau de severidade (IRSHAD et al., 2014). Além disso, ruídos podem ser inseridos durante a preparação das lâminas histológicas, tornando necessários métodos mais robustos para tratar esses fatores (BASU et al., 2023).

Nos últimos anos, as métodos de DL baseados em Redes Neurais Convolucionais (do inglês - *Convolutional Neural Network* (CNN)) e redes adversárias generativas (do inglês, *Generative Adversarial Network* (GAN)) se destacaram como alternativas promissoras para superar os desafios das abordagens tradicionais (AGGARWAL et al., 2018). Dentro desse contexto, métodos de segmentação semântica, via instâncias e panóptica têm

sido amplamente explorados em aplicações que envolvem imagens histológicas (MINAEE et al., 2021). Embora estudos mostrem que modelos de DL apresentem resultados relevantes para segmentação de estruturas histológicas, como visto em Kumar et al. (2019), Martino et al. (2020), Rasool, Fraz e Javed (2021), diversos desafios existem para o uso dessas técnicas para a etapa de segmentação, como: (1) baixa taxa de convergência; (2) pouca flexibilidade para tarefas complexas; (3) quantidade massiva de informações armazenadas. No contexto de tecidos de DEOs, investigar e propor modelos que permitam a separação de forma mais efetiva os núcleos para um processo de análise ainda é um tópico em aberto principalmente em tecidos de DEO com diferentes graus (leve, moderada e severa) (SHEPHARD et al., 2021; BASU et al., 2023).

1.1 Motivação

O diagnóstico precoce e a detecção precisa do grau de DEO são importantes para que o paciente receba o tratamento adequado antes que a lesão evolua para um carcinoma invasivo (KUMAR; ABBAS; ASTER, 2013). Segundo o estudo apresentado por Tilakaratne et al. (2019), um sistema eficaz deve ser capaz de identificar as estruturas celulares em imagens de DEO, possibilitando diferenciar de forma precisa os graus de severidade, e minimizando variações intra e interobservador. Apesar do avanço das técnicas para o diagnóstico, os estudos mostram que o domínio ainda apresenta dificuldade na análise dessas lesões (KADEMANI, 2019; TILAKARATNE et al., 2019). Esse cenário aponta para a necessidade de métodos mais precisos, robustos e acessíveis, que possam ser integrados à prática clínica com confiabilidade.

Uma forma de contribuir para diminuir o número de casos de carcinoma na cavidade bucal é a realização de diagnósticos em lesões com potencial de transformação maligna. Exemplos dessas lesões são: a leucoplasia e a eritroplasia, que exibem microscopicamente DEOs em diferentes graus, sendo a eritroplasia uma lesão que, na maioria das vezes, apresenta DEOs graves (TILAKARATNE et al., 2019). As DEOs são alterações no tamanho e formato das estruturas celulares que se manifestam em conjunto com as outras lesões, apresentando potencial de transformação maligna (KADEMANI, 2019). Há relatos de que as leucoplasias podem apresentar DEOs, sendo que leucoplasias orais com DEOs severas são relacionadas a transformação maligna, podendo haver menor taxa de sobrevida (SHEARSTON et al., 2019; JAYASOORIYA et al., 2020). Por esse motivo, a análise por meio da histopatologia pode contribuir para a identificação e gradação de DEOs, permitindo saber quais delas apresentam riscos de transformação maligna (KADEMANI, 2019).

O desenvolvimento de ferramentas computacionais para diagnóstico histopatológico tem o potencial de preencher essa lacuna ao fornecer análises mais objetivas e reprodutíveis. Dentre as técnicas emergentes, os Sistemas CAD baseados DL têm mostrado resultados promissores, pois são capazes de lidar com grandes volumes de dados e identifi-

car padrões complexos que podem não ser perceptíveis ao especialista (SILVA, 2019). Um dos principais desafios para a adoção de modelos de DL em sistemas CAD está relacionado à variação na aparência das lesões histopatológicas e à necessidade de segmentações precisas e consistentes das estruturas celulares. Nesse contexto, comitês de CNNs têm se destacado como uma solução promissora, devido à sua capacidade de combinar os pontos fortes de múltiplos modelos, reduzindo erros e melhorando a generalização (GHOSH et al., 2021; KATO; HOTTA, 2021). Embora avanços significativos tenham sido alcançados em técnicas de comitês para segmentação, a maior parte dos estudos se concentra em abordagens semânticas, com poucas investigações sobre segmentação via instâncias (ADEL et al., 2018; KATO; HOTTA, 2021). A ausência de trabalhos que explorem comitês de segmentadores para segmentação via instâncias representa uma lacuna importante na literatura, pois essas técnicas permitem isolar objetos individuais em meio a aglomerações celulares (XIE et al., 2018). A exploração dessa abordagem pode proporcionar aprimoramentos significativos na análise de lesões de DEO, auxiliando diretamente no diagnóstico e prognóstico dos pacientes.

Dessa forma, é destacada a necessidade de desenvolver soluções que combinem modelos de CNN de maneira a explorar suas vantagens, especialmente em aplicações que demandam alta precisão, como é o caso da histopatologia. Essas abordagens tem o potencial para aprimorar a confiabilidade de um sistema CAD, tornando a ferramenta ainda mais eficaz no auxílio ao patologista. Este trabalho propõe explorar e aprimorar essas abordagens, investigando o uso de comitês de redes neurais para segmentação de estruturas histológicas em imagens de lesões de DEO, com o objetivo de oferecer uma solução inovadora e prática para os desafios descritos. Com isso, torna-se necessário investigar novas abordagens com o uso de comitês de CNNs para segmentação semântica, via instâncias ou híbrida dos modelos para núcleos celulares em imagens de lesões de DEO, com o objetivo de auxiliar o processo de tomada de decisão dos profissionais da área (GU et al., 2022; JI et al., 2023).

1.2 Objetivos e Desafios da Pesquisa

O objetivo principal deste trabalho é o desenvolvimento de algoritmos para a segmentação de núcleos celulares em imagens histológicas de DEOs, utilizando modelos CNNs e abordagens de comitês. Essa proposta visa auxiliar na análise de DEOs de forma padronizada e robusta. Para atingir este objetivo, as seguintes etapas foram investigadas:

- construção de um banco de imagens de tecidos histológicos com tecidos saudáveis e com os diferentes graus de DEO, incluindo anotações realizadas por um especialista para validação das abordagens de segmentação semântica e via instâncias;

- ❑ avaliação de diferentes arquiteturas de CNN para segmentação de núcleos em imagens histológicas de DEO;
- ❑ desenvolvimento de comitês de segmentadores baseados em DL para melhorar a precisão e reduzir a variabilidades nos resultados, englobando diferentes paradigmas de segmentação;
- ❑ estudo do impacto de técnicas de aumento de dados para mitigar a escassez de imagens rotuladas e melhorar o desempenho das abordagens baseadas em DL;
- ❑ validação das abordagens propostas em relação a modelos individuais e ao padrão ouro definido pelo especialista, incluindo experimentos em bases públicas para avaliar a generalização da metodologia.

1.3 Hipóteses

Considerando os desafios e limitações relacionadas à segmentação de imagens de DEO, o presente trabalho foi desenvolvido considerando as seguintes hipóteses:

- ❑ comitês de segmentadores baseados em CNNs apresentam melhor desempenho do que modelos individuais em termos de acurácia e estabilidade dos resultados;
- ❑ métodos de segmentação via instâncias são mais eficazes para a identificação de núcleos em comparação à segmentação semântica, especialmente em casos de DEO;
- ❑ uma composição híbrida, combinando modelos de segmentação semânticos e via instâncias, permite o aprimoramento dos resultados, superando abordagens exclusivamente semânticas ou via instâncias;
- ❑ estratégias de aumento de dados melhoram o desempenho dos modelos CNNs e comitês, reduzindo o impacto relacionado à escassez de imagens rotuladas e aumentando a generalização em bases diversas.

1.4 Contribuições

As principais contribuições deste trabalho são:

- ❑ construção de um banco de imagens histológicas de DEO com anotações semânticas e via instâncias realizadas por especialistas, disponibilizado publicamente para a comunidade científica. Esse material pode ser obtido em: <<https://github.com/LIPAIGroup/OralEpitheliumDB>>;

- ❑ desenvolvimento de estratégias de comitês baseados em DL para a segmentação de núcleos celulares, destacando o impacto da segmentação semântica e via instâncias nos diferentes graus de DEO;
- ❑ desenvolvimento de uma abordagem inovadora para comitês híbridos, combinando os resultados de segmentação de modelos por instâncias e semânticos;
- ❑ proposta e aplicação de técnicas de pós-processamento para refinar os resultados de segmentação dos modelos de DL, favorecendo a consistência morfológica dos núcleos;
- ❑ investigação do impacto de técnicas de aumento de dados no desempenho dos modelos e dos comitês, considerando a escassez de imagens rotuladas disponíveis;
- ❑ análise de desempenho entre modelos de CNN individuais e comitês, utilizando métricas de avaliação e explorando os diferentes graus de DEO;
- ❑ avaliação das estratégias propostas para comitês em bases de imagens públicas, com validação dos resultados obtidos em relação a métodos presentes na literatura.

1.5 Organização desta Tese

Este capítulo apresentou brevemente os objetivos da metodologia proposta e a motivação para o desenvolvimento deste trabalho. Os próximos capítulos deste estudo são organizados da seguinte forma:

- ❑ **Capítulo 2 - Fundamentação Teórica:** discute os conceitos de lesões pré-câncer, sistemas CAD, segmentação, redes neurais e comitês, fornecendo o embasamento necessário para este estudo;
- ❑ **Capítulo 3 - Estado da Arte:** explora os trabalhos relacionados à segmentação de núcleos em imagens histológicas, com ênfase no uso de CNNs e comitês;
- ❑ **Capítulo 4 - Metodologia:** descreve o processo de criação do banco de imagens utilizado, os modelos desenvolvidos para a segmentação, o pós processamento e o processo de criação dos comitês;
- ❑ **Capítulo 5 - Experimentos e Análise dos Resultados:** apresenta os resultados obtidos em cada etapa dos experimentos, incluindo análises quantitativas e qualitativas em relação ao padrão ouro e bases públicas;
- ❑ **Capítulo 6 - Conclusões:** apresenta as conclusões do trabalho, destacando as contribuições realizadas, a produção bibliográfica gerada e perspectivas para trabalhos futuros.

Fundamentação Teórica

Neste capítulo são apresentados os conceitos fundamentais que sustentam as teorias e metodologias aplicadas neste trabalho. Inicialmente, são abordados os aspectos relacionados à histologia de câncer e DEOs, destacando a importância do diagnóstico precoce e o papel da análise histológica no entendimento dessas condições. Em seguida, são introduzidos os processos de preparação e digitalização de tecidos histológicos, essenciais para garantir a qualidade das imagens utilizadas nas etapas posteriores de análise. Além disso, o capítulo explora os avanços em sistemas computacionais de apoio ao diagnóstico, com foco no uso de CNNs para segmentação de imagens médicas, e os comitês. Essa base teórica é essencial para contextualizar as lacunas e inovações propostas neste estudo, particularmente no que diz respeito ao uso de comitês para segmentação em imagens histológicas.

2.1 Câncer

Lesões tumorais são caracterizadas pelo crescimento anormal de células que formam massas ou proliferações de tecido, conhecidas como tumores. Essas lesões podem ser classificadas como benignas ou malignas, dependendo de suas características biológicas e comportamento clínico (KUMAR; ASTER; ABBAS, 2010). Tumores benignos apresentam crescimento limitado, não invadem tecidos adjacentes e, geralmente, não ameaçam a vida do paciente (KUMAR; ABBAS; ASTER, 2013). Por outro lado, tumores malignos, conhecidos como câncer, possuem comportamento agressivo, com capacidade de invadir tecidos próximos e se disseminar para outras partes do corpo através de metástases (SANTOS, 2023).

O câncer é uma desordem genética resultante de mutações no DNA, que podem ocorrer espontaneamente ou devido à exposição a fatores ambientais, como tabagismo, consumo de álcool, infecções virais e radiação (KADEMANI, 2019). O câncer da cavidade oral é o sexto tipo mais comum, com uma taxa de incidência crescente, sendo frequentemente diagnosticado em estágios avançados (KADEMANI, 2019). A taxa de sobrevivência para

este tipo de câncer é de cerca de 50% em cinco anos após o diagnóstico (SANTOS, 2023), destacando a necessidade de estratégias mais eficazes para a detecção precoce e o tratamento.

No câncer, células mutantes obtêm vantagens seletivas em crescimento e sobrevivência por meio do processo de seleção natural. Elas podem se proliferar rapidamente, substituindo células normais e dominando a população celular (KUMAR; ABBAS; ASTER, 2013). Essas anormalidades malignas geralmente apresentam características morfológicas distintas, como variações no tamanho, forma e coloração das células em relação às saudáveis (KADEMANI, 2019). Além disso, o crescimento tumoral é frequentemente proporcional ao grau de alterações sofridas pelas células, resultando em lesões invasivas que infiltram tecidos adjacentes e destroem estruturas próximas em um curto intervalo de tempo. Por essa razão, intervenções cirúrgicas geralmente incluem a remoção de margens amplas de tecido saudável ao redor da lesão, visando minimizar o risco de recidiva tumoral (KUMAR; ABBAS; ASTER, 2013).

Lesões benignas, por outro lado, apresentam características não invasivas, taxas de crescimento reduzidas e, geralmente, permanecem localizadas. No entanto, algumas lesões benignas podem evoluir para câncer e são classificadas como lesões pré-cancerosas (WARNAKULASURIYA et al., 2008). Embora a maioria das lesões pré-cancerosas não evolua para malignidade, elas criam um ambiente propício para a proliferação celular anormal e inflamação, fatores que aumentam o risco de transformação maligna. Essas lesões podem permanecer estáveis por muitos anos, mas, em alguns casos, estímulos externos, como traumas, infecções ou tabagismo, podem induzi-las a transformações malignas (KUMAR; ASTER; ABBAS, 2010).

Exemplos de lesões pré-cancerosas incluem leucoplasia, eritroplasia e fibrose submucosa oral (TILAKARATNE et al., 2019). Em muitos casos, essas lesões podem se manifestar em conjunto com DEOs, sendo um fator importante na progressão para o câncer. Estudos mostram que DEOs severas, especialmente quando associadas a lesões como leucoplasia, têm um maior potencial de transformação maligna (KADEMANI, 2019). Dessa forma, a identificação precoce e a caracterização precisa da severidade dessas lesões são importantes para a saúde do paciente, permitindo intervenções terapêuticas adequadas e mais eficazes (KUMAR; ASTER; ABBAS, 2010; SANTOS, 2023).

2.2 Displasias Epiteliais Orais

Em histopatologia, a DEO refere-se a um crescimento desordenado caracterizado por anormalidades na maturação celular e alterações citológicas. Quando essas alterações ocorrem no epitélio da cavidade oral, são classificadas DEOs e consideradas lesões pré-cancerosas (KADEMANI, 2019). As DEOs apresentam alterações celulares, como crescimento anormal, deformações nucleares, hiper Cromasia e perda de uniformidade celu-

lar (KUMAR; ASTER; ABBAS, 2010). Além disso, essas lesões frequentemente exibem mudanças na histoarquitetura do tecido epitelial. Por exemplo, a progressão normal das células basais para células superficiais pode ser interrompida, sendo substituída por células basófilas em toda a extensão do tecido (KUMAR; ABBAS; ASTER, 2013).

O diagnóstico de DEOs segue a classificação proposta pela Organização Mundial da Saúde (OMS), em 2005, que analisa a intensidade das alterações celulares e a extensão do tecido epitelial comprometido (KADEMANI, 2019). As DEOs leves apresentam alterações discretas, limitadas a até um terço da espessura do epitélio. As DEOs severas comprometem mais de dois terços do epitélio, com alterações celulares intensas (WARNAKULASURIYA et al., 2008). O grupo intermediário, as lesões moderadas, afetam entre um e dois terços do epitélio e demandam avaliação detalhada, pois podem ser classificadas como leves ou severas dependendo da intensidade das alterações (SMITH et al., 2009). Na Figura 1, observa-se exemplos de tecidos saudáveis e diferentes níveis de DEO. É possível observar, pela Figura 1(b), que a DEO leve exibe alterações nucleares na camada basal, como indicado pela seta na cor vermelha, enquanto as Figuras 1(c) e 1(d) mostram o comprometimento progressivo em estágios mais avançados.

As DEOs apresentam uma probabilidade de 6% a 36% de evolução para câncer (SMITH et al., 2009). Fatores como tabaco e álcool aumentam significativamente o risco de malignização. O consumo de tabaco está diretamente associado ao desenvolvimento de DEOs, com fumantes ativos apresentando maior probabilidade de alterações displásicas em relação a não fumantes (PEREIRA et al., 2011; DOST et al., 2014). O risco aumenta ainda mais quando o tabaco é combinado com o consumo de álcool, atingindo cerca de 65% de chances de evolução maligna para DEOs moderadas e severas (SMITH et al., 2009; SANTOS, 2023). O assoalho bucal é a região mais propensa ao desenvolvimento dessas lesões, devido à acumulação de produtos derivados do tabaco e álcool na saliva, combinada à fina espessura do epitélio (KADEMANI, 2019). Sem intervenção, essas lesões podem evoluir para carcinoma de células escamosas, que possui uma taxa de mortalidade de aproximadamente 50%. Por isso, a detecção precoce e a caracterização da severidade das DEOs são importantes para melhorar o prognóstico do paciente (KUMAR; ASTER; ABBAS, 2010; SANTOS, 2023).

Embora critérios como extensão do epitélio afetado e intensidade das alterações nucleares sejam essenciais para a classificação de DEOs, essas características podem ocorrer de forma independente, introduzindo subjetividade na análise (SMITH et al., 2009). Por exemplo, duas lesões com igual comprometimento da espessura epitelial podem diferir em intensidade de alterações nucleares, resultando em diagnósticos variados dependendo do critério utilizado (WARNAKULASURIYA et al., 2008; SMITH et al., 2009).

A subjetividade no diagnóstico contribui para uma taxa de divergência inter-observador de 19% a 38% (MÜLLER, 2018), dificultando a consistência na identificação dessas lesões. Embora análises conjuntas entre especialistas possam reduzir essa divergência, essa abor-

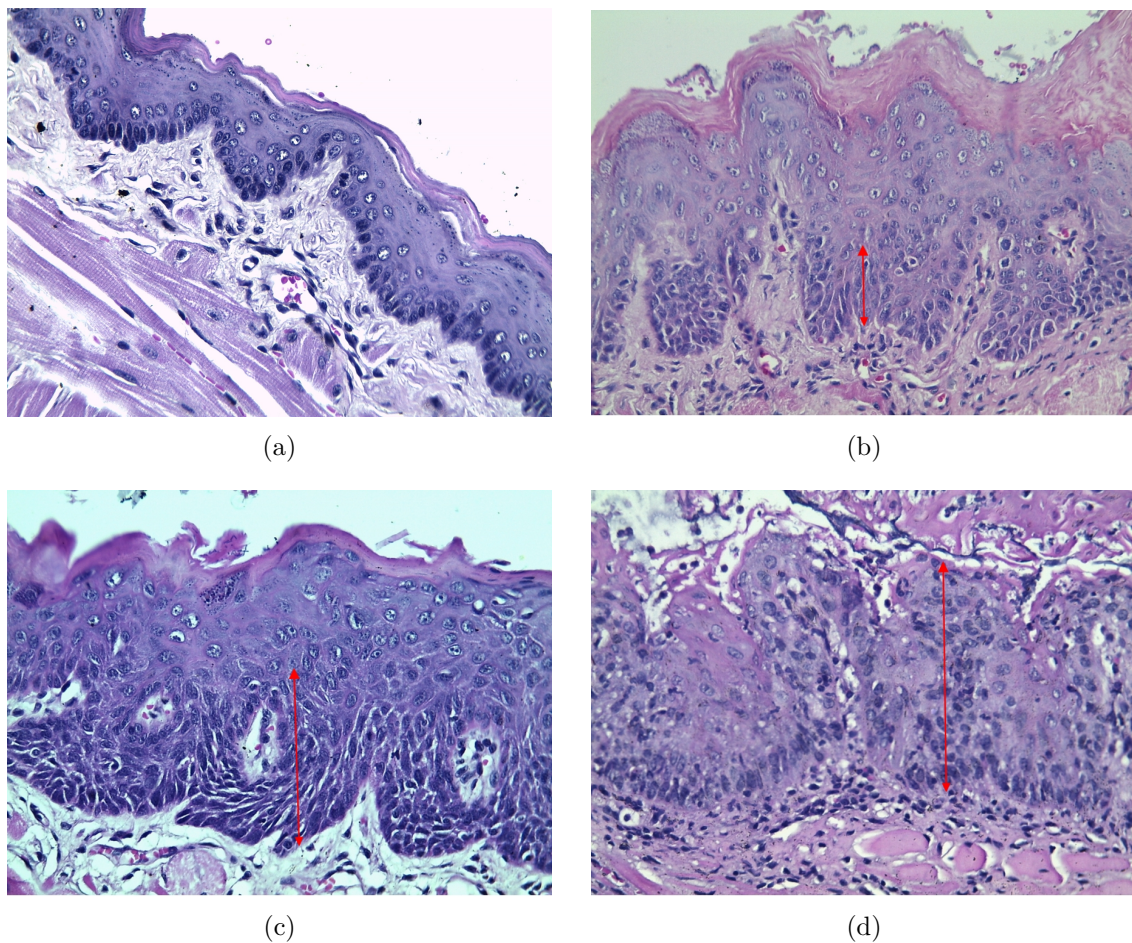


Figura 1 – Exemplos de tecidos histológicos da cavidade bucal: (a) tecido saudável, (b) DEO leve, (c) DEO moderada e (d) DEO severa. As setas vermelhas indicam a proporção do epitélio que apresenta núcleos com alterações morfológicas.

dagem é mais custosa, pois exige maior tempo dos profissionais (SPEIGHT et al., 2015; MÜLLER, 2018). Apesar dessas dificuldades, a análise histológica permanece como uma das metodologias mais eficazes para caracterizar os níveis de DEO (FONSECA-SILVA et al., 2016).

2.3 Preparação e Digitalização de Tecidos Histológicos

A histologia é o estudo dos tecidos e células, suas funções e como se organizam para formar os órgãos. Além disso, analisa dimensões celulares e suas estruturas internas com o auxílio de microscópios (GARTNER, 2007; JUNQUEIRA; CARNEIRO, 2013). A análise histológica é essencial em áreas como a oncologia, permitindo identificar alterações celulares e morfológicas associadas a lesões pré-cancerosas e tumores malignos (KADEMANI, 2019). Para que essa análise seja realizada, é necessário transformar o material biológico em cortes histológicos, que são secções que mantêm o aspecto natural (*in vivo*)

do tecido (GARTNER, 2007; JUNQUEIRA; CARNEIRO, 2013).

O processo de obtenção das lâminas histológicas envolve etapas bem definidas, como descritas em (JOHNSON, 1991; GARTNER, 2007; JUNQUEIRA; CARNEIRO, 2013), e sintetizadas aqui:

- a) **fixação**: o tecido é imerso em soluções químicas, como formaldeído ou álcool, para preservar sua integridade molecular e estrutural, evitando degradação celular;
- b) **desidratação e diafanização**: o tecido é tratado com uma série de banhos em álcoois de concentrações crescentes, seguido pelo uso de xilol ou solventes similares para torná-lo translúcido.
- c) **inclusão**: o material é embebido em parafina líquida, que solidifica ao esfriar, formando um bloco rígido que conserva o tecido e facilita seu corte.
- d) **microtomia**: o bloco de parafina é fatiado em cortes finos com espessura de 1 a 10 micrômetros usando um micrótomo. Essas secções permitem a visualização detalhada das estruturas celulares.
- e) **montagem e coloração**: os cortes são colocados em lâminas de vidro e submetidos à coloração, geralmente H&E. A hematoxilina cora os núcleos celulares de violeta, enquanto a eosina dá uma tonalidade rosada às estruturas citoplasmáticas e extracelulares, destacando componentes do tecido. Ao fim da coloração, o tecido é colocado entre lâminas de vidro.

Após a coloração, as lâminas são analisadas diretamente sob o microscópio ou digitalizadas utilizando dispositivos equipados com câmeras de alta resolução (BELSARE; MUSHRIF, 2012). As imagens resultantes são armazenadas em arquivos digitais, geralmente no modelo de cores RGB. Esse processo permite que especialistas analisem as imagens de qualquer lugar, facilitando consultas e diagnósticos colaborativos, além de permitir que técnicas de processamento de imagens sejam empregadas para detecção de anormalidades, como padrões displásicos e características tumorais (KRISHNAN et al., 2011a; BOROVEC, 2013).

Apesar dos avanços, existem desafios associados ao processo, como artefatos durante a coloração, dificuldades em cortes uniformes ou limitações na resolução das imagens capturadas. Apesar disso, a digitalização histológica tem se mostrado como uma ferramenta poderosa para diagnósticos, monitoramento de doenças e pesquisas biomédicas (PALLUA et al., 2020).

2.4 Sistemas Computacionais para Apoio ao Diagnóstico

Os sistemas CAD combinam o conhecimento de medicina e sistemas computacionais para aplicar regras de diagnóstico em grande quantidade de dados, simulando a tomada

de decisões clínicas por um especialista (YANASE; TRIANTAPHYLLOU, 2019). Fatores como a complexidade do processo de diagnóstico, a grande quantidade de dados clínicos a serem analisados e a evolução da área de inteligência artificial levaram ao uso de sistemas computacionais na medicina (BOROVEC, 2013; LYNCH; LISTON, 2018). Sistemas CAD mais avançados podem extrair conhecimento a partir de dados clínicos e utilizar esse conhecimento para aprimorar seus resultados (OBERMEYER; EMANUEL, 2016).

No uso desse sistema, as imagens digitalizadas são processadas pelo algoritmo e os resultados gerados são utilizados para auxiliar o especialista como uma segunda opinião (CHAN; HADJIISKI; SAMALA, 2020). Dentre as aplicações de sistemas CAD, a análise de imagens histológicas vem recebendo cada vez mais atenção dos especialistas, pois pode aprimorar o diagnóstico médico e fornecer avaliações precisas (LYNCH; LISTON, 2018). Com a evolução do processamento de imagens digitais, muitas técnicas foram desenvolvidas com o intuito de melhorar todo o processo de identificação de anomalias nesse tipo de dado (IRSHAD et al., 2014). Com uso dessas técnicas, problemas no diagnóstico causados por ruídos, manchas de pigmentação e divergência na análise realizada por diferentes patologistas podem ser minimizadas (KRISHNAN et al., 2011b).

Durante a análise computacional de imagens histológicas, as etapas de pré-processamento, segmentação, extração de características e classificação são comumente empregadas, conforme ilustrado na Figura 2. O pré-processamento é utilizado para preparar a imagem para as operações de segmentação e análise. Sua função é a eliminação de ruídos e variações na pigmentação, realce de contraste e bordas, entre outras operações (YANASE; TRIANTAPHYLLOU, 2019). Geralmente, nessa etapa são empregados algoritmos para normalização de corantes, filtros para eliminação de ruídos e operações de suavização com objetivo de realçar estruturas (KAUSHAL et al., 2019). A etapa de segmentação é importante para a área da histologia por separar as estruturas histológicas da região de fundo, evitando que essas regiões influenciem negativamente na análise (YANASE; TRIANTAPHYLLOU, 2019). Em sistemas modernos, modelos de DL, podem ser aplicados para realizar essa tarefa (XIE et al., 2018; CHAN; HADJIISKI; SAMALA, 2020).

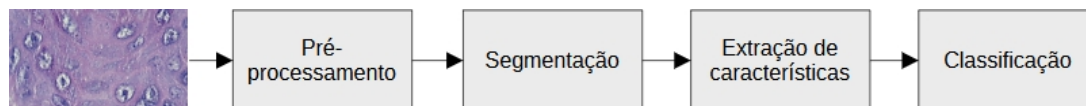


Figura 2 – Etapas comumente empregadas em tarefas de análise computacional de imagens histológicas.

A etapa de extração de características permite computar os atributos dos núcleos segmentados de maneira que anomalias possam ser identificadas (IRSHAD et al., 2014). As características comumente usadas são relacionadas à intensidade, morfologia e textura dos núcleos (KAUSHAL et al., 2019). Essas características são utilizadas para medir alterações nos tecidos e identificar o surgimento de lesões (BEJNORDI et al., 2015). A tarefa de classificação é empregada para associar regiões e objetos às classes predefinidas. Nessa

etapa, classificadores são empregados para categorizar as estruturas em diferentes classes de acordo com o tipo de tecido e o nível da lesão (YANASE; TRIANTAPHYLLOU, 2019). Diversas técnicas, como regressão logística, *K-Nearest Neighborhood*, e Redes Neurais Artificiais (RNA) podem ser aplicadas em imagens histopatológicas para classificar tecidos como saudáveis ou lesionados (KAUSHAL et al., 2019).

2.5 Segmentação

A segmentação de imagens histológicas é uma etapa relevante no processamento digital para sistemas CAD. Essa tarefa permite a identificação de regiões ou objetos específicos dentro da imagem, facilitando análises subsequentes (MINAEE et al., 2021). No contexto da histopatologia digital, o diagnóstico frequentemente depende da análise detalhada dos componentes celulares, sendo o formato, a aparência e a disposição espacial dos núcleos elementos-chave no exame histopatológico (KUMAR et al., 2019). Assim, a segmentação de estruturas celulares, especialmente núcleos, desempenha um papel importante na histologia digital, permitindo a identificação de anomalias teciduais e contribuindo para caracterização de lesões (VUOLA; AKRAM; KANNALA, 2019; LIU et al., 2020).

Em histopatologia, a segmentação é desafiadora devido à irregularidade visual dos tecidos e núcleos celulares, o que pode resultar em métodos com desempenho inconsistente para diferentes tipos de lesões (GONZALEZ; WOODS, 2018; KUMAR et al., 2019). O diagnóstico de DEOs, por exemplo, frequentemente envolve a análise de alterações no formato e no tamanho dos núcleos celulares (KADEMANI, 2019) e, nesse contexto, a caracterização morfológica e espacial dos núcleos é essencial para o diagnóstico e monitoramento dessas lesões, reforçando a necessidade de métodos robustos e eficientes para segmentação automática (HESAMIAN et al., 2019).

Métodos de segmentação comumente utilizados são baseados em limiarização, detecção de bordas e agrupamento de regiões similares. Os métodos baseados em limiarização utilizam valores de intensidade como limites que separam classes de objetos. Essa separação pode ser realizada utilizando apenas um valor de numérico para segmentação ou mais valores para essa operação. Ao utilizar um valor de limiar, os *pixels* com intensidade abaixo desse valor são atribuídos à uma classe e *pixels* com valores acima são atribuídos à outra classe. Um número maior de limiares pode ser empregado para definir mais classes na imagem (GONZALEZ; WOODS, 2018). Essa segmentação pode ser realizada de maneira global ou local. Na limiarização global, os valores definidos como limiares são aplicados na imagem inteira, sendo útil em situações com alta distinção entre objetos. Na limiarização local, a imagem é dividida em regiões e os valores de limiar são definidos para cada uma dessas regiões, sendo utilizada quando existem alterações de brilho em diferentes regiões da imagem (BALI; SINGH, 2015). Técnicas de limiarização comumente utilizadas são o Otsu e o *Gaussian Mixture Model*. Em imagens histológicas, esses méto-

dos calculam os valores de limiar automaticamente buscando maximizar a variância entre as classes de objetos (LEE; LIEW; THONG, 2015).

Uma outra abordagem para segmentação é a detecção de bordas, que busca a identificação de mudanças bruscas nos níveis de intensidade da imagem. Os *pixels* nessas regiões com alterações bruscas são considerados *pixels* de borda (LEE; LIEW; THONG, 2015). Os algoritmos dessa categoria geralmente empregam operadores de gradiente para identificar essas alterações nos sentidos horizontal, vertical e diagonal e definir se pertencem à borda de um objeto (ASLAM; KHAN; BEG, 2015). Exemplos de algoritmos utilizados para identificar bordas em regiões de imagens são os detectores *Sobel*, *Canny* e *Laplacian of Gaussian*.

Na terceira categoria, os métodos baseados em regiões utilizam critérios de similaridade predefinidos para identificar regiões com características semelhantes e atribuí-las à classes de objetos (SOLOMON; BRECKON, 2000). Essas regiões podem ser definidas através do agrupamento de regiões com características similares ou pela divisão da imagem em regiões com propriedades dissimilares. Métodos de agrupamento buscam regiões com propriedades similares e as agrupam entre si, gerando regiões maiores. Os métodos de divisão procuram por áreas que apresentam grande variação de intensidade e as separam em regiões distintas, gerando regiões menores (BALI; SINGH, 2015). Em ambos, o número de repetições predefinido vai determinar o tamanho das regiões identificadas e, conseqüentemente, a precisão da segmentação. Métodos comumente utilizados dessa categoria são denominados crescimento de regiões, *K-means* e *fuzzy clustering*.

Modelos baseados em DL têm emergido como uma solução promissora para tarefas de extração de características, segmentação e classificação de imagens médicas. Com esses métodos, tarefas complexas podem ser realizadas automaticamente com alta precisão (SHIRAZI et al., 2020). A segmentação pode ser abordada de três formas: semântica, por instâncias ou panóptica, como ilustrado na Figura 3. Na segmentação semântica, os *pixels* pertencentes a uma mesma classe recebem o mesmo rótulo, identificando componentes como núcleos, citoplasma e queratina (MINAEE et al., 2021). Essa abordagem permite reconhecer as regiões da imagem e proporciona uma visão geral da estrutura do tecido. No entanto, todos os objetos de uma determinada classe são identificados com o mesmo rótulo, o que representa uma limitação em imagens com núcleos ou estruturas celulares sobrepostas.

A segmentação de instâncias combina a detecção de objetos com a segmentação para gerar máscaras individuais para cada objeto dentro de uma mesma classe (SULTANA; SUFIAN; DUTTA, 2020). Esse processo envolve a identificação dos objetos nos mapas gerados pelas arquiteturas de DL, a definição de caixas delimitadoras e a aplicação da segmentação semântica dentro dessas regiões, resultando em uma máscara única para cada objeto identificado (SULTANA; SUFIAN; DUTTA, 2020). Esse tipo de segmentação permite isolar núcleos sobrepostos e analisar características morfológicas deles.

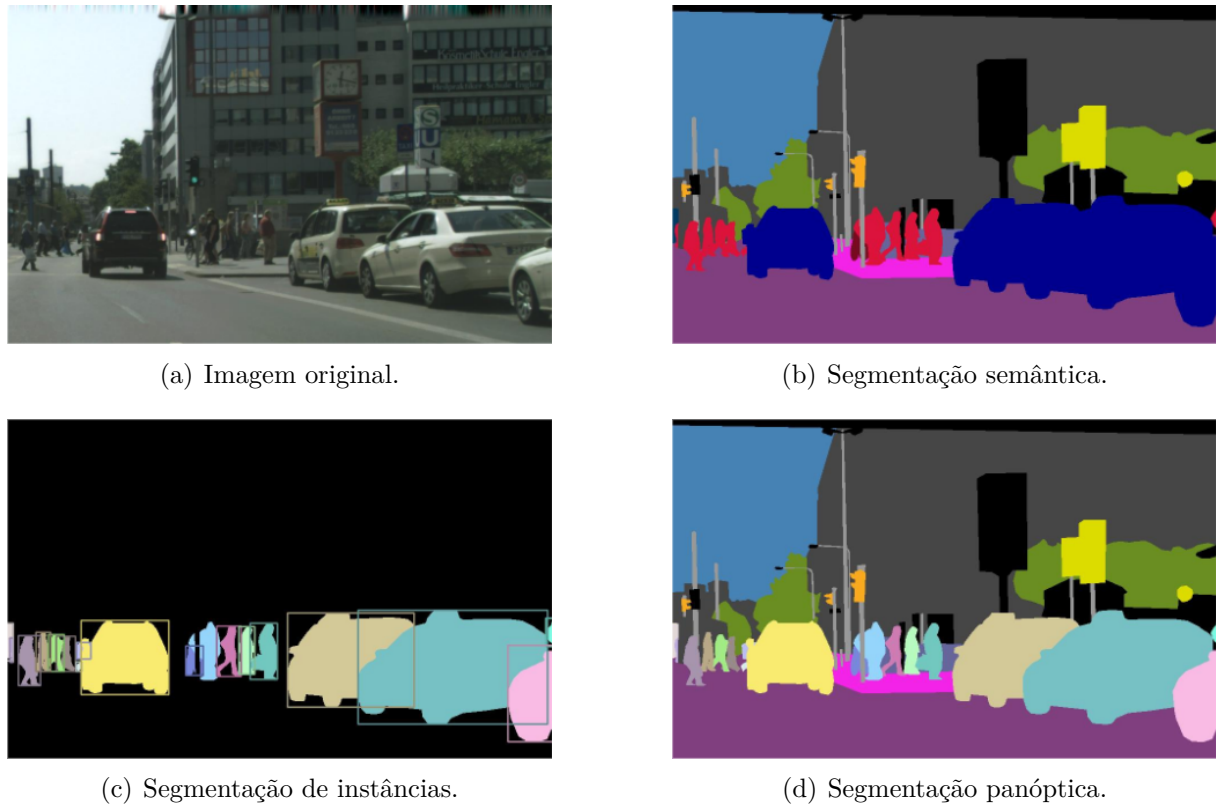


Figura 3 – Exemplos dos diferentes tipos de segmentação: (a) imagem original, (b) segmentação semântica (rotulação por classes de *pixels*), (c) segmentação de instâncias (rotulação por objetos individuais) e (d) segmentação panóptica (rotulação por classes de *pixels* e por instâncias de objetos). Imagem adaptada de (KIRILLOV et al., 2019)).

Já a segmentação panóptica combina as características das abordagens anteriores, proporcionando uma visão completa da imagem ao integrar informações semânticas e das instâncias (LI; CHEN, 2021). Essa abordagem gera uma máscara com rótulos semânticos e outra para as instâncias dos objetos de interesse. Dessa forma, regiões homogêneas são identificadas e objetos distintos dentro dessas regiões são segmentados separadamente (KIRILLOV et al., 2019). Essa abordagem possibilita segmentar regiões de tecido na imagem e, ao mesmo tempo, identificar núcleos individuais.

2.6 Redes Neurais Artificiais

As RNAs são modelos inspirados na estrutura e funcionamento do cérebro humano, em que os neurônios biológicos processam estímulos externos e realizam tarefas complexas de análise e aprendizado (HAYKIN et al., 2009). Esses modelos buscam emular os elementos fundamentais do sistema nervoso: os dendritos, que recebem sinais de outros neurônios; o corpo celular, que processa os sinais recebidos; os axônios, que transmitem novos sinais; e as sinapses, responsáveis pelas conexões entre neurônios (JUNQUEIRA; CARNEIRO,

2013). Analogamente, um neurônio artificial é composto por três componentes principais: (i) pesos sinápticos, que representam a força das conexões; (ii) uma função de soma, que combina os sinais recebidos; e (iii) uma função de ativação, que transforma a saída em um valor limitado e não-linear (HAYKIN et al., 2009). Na Figura 4, é exibido um modelo de neurônio artificial.

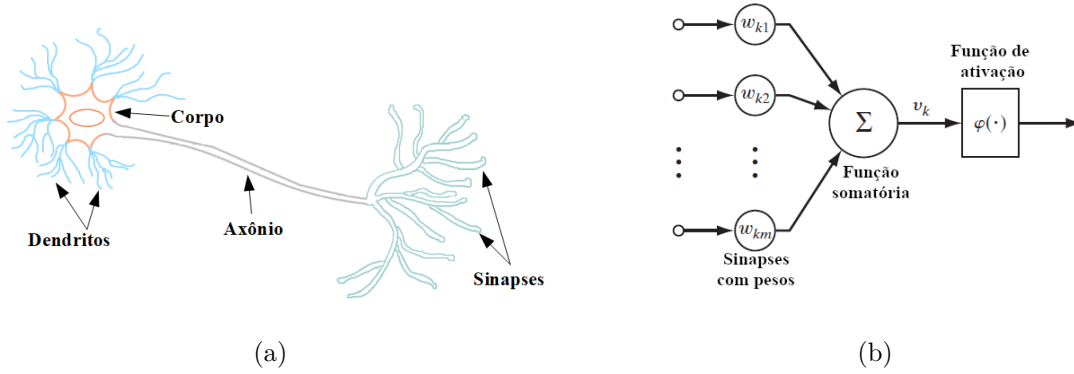


Figura 4 – Exemplos de neurônios: (a) neurônio biológico (fonte: (SILVA, 2019)) e (b) neurônio artificial (adaptada de (HAYKIN et al., 2009)).

Elas são organizadas em, no mínimo, uma camada de entrada, que recebe os dados iniciais; uma camada de saída, que realiza a classificação ou regressão de acordo com padrões aprendidos; e camadas ocultas, responsáveis por extrair características complexas (AGGARWAL et al., 2018). Durante o treinamento, um grande número de exemplos são apresentados para o modelo de rede e os pesos sinápticos são ajustados iterativamente para minimizar o erro, utilizando algoritmos como a retropropagação do erro (BUDUMA; LOCASCIO, 2017).

Nos últimos anos, RNAs têm sido amplamente aplicadas na área da saúde, incluindo a análise de imagens médicas, demonstrando resultados promissores em tarefas de CAD, como detecção de anomalias e segmentação de estruturas biológicas (DENG et al., 2020; MARTINO et al., 2020). Entretanto, devido aos desafios associados à segmentação de imagens, modelos de CNN tornaram-se a abordagem dominante para essa tarefa. Essas arquiteturas possuem a capacidade de extrair automaticamente características relevantes diretamente das imagens de entrada, eliminando a necessidade de engenharia manual de atributos (GOODFELLOW; BENGIO; COURVILLE, 2016).

As CNNs são uma evolução das RNAs, projetadas especificamente para lidar com dados estruturados em forma de grades, como imagens (GOODFELLOW; BENGIO; COURVILLE, 2016). Diferentemente das RNAs tradicionais, as arquiteturas de CNNs utilizam camadas convolucionais para explorar propriedades espaciais e hierárquicas das imagens, extraindo características relevantes em diferentes escalas. A arquitetura das CNNs geralmente consiste nas operações de convolução, ativação e *pooling* (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Durante a convolução, filtros percorrem a imagem e realizam operações matemáticas, calculando a média ou os valores máximos de regiões e gerando mapas de

características que destacam padrões como bordas, texturas e formas. Exemplos dessa operação são mostrados na Figura 5. Cada camada convolucional aprende representações progressivamente mais abstratas da imagem. Enquanto as camadas iniciais detectam elementos simples, como formas e bordas, as camadas intermediárias e finais combinam essas informações para identificar informações complexas de textura (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

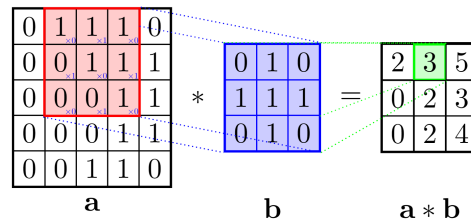


Figura 5 – Exemplo de uma operação de convolução. Os elementos da imagem a são multiplicados pelos elementos do filtro b e os resultados são somados, gerando o mapa resultante.

A função de ativação introduz não linearidade ao modelo, permitindo que a rede aprenda padrões complexos nos dados. Entre as funções de ativação mais comuns, destacam-se a ReLU (*Rectified Linear Unit*), que acelera o treinamento e reduz o problema do desaparecimento do gradiente, e a *softmax*, utilizada na camada de saída para converter os valores da rede em uma distribuição de probabilidades entre as classes possíveis (AGGARWAL et al., 2018).

Além disso, operações de *pooling*, como *max-pooling* e *average-pooling*, são usadas para reduzir a dimensionalidade dos mapas de características, preservando as informações mais importantes enquanto descartam redundâncias (AGGARWAL et al., 2018). O *max-pooling* seleciona o valor máximo dentro de uma região da imagem, enfatizando características mais fortes, como contornos e padrões de textura. Já o *average-pooling* calcula a média dos valores dentro da região, suavizando os detalhes e tornando a representação da imagem mais robusta a variações locais. Essas operações aumentam a capacidade de generalização dos modelos, tornando-o menos sensível a pequenas mudanças na imagem de entrada (ALBAWI; MOHAMMED; AL-ZAWI, 2017). Exemplos dessas operações são ilustrados na Figura 6.

A etapa final das CNNs geralmente consiste em camadas totalmente conectadas, que recebem os mapas de características extraídos pelas camadas anteriores e combinam essas informações para realizar a rotulação final. Diferentemente das camadas convolucionais, onde cada neurônio está conectado apenas a uma região específica da entrada, as camadas totalmente conectadas possuem conexões entre todos os neurônios, permitindo a integração global das características aprendidas. Essa estrutura possibilita gerar o mapa de segmentação com base em todas as informações extraídas ao longo das camadas convolucionais. No entanto, devido ao grande número de parâmetros envolvidos, essas camadas

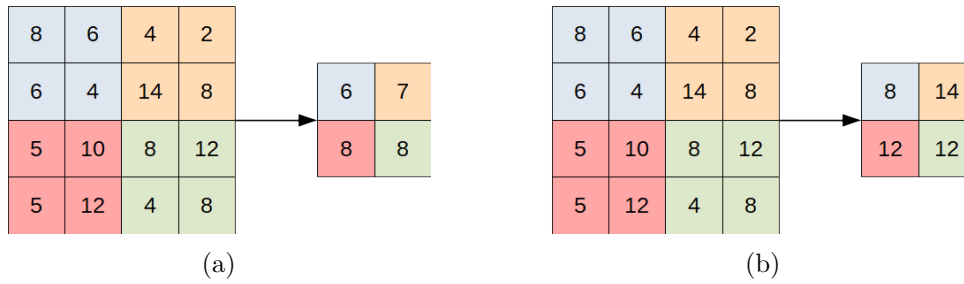


Figura 6 – Exemplos de operações de *pooling*: (a) *average pooling* e (b) *max pooling* (imagens retiradas de: (SILVA, 2019)).

também aumentam a complexidade do modelo, exigindo estratégias como regularização para evitar o *overfitting* (GOODFELLOW; BENGIO; COURVILLE, 2016).

O treinamento de CNNs demanda grandes volumes de dados rotulados e considerável poder computacional. Além disso, a necessidade de anotações manuais detalhadas para treinar esses modelos representa um desafio, pois a rotulagem de imagens médicas é um processo demorado e depende de especialistas. Ainda assim, avanços recentes em técnicas de aprendizado auto-supervisionado e transferência de aprendizado têm buscado mitigar essas limitações, permitindo que CNNs sejam aplicadas de maneira mais eficiente em cenários com dados limitados (MINAEE et al., 2021).

2.7 Comitês de Aprendizagem Profunda

Comitês referem-se à combinação dos resultados de dois ou mais algoritmos de aprendizado de máquina com o objetivo de obter previsões mais robustas e confiáveis. Essa abordagem tem se mostrado eficaz em diversas aplicações, pois permite a mitigação de erros individuais dos modelos, combinando seus acertos para melhorar a generalização e a estabilidade das previsões (KASSANI et al., 2019; SAGI; ROKACH, 2018). No contexto de DL, comitês de CNNs têm sido amplamente utilizados para aumentar o desempenho em tarefas complexas, como classificação e segmentação de imagens médicas (DOLZ et al., 2020). Um exemplo do funcionamento de um comitê é apresentado na Figura 7, que ilustra o limiar de decisão para rotulação em três modelos, a agregação através de uma operação de somatório e o limiar de decisão final após essa agregação.

A motivação para o uso de comitês está na premissa de que os erros produzidos por um modelo podem ser compensados por outros modelos dentro do grupo. Dessa forma, o uso de múltiplas abordagens, quando adequadamente combinadas, pode minimizar limitações individuais e melhorar a generalização (DOLZ et al., 2020; GHOSH et al., 2021). Em problemas complexos, como a análise de imagens médicas, onde a variabilidade dos dados pode comprometer o desempenho de um único modelo, o uso de comitês tem demonstrado ser uma abordagem eficaz para aumentar a confiabilidade das previsões (GHOSH et al., 2021). Além disso, os comitês ajudam a mitigar problemas como *overfitting*, uma vez que

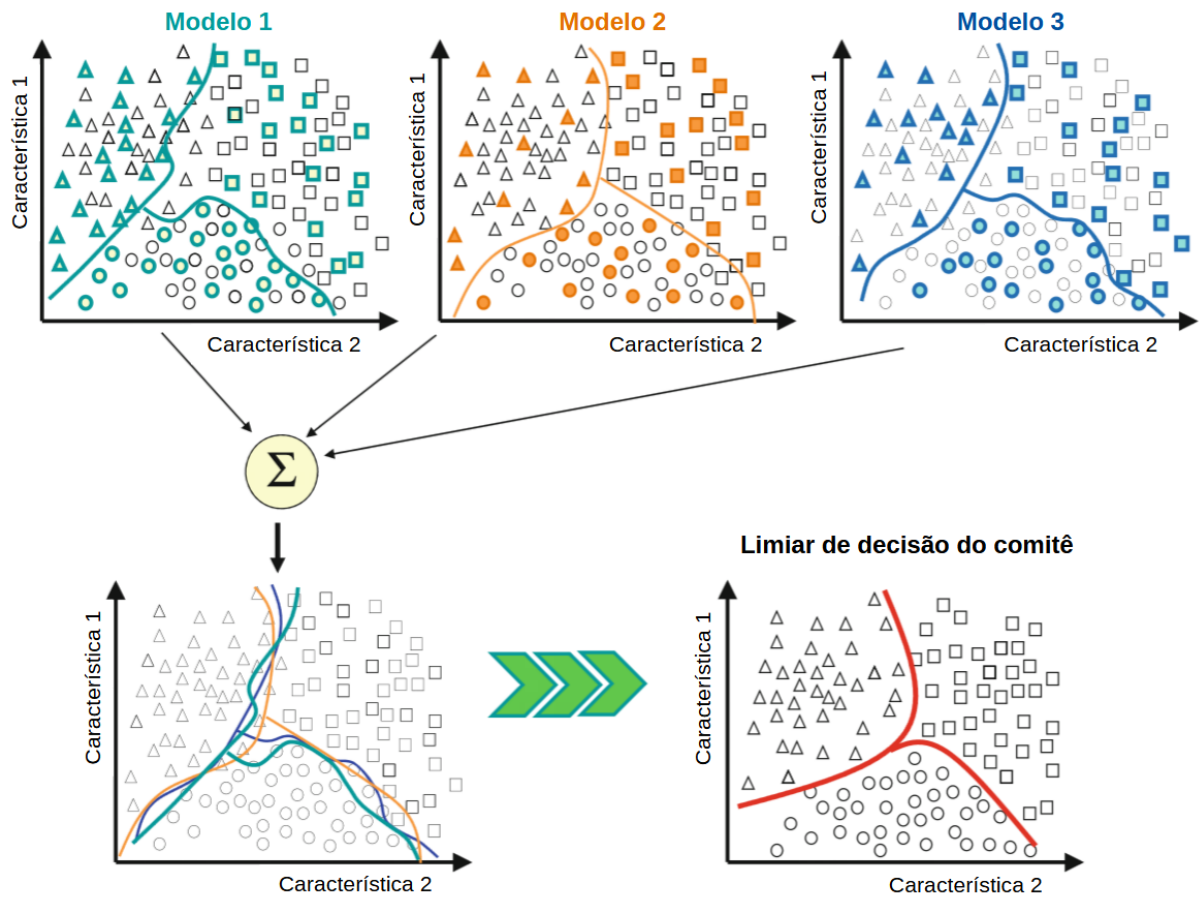


Figura 7 – Exemplo de funcionamento de um comitê com três modelos de DL. Os pesos e limiares de decisão dos modelos são agregados através de uma operação de somatório, gerando um novo limiar para o comitê. Imagem adaptada de (POLIKAR, 2012).

combinam características de modelos que podem generalizar de maneira distinta sobre os dados de treinamento (SAGI; ROKACH, 2018). Outra vantagem está na redução da variação das previsões, o que pode ser benéfico em cenários com classes desbalanceadas ou dados com alto nível de ruído (SAGI; ROKACH, 2018).

Para que os comitês sejam eficazes, é importante que seus membros apresentem diversidade, pois modelos muito similares tendem a gerar erros correlacionados, o que reduz o impacto positivo da combinação. Essa diversidade pode ser introduzida por meio da variação nos dados de treinamento e no uso de diferentes arquiteturas e hiperparâmetros (NANNI et al., 2022). Essa abordagem visa criar uma compensação entre as diferentes fontes de erro, de modo que os erros de um modelo possam ser corrigidos pelos demais, resultando em previsões mais confiáveis.

Dentre as principais estratégias utilizadas para a construção de comitês, destacam-se o *bagging* e o *boosting*. O *bagging* (*Bootstrap Aggregating*) é uma técnica em que múltiplos modelos são treinados de forma independente, utilizando subconjuntos distintos dos dados de treinamento, obtidos por amostragem com reposição. A previsão final

é gerada a partir da combinação dos resultados individuais dos modelos, reduzindo a variância e tornando o sistema mais robusto (MOHAMMED; KORA, 2023). Por outro lado, o *boosting* constrói o comitê de forma sequencial, onde cada novo modelo é treinado para corrigir os erros cometidos pelos modelos anteriores. Essa estratégia melhora a capacidade de aprendizado ao reduzir o viés das previsões, sendo métodos como *AdaBoost* e *Gradient Boosting* amplamente utilizados em diversos domínios (GANAIE et al., 2022).

No contexto de DL, estratégias de comitês foram desenvolvidas para combinar múltiplas CNNs e melhorar o desempenho dos modelos individuais. Essas estratégias podem ser agrupadas em dois níveis: manipulação de dados e manipulação de modelos (MOHAMMED; KORA, 2023). No nível dos dados, diferentes conjuntos de treinamento podem ser gerados por amostragem ou validação cruzada, permitindo que modelos distintos aprendam características complementares dos dados. Já no nível dos modelos, abordagens exploram variações estruturais em CNNs, como modificar os hiperparâmetros ou utilizar arquiteturas distintas sobre os mesmos dados. Isso permite reduzir a variação dos modelos individuais e aumentar a robustez da previsão (SALEH et al., 2022). No entanto, a complexidade dos modelos de CNN, que possuem milhões de hiperparâmetros, representa um desafio adicional para a construção de comitês eficientes.

Os comitês de DL podem ser organizados seguindo quatro abordagens principais, conforme ilustrado na Figura 8. Na primeira abordagem, o comitê utiliza os mesmos dados e o mesmo modelo para treinamento, mas variando parâmetros internos, como inicializações diferentes (MOHAMMED; KORA, 2023). Na segunda abordagem, o mesmo conjunto de dados é utilizado para treinar modelos distintos, explorando diferentes arquiteturas para capturar representações variadas (NANNI; LUMINI; FANTOZZI, 2023). Para a terceira abordagem, diferentes amostras de dados são utilizadas para treinar um mesmo modelo, permitindo que ele aprenda padrões complementares em diferentes amostras (LEE et al., 2024). Já na quarta abordagem, modelos distintos são treinados utilizando diferentes amostras de dados, maximizando a diversidade tanto na representação quanto na generalização (MOHAMMED; KORA, 2023).

Além das técnicas de treinamento, diferentes regras de votação podem ser empregadas para combinar as previsões dos modelos que compõem o comitê. Uma das abordagens mais comuns é a votação da maioria, na qual a classe mais frequentemente predita pelos modelos é selecionada como a previsão final. Essa técnica é útil em problemas de classificação, pois permite uma decisão simples e direta baseada na contagem de votos (KRAWCZYK et al., 2017). Outra estratégia amplamente utilizada é a regra da soma, em que as probabilidades atribuídas a cada classe pelos modelos do comitê são somadas, resultando em uma previsão mais robusta ao minimizar o impacto de previsões errôneas individuais (SAGI; ROKACH, 2018). Além disso, pode-se utilizar a soma ponderada, uma variação da regra da soma em que pesos específicos são atribuídos a cada modelo com base em seu desempenho, garantindo que modelos mais confiáveis tenham

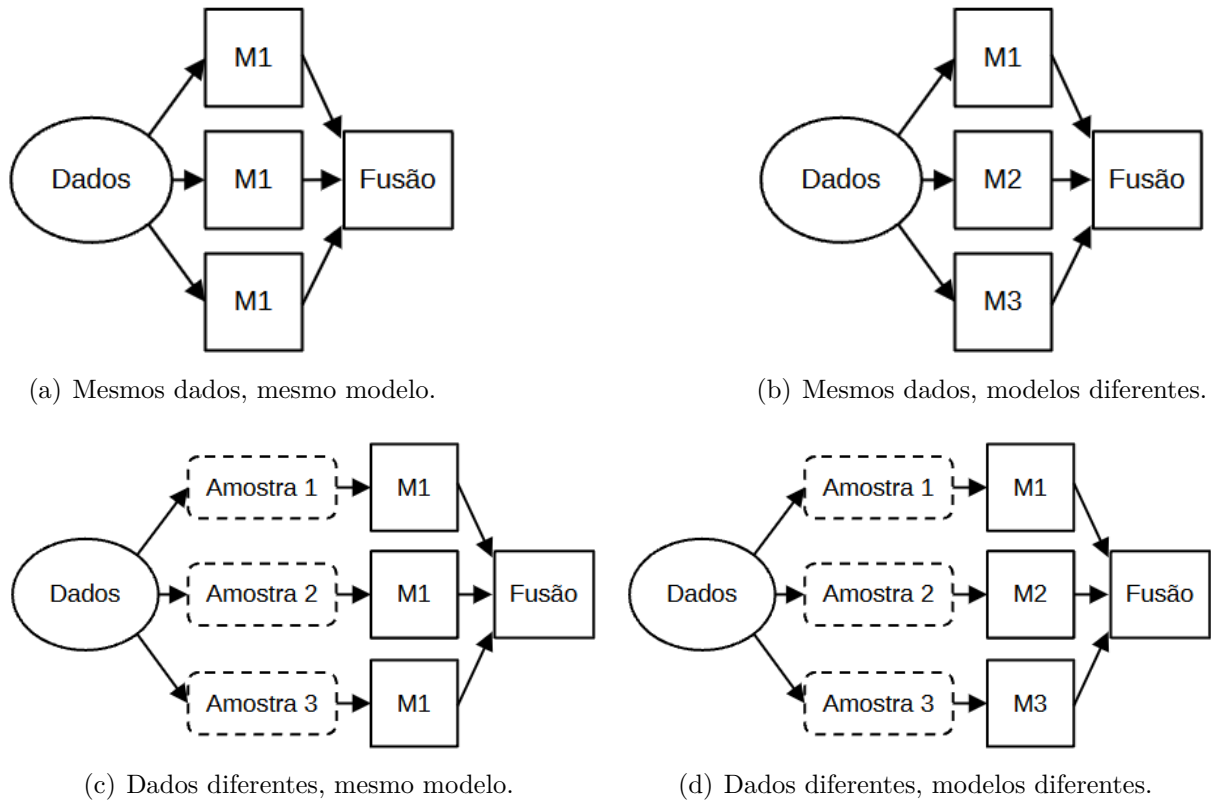


Figura 8 – Exemplos de diferentes composições para comitês de modelos de DL. Imagem adaptada de Mohammed e Kora (2023).

maior influência sobre a decisão final (NANNI; LUMINI; GHIDONI, 2018).

No campo da segmentação de imagens médicas, os comitês têm se destacado como uma estratégia robusta para lidar com os desafios inerentes a esse tipo de dado (NANNI; LUMINI; FANTOZZI, 2023). Problemas como ruído, baixa qualidade, sobreposição de estruturas e variabilidade interobservadores tornam a segmentação de imagens médicas uma tarefa desafiadora. O uso de comitês nessa área não apenas melhora a robustez das predições, mas também oferece uma alternativa eficaz para aumentar a generalização dos resultados, o que é importante em aplicações clínicas que exigem alta precisão, como o diagnóstico e o planejamento terapêutico (DONG et al., 2020).

Em muitos casos, a fusão das máscaras de segmentação geradas por diferentes modelos permite obter um resultado aprimorado (NANNI et al., 2022). No entanto, métodos estatísticos também podem ser empregados para ponderar as contribuições individuais, ajustando as predições conforme a confiança de cada modelo (MOHAMMED; KORA, 2023). Além disso, abordagens emergentes exploram o uso de RNAs dedicadas à fusão de múltiplas segmentações, permitindo uma integração mais sofisticada dos resultados e aumentando a robustez das predições (YANG; LV; CHEN, 2021). Assim, o uso de comitês continua a evoluir como uma solução promissora para os desafios encontrados nesse domínio.

2.8 Métricas de Avaliação para Segmentação

A avaliação da segmentação, tanto semântica quanto por instâncias, é realizada por meio de métricas que quantificam a capacidade de detecção e a sobreposição das regiões segmentadas pelas abordagens em relação ao padrão ouro. Para isso, comumente utiliza-se a análise de quatro medidas principais: Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN). Essas medidas permitem validar o desempenho dos métodos propostos e identificar possíveis melhorias (CHANG et al., 2014). Com base nessas medidas, diversas métricas podem ser obtidas, incluindo quatro métricas são amplamente utilizadas na literatura: *Acurácia* (A_{CC}), *Sensibilidade* (S_E), *Especificidade* (E_S) e *Coeficiente Dice* (D_C) (EELBODE et al., 2020). Cada uma dessas métricas é projetada para capturar aspectos específicos do desempenho de um método de segmentação.

A extração dessas métricas depende do tipo de segmentação realizada. Para a segmentação semântica, a avaliação é feita por meio da comparação direta entre os *pixels* da máscara segmentada e do padrão ouro, considerando toda a imagem como uma única unidade de análise (EELBODE et al., 2020). Já na segmentação por instâncias, a correspondência entre os objetos segmentados e os do padrão ouro é importante. Assim, a avaliação é realizada apenas nos *pixels* pertencentes a objetos que apresentam sobreposição significativa entre a segmentação predita e imagem de referência (WANG; WANG; ZHU, 2020).

A A_{CC} mede a proporção de rotulações corretas em relação ao total de classificações realizadas (SUNDARA; AARTHI, 2019). É uma métrica que avalia o desempenho geral do modelo, sendo útil para obter uma visão ampla sobre o resultado da segmentação. Ela é calculada pela equação:

$$A_{CC} = \frac{\sum_i (VP_i + VN_i)}{\sum_i (VP_i + FP_i + VN_i + FN_i)}, \quad (1)$$

onde VP_i , VN_i , FP_i e FN_i representam, respectivamente, os VP, VN, FP e FN para cada instância i . Na segmentação semântica, o índice i equivale a 1, pois avalia a sobreposição considerando toda a imagem como uma única instância. Já na segmentação via instâncias, o somatório ocorre sobre todas as instâncias identificadas na imagem.

A S_E mede a proporção de *pixels* corretamente rotulados como pertencentes à classe de interesse (VP) em relação ao total de *pixels* dessa classe na imagem de referência (WANG; WANG; ZHU, 2020). Ela é importante para avaliar a capacidade do modelo de detectar as regiões de interesse (do inglês, *Region of Interest* (ROI)), sendo definida como:

$$S_E = \frac{\sum_i VP_i}{\sum_i (VP_i + FN_i)}. \quad (2)$$

Para a segmentação semântica, o cálculo considera toda a máscara segmentada como uma única classe de interesse. Na segmentação via instâncias, o cálculo é realizado para

cada instância separadamente e, em seguida, os resultados são agregados por meio do somatório.

A E_S avalia a proporção de *pixels* corretamente rotulados como região de fundo, ou seja, os VN (WANG; WANG; ZHU, 2020). É uma métrica complementar à sensibilidade que indica a capacidade do modelo de evitar falsos positivos, sendo definida como:

$$E_S = \frac{\sum_i VN_i}{\sum_i (VN_i + FP_i)}. \quad (3)$$

Assim como as outras métricas, o cálculo para segmentação semântica considera toda a imagem como uma única classe, enquanto na segmentação por instâncias o cálculo é realizado para cada instância e os resultados são somados.

O D_C é amplamente utilizado para medir a similaridade entre duas máscaras binárias, quantificando a sobreposição entre as regiões segmentadas pelo modelo e a imagem de referência (SUNDARA; AARTHI, 2019). Para a segmentação semântica, o D_C é calculado como:

$$D_C = \frac{\sum_i (2 \times VP_i)}{\sum_i (2 \times VP_i + FP_i + FN_i)}. \quad (4)$$

Para a segmentação via instâncias, o cálculo considera cada instância separadamente. Isso permite medir a precisão da segmentação em nível individual, garantindo que as instâncias segmentadas correspondam adequadamente às instâncias do padrão-ouro.

Além das métricas de desempenho, a análise estatística é importante para avaliar a significância dos resultados em segmentação de imagens médicas. O teste de Wilcoxon, ou *Wilcoxon Signed-Rank Test*, é um método estatístico não paramétrico amplamente utilizado para comparar pares de amostras dependentes, como os resultados de dois métodos de segmentação aplicados ao mesmo conjunto de imagens (LIU et al., 2021; EHSAEYAN, 2023; MA et al., 2024). Ele avalia a hipótese nula de que a mediana das diferenças entre os pares de amostras é zero, sendo útil para dados que não seguem uma distribuição normal (WILCOXON, 1992). O cálculo baseia-se nas diferenças $d_i = x_i - y_i$, onde x_i e y_i representam os desempenhos dos dois métodos. As diferenças são ranqueadas em valores absolutos, ignorando os pares onde $d_i = 0$. A estatística W é definida como:

$$W = \sum_{i=1}^n R_i \cdot \text{sign}(d_i), \quad (5)$$

onde R_i é o ranque de $|d_i|$ e $\text{sign}(d_i)$ representa o sinal de d_i .

O valor ρ , derivado da distribuição dos ranques, permite determinar se as diferenças observadas são estatisticamente significativas (DERRAC et al., 2011). Um valor ρ menor que um nível de 0,05 indica evidências suficientes para rejeitar a hipótese nula, sugerindo que as diferenças entre dois métodos são estatisticamente relevantes. Essa abordagem permite validar a eficácia de métodos de segmentação, destacando sua importância na análise comparativa e na escolha do modelo mais adequado para aplicações específicas (EHSAEYAN, 2023).

Essas métricas podem ser empregadas em ambos os contextos, semântico e por instâncias. Na segmentação semântica, onde o objetivo é identificar áreas ou regiões homogêneas, o cálculo é realizado sobre toda a máscara segmentada. Já na segmentação via instâncias, cujo objetivo é identificar e delinear objetos individuais, o cálculo é feito para cada instância separadamente e os resultados são agregados por meio do somatório.

Estado da Arte

Nesta seção são apresentados estudos relacionados ao processamento de imagens histológicas utilizando DL e comitês. Para embasar o estudo proposto, foram considerados apenas artigos aplicados a imagens histológicas, priorizando aqueles que utilizam DL ou comitês para segmentação de imagens de DEO. Dado que poucos estudos focam nesse tópico (ADEL et al., 2018), não foram encontrados artigos que empregassem comitês para segmentação de DEOs. Assim, a busca de artigos de comitês foi expandida para incluir estudos que abordam histologia geral.

A busca bibliográfica foi realizada nas bases de dados *Google Scholar*, *Scopus*, *Web of Science*, *ScienceDirect*, *PubMed* e *IEEE Xplore*, utilizando combinações de palavras-chave e operadores lógicos, com a restrição de publicações a partir de 2020. No total, 262 artigos foram obtidos, conforme detalhado na Tabela 1. Em seguida, foi realizada a seleção dos trabalhos mais relevantes, descartando artigos de revisão e aqueles que não empregavam DL e comitês para segmentação de núcleos em imagens histológicas. Quando foram encontradas múltiplas publicações de um mesmo autor, apenas a mais recente foi considerada. O processo de seleção dos artigos ocorreu em três etapas: leitura dos títulos, análise dos resumos e leitura completa dos artigos considerados relevantes. Esse levantamento proporciona um embasamento sólido para a discussão das abordagens mais recentes de comitês de segmentação em imagens histológicas de DEO e para a justificativa da metodologia proposta.

3.1 Aprendizagem Profunda para Segmentação de Núcleos em Tecidos de DEO

Técnicas de DL empregadas em segmentação são capazes de generalizar estruturas presentes em imagens ao mesmo tempo que mantém informações de localização, o que garante boas representações e resultados precisos. Por isso, o uso desse tipo de modelo para a segmentação em imagens médicas vem ganhando atenção nos últimos anos.

Tabela 1 – Combinações de palavras-chave utilizadas na busca bibliográfica.

Conceito	Palavras-chave	Artigos retornados
CNNs para segmentação semântica	"semantic segmentation" AND ("deep learning" OR "neural networks") AND ("OED" OR "oral epithelial dysplasia") AND ("nuclei" OR "nucleus")	74
CNNs para segmentação via instâncias	"instance segmentation" AND ("deep learning" OR "neural networks") AND ("OED" OR "oral epithelial dysplasia") AND ("nuclei" OR "nucleus")	52
Comitês de segmentadores semânticos	("CNN ensemble" OR "deep learning ensemble") AND "semantic segmentation" AND "histology" AND ("nuclei" OR "nucleus")	62
Comitês de segmentadores via instâncias	("CNN ensemble" OR "deep learning ensemble") AND "instance segmentation" AND "histology" AND ("nuclei" OR "nucleus")	42
Comitês de segmentadores híbridos	("CNN ensemble" OR "deep learning ensemble") AND ("instance segmentation" AND "semantic segmentation") AND "histology" AND ("nuclei" OR "nucleus")	32

Dado o número reduzido de trabalhos voltados especificamente para segmentação de núcleos em imagens de DEO, foram incluídos nesta seção alguns estudos desenvolvidos por pesquisadores do mesmo grupo de pesquisa responsável por este trabalho, que contribuíram de forma significativa para o avanço da área.

No estudo apresentado por Santos et al. (2020), um método foi proposto para a segmentação semântica de núcleos em tecidos histológicos com DEOs. O trabalho utilizou imagens histológicas de lesões orais de camundongos. Como pré-processamento, operações morfológicas de aumento de dados foram empregadas para gerar um número maior de imagens de treinamento. Um modelo baseado na U-Net foi treinado durante 500 épocas na base de imagens aumentada e, após a segmentação, o método de limiarização de Otsu foi aplicado para binarizar a máscara resultante. Os testes foram realizados em 120 imagens de DEOs, sendo 30 para cada classe da lesão. O método apresentou acurácia de 88,2% para a tarefa de segmentação de núcleos.

O trabalho de Silva et al. (2022) apresentou um método para a segmentação de instâncias de núcleos celulares em imagens histológicas de DEO. Quatro técnicas de normalização de cores foram empregadas nas imagens como pré-processamento e a arquitetura Mask R-CNN foi empregada para a segmentação. No estudo, a ResNet50 foi utilizada como *backbone* para extração de características. Os mapas foram empregados para a construção de uma pirâmide de características. Essa RNA foi pré-treinada na base de imagens ImageNet e foi empregada a etapa de ajuste durante 40 épocas. No pós-processamento, operações morfológicas foram aplicadas para remover falhas nos contornos dos núcleos e remover pequenos artefatos. A metodologia foi empregada em uma base de dados com

120 imagens de DEO. A metodologia apresentou A_{CC} de 88,7% para a segmentação de núcleo em imagens de DEO.

Em Shephard et al. (2024), uma metodologia para segmentação de instâncias de núcleos em imagens de mucosa oral foi apresentada. A arquitetura proposta, chamada HoVerNet+, foi composta por uma sequência de *encoder*, composta por 50 camadas residuais, e três sequências de *decoder*. O primeiro *decoder* foi responsável por realizar a segmentação semântica das regiões de *pixel*. A segunda sequência empregou esse mapa na segmentação de instâncias dos núcleos celulares. O terceiro *decoder* segmentou os diferentes tecidos presentes na imagem. O modelo foi treinado sob uma base de dados com 43 imagens histológicas da região da boca, sendo 38 com DEO, durante 40 épocas. O método apresentou D_C de 87% e 82%, respectivamente, para a segmentação de camadas e de núcleos celulares.

O estudo de Aljuaid et al. (2024) propôs uma abordagem para segmentação semântica de núcleos celulares utilizando a RNA U-Net em um modelo de aprendizado multitarefa. Nesse estudo, duas variantes da U-Net foram exploradas: uma com o *encoder* pré-treinado VGG16 e outra com o InceptionV3. Inicialmente, um modelo de classificação foi treinado e utilizado como base para a etapa de segmentação semântica, garantindo o compartilhamento de características. O treinamento foi realizado com 200 iterações, empregando o otimizador Adam, além de técnicas de *early stopping* e ajuste dinâmico da taxa de aprendizado para evitar *overfitting*. A metodologia foi empregada em uma base de dados com 1299 imagens da cavidade oral, sendo 719 de tecidos saudáveis e 580 de tecidos com DEO. A avaliação do método mostrou que a U-Net com a VGG16 apresentou o melhor desempenho, atingindo um D_C de 90,0%, enquanto a versão com a InceptionV3 obteve 89,7%.

3.2 Comitês de Aprendizagem Profunda Aplicados em Imagens Médicas

Embora métodos baseados em DL apresentem um desempenho promissor para segmentação em imagens médicas, essas arquiteturas são sujeitas a *overfitting* e variação nos resultados. Essa limitação geralmente ocorre devido ao baixo número de imagens disponíveis para treinar os modelos. Assim, estudos recentes tem explorado a combinação de RNAs para reduzir esses efeitos durante a segmentação.

O estudo de Ahmed et al. (2023) propôs uma abordagem para segmentação semântica de núcleos celulares em imagens histológicas de fluorescência. A arquitetura empregada foi composta por duas redes U-Net empilhadas, utilizando os modelos VGG-19 e EfficientNetB7 como *backbone* para extração de características em múltiplas escalas. Além disso, o modelo incorporou conexões de atalho e um mecanismo de tripla atenção para refinar a propagação de informações relevantes. No final da rede, operações convolucio-

nais residuais guiadas por atenção foram aplicadas para mitigar problemas de gradiente. Para o estudo, uma base com 670 imagens histológicas de diferentes órgãos foi empregada em conjunto de técnicas de aumento de dados baseados em morfologia. O modelo foi treinado utilizando uma taxa de aprendizado inicial de 0,0001, ajustada dinamicamente durante o treinamento. A abordagem proposta atingiu um D_C de 95,76% na segmentação de núcleos celulares.

Uma abordagem baseada em DL para segmentação de instâncias de núcleos em imagens histológicas foi proposta por Aghera et al. (2023). O método utilizou a arquitetura Squeeze U-Net, uma variante da U-Net composta por blocos de atenção para realçar seletivamente as características relevantes das imagens. O modelo foi treinado na base CryoNuSeg, que contém 30 imagens histológicas, e avaliado nas bases TNBC e MoNuSeg, compostas, respectivamente, por 50 e 51 imagens histológicas de múltiplos órgãos. O treinamento foi realizado durante 300 épocas, com 10 iterações por época, o otimizador Adam e taxa de aprendizado de 0,001, com a função de perda baseada no D_C . Além disso, foi aplicada uma abordagem de comitê homogêneo, na qual cinco modelos Squeeze U-Net foram treinados separadamente com diferentes inicializações de pesos para cada conjunto de treino e teste em uma validação cruzada com $K = 5$. O método obteve um D_C de 76,62% para a segmentação das instâncias dos núcleos.

No trabalho de Yıldız, Memiş e Varlı (2024), um comitê foi proposto para a segmentação de instâncias de núcleos em imagens histológicas. O método explorou o modelo U-Net treinado com oito funções de perda diferentes: *Cross-Entropy*, *Dice*, *Focal*, *Tversky*, *Focal Tversky*, *Log-Cosh Dice*, L1 e *Mean Squared Error*. Após a segmentação, o algoritmo de *Watershed* foi aplicado às máscaras para separar núcleos sobrepostos. As predições foram combinadas por meio de um mecanismo de votação, considerando duas estratégias: uma abordagem probabilística e um método de votação direta. Além da estratégia tradicional de votação em grupo, o estudo propôs um método de seleção de modelos baseado no coeficiente Kappa, visando a escolha dos modelos mais complementares para compor o comitê. A seleção foi realizada iterativamente, iniciando pelo modelo com melhor desempenho na validação e incorporando sucessivamente modelos com menor similaridade em relação aos já selecionados. No estudo, uma base com 4.981 imagens de histologia colorretal foram empregadas, organizadas em sete classes de tecidos e divididas em 80% para conjunto de treinamento e 20% para conjunto de testes. Os modelos foram treinados nesta base durante 40.000 iterações, com um *batch size* de 32, utilizando o otimizador Adam e taxa de aprendizado de 0,0006. A abordagem proposta obteve uma qualidade panóptica média (mPQ) de 52,12% para a segmentação de instâncias de núcleos.

3.3 Considerações

No contexto da histopatologia digital, têm-se observado avanços na identificação e segmentação de estruturas presentes em tecidos histológicos. Devido ao potencial de evolução para lesões cancerígenas, o diagnóstico preciso de DEOs é essencial para a definição do tratamento adequado ao paciente (KADEMANI, 2019). Apesar disso, poucos estudos na literatura focam na segmentação de núcleos em DEOs, e nenhum dos trabalhos levantados emprega comitês de segmentadores nesse contexto.

Um desafio recorrente em DL é o *overfitting*, que ocorre quando o modelo apresenta alta acurácia nos dados de treinamento, mas desempenho inferior no conjunto de testes. Esse problema ocorre em situações onde poucos dados de treinamento estão disponíveis, como ocorre frequentemente na histopatologia digital, em que a obtenção de tecidos histológicos é limitada (YING, 2019). Além disso, as classes frequentemente apresentam desbalanceamento, e as estruturas podem variar significativamente em diferentes regiões do tecido (IRSHAD et al., 2014). Tais desafios reforçam a necessidade de estudos para reduzir a variação dos resultados nos modelos de CNN. Entretanto, definir quais arquiteturas irão compor o comitê e qual sistema de votação será empregado para combinar os resultados é uma tarefa complexa, pois cada abordagem apresenta particularidades dependentes do contexto de aplicação (DONG et al., 2020).

Outro ponto relevante é o impacto do aumento de dados, que tem sido explorado nos estudos analisados. Essas técnicas ajudam a mitigar o *overfitting* e a escassez de imagens, permitindo maior variabilidade no conjunto de treinamento. Contudo, não há um padrão nas técnicas empregadas nos estudos levantados, indicando a necessidade de investigar o impacto de diferentes abordagens no contexto de imagens de DEO. Observou-se ainda que a maioria dos estudos sobre comitês emprega segmentação semântica, com apenas um estudo utilizando comitês baseados em segmentação por instâncias e nenhum trabalho combinando as abordagens semânticas e por instâncias. Essa lacuna representa uma oportunidade para explorar novas soluções para ambas as abordagens, aprimorando a segmentação de regiões complexas em imagens histológicas.

Com base das lacunas observadas, é observado que há desafios significativos para o contexto de DEOs, sendo necessários novas pesquisas explorando o uso de diferentes abordagens de comitês, visando avanços no processo de identificação e segmentação de núcleos celulares em tecidos desse tipo de anormalidade.

Metodologia

Nesta seção são descritas as etapas para elaboração de um banco de imagens, bem como as etapas exploradas para o desenvolvimento dos métodos computacionais propostos. Os principais estágios desses métodos estão relacionados às etapas de segmentação de núcleos celulares com modelos de DL, pós-processamento e a construção dos sistemas de comitês, conforme ilustrado na Figura 9. Os modelos de CNNs para segmentação e as técnicas para aumento de dados foram implementadas na linguagem Python, com as bibliotecas TensorFlow, PyTorch e Albumentation. Os sistemas de comitês e as avaliações experimentais foram implementadas com uso da linguagem MATLAB. Esses modelos foram treinados em um computador com processador Ryzen 5, GPU NVidia RTX 2070 e 32 GB de RAM.

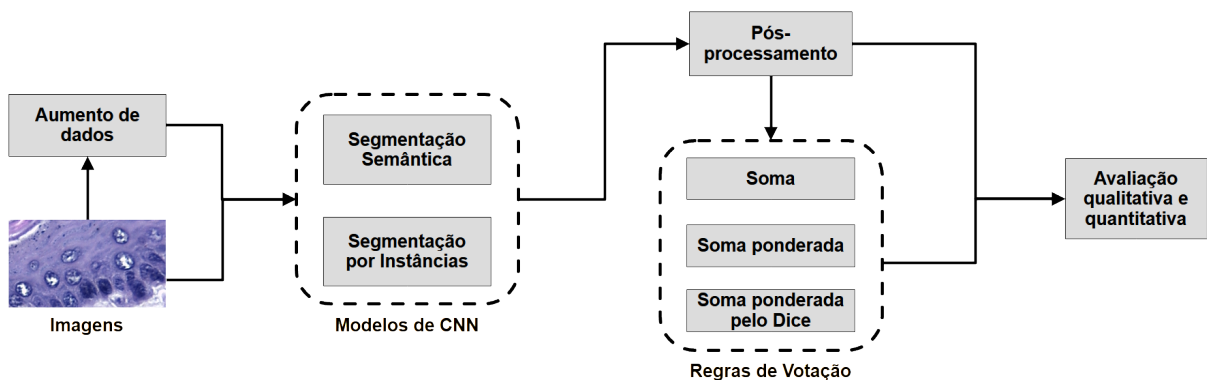


Figura 9 – Fluxograma com as principais etapas para a segmentação de núcleos celulares em tecidos histológicos de DEO.

4.1 Bancos de imagens

Neste estudo, um banco de imagens de tecidos de DEO foi construído para a aplicação da metodologia proposta. Esse banco foi construído a partir de lâminas contendo lesões induzidas experimentalmente em língua de camundongos com o uso de um carcinógeno. O experimento *in vivo* foi devidamente aprovado pelo Comitê de Ética na Utilização de

Animais sob número de protocolo 038/09 e realizado no período de 2009 a 2010, sendo utilizado nos trabalhos apresentados por (FARIA et al., 2011; SANT'ANA et al., 2011; MENDONÇA et al., 2012; SOUZA et al., 2014). Nesse protocolo de indução de lesões em língua de camundongos, 30 animais tiveram acesso ao carcinógeno por 16 semanas ininterruptas, sendo 50% sacrificados logo após esse período e os outros 50% sacrificados 16 semanas depois. Após o sacrifício, as línguas foram removidas, incluídas em parafina para confecção dos blocos, recortadas e coradas com H&E para estudo histopatológico. As lâminas foram classificadas em tecido saudável, DEO leve, DEO moderada e DEO severa, conforme descrito por Lumerman, Freedman e Kerpel (1995).

As imagens histológicas digitais foram obtidas com o uso do microscópio óptico Leica DM500, em magnificação de 400x, e armazenadas em arquivos em formato TIFF, utilizando o modelo de cores RGB, com quantificação em 8 *bits* e resolução de 2048×1536 *pixels*. Ao todo, 43 lâminas foram digitalizadas com a ajuda de um patologista. As imagens foram recortadas em ROIs com tamanho de 450×250 *pixels*, sendo 114 ROIs para cada classe de tecido, totalizando 456 ROIs. Os núcleos dessas imagens foram inicialmente anotados manualmente e seus contornos foram posteriormente validados e revisados por um patologista, conforme exibido na Figura 10. Esse processo resultou nas máscaras binárias nas quais a região de núcleos é colorida com a cor branca e o fundo do tecido é marcado na cor preta. Na Figura 11 são apresentados exemplos das ROIs obtidas de cada uma das classes investigadas neste estudo e seus respectivos padrão ouro.

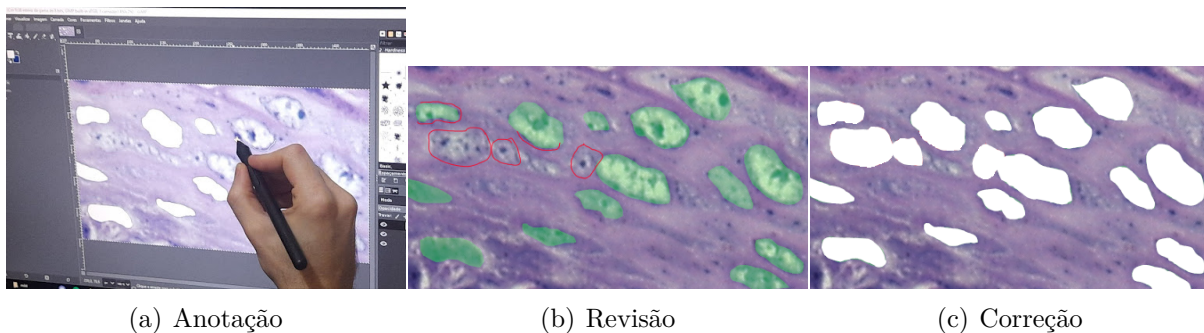


Figura 10 – Processo de criação do padrão-ouro de segmentação: (a) anotação manual, (b) revisões apontadas pelo patologista e (c) anotações revisadas.

Para avaliar os métodos computacionais propostos com os bancos de dados de domínio público, bem como para analisar seu desempenho em relação aos trabalhos da literatura, os bancos de imagens CryoNuSeg (MAHBOD et al., 2021), MoNuSeg 2018 (KUMAR et al., 2019) e GlaS (SIRINUKUNWATTANA et al., 2017) foram utilizados. O banco CryoNuSeg contém 30 imagens de tamanho 512×512 *pixels* de 10 órgãos humanos. O conjunto MoNuSeg 2018 é composto de 51 imagens de tumores de diferentes órgãos e com tamanho de 1000×1000 *pixels*. O conjunto de dados GlaS é composto por 165 imagens de tamanho 775×522 *pixels* de tecidos de câncer colorretal. Na Figura 12 são apresentados exemplos de imagens e marcações de padrão ouro desses bancos.

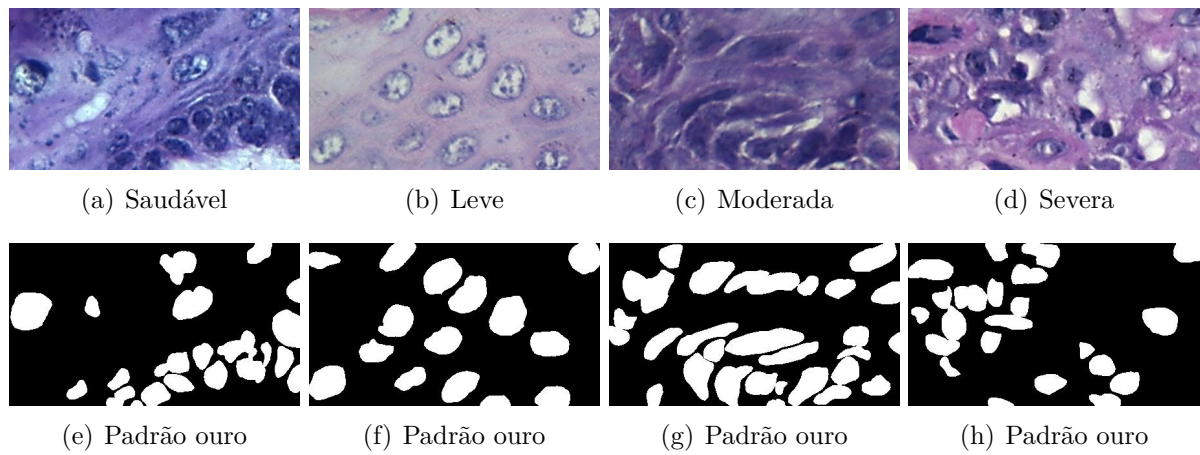


Figura 11 – Exemplos de tecidos histológicos utilizados na construção da base de dados: (a) tecido saudável, (b) DEO leve, (c) DEO moderada, (d) DEO severa, (e-h) máscaras do padrão ouro para cada classe das lâminas digitalizadas.

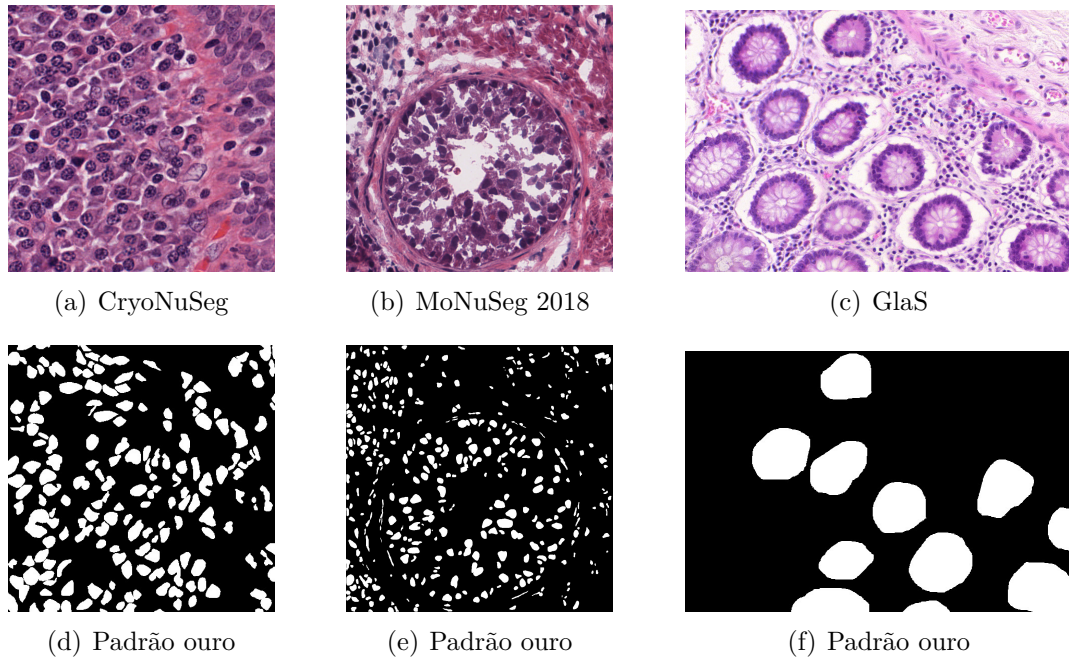


Figura 12 – Exemplos de tecidos histológicos da literatura utilizados para avaliação do método proposto: (a) MoNuSeg 2018, (b) CryoNuSeg, (c) GlaS, (d-f) máscaras do padrão ouro.

4.2 Aumento de dados

O aumento de dados é uma estratégia utilizada para ampliar os conjuntos de treinamento das RNAs, especialmente em cenários com escassez de imagens rotuladas, como a histopatologia. Além de mitigar problemas de *overfitting*, o aumento de dado contribui para o balanceamento das classes, favorecendo o aprendizado de características relevantes mesmo em categorias com menor representatividade (XUE et al., 2021). Neste trabalho, duas abordagens distintas de aumento de dados foram investigadas: uma baseada no uso

de GANs, com o método HistoGAN, que foi escolhida por ter sido desenvolvida especificamente para imagens histológicas e apresentar resultados relevantes (XUE et al., 2021). A segunda abordagem, chamada RCAug (SANTOS et al., 2023), aplica transformações geométricas e de aparência de forma dinâmica. Essa foi selecionada por apresentar bons resultados em imagens histológicas com carcinoma da região da cavidade oral (WARIN; SUEBNUKARN, 2024). Ambas as abordagens foram empregadas nos dados de treinamento, gerando duas vezes o número de imagens do conjunto original, os detalhes estão descritos na Seção 4.3.4. Um total de 1.041 imagens sintéticas foi obtido com cada abordagem, permitindo gerar conjuntos com 1.537 imagens.

4.2.1 HistoGAN

O modelo HistoGAN, proposto por Xue et al. (2021), foi composta por duas partes principais: o gerador, responsável por criar imagens sintéticas, e o discriminador, encarregado de distinguir entre as imagens reais e as imagens sintéticas.

O gerador é uma RNA que recebeu como entrada um vetor de ruídos aleatório e o utilizou para gerar imagens sintéticas dos tecidos histológicos. Esse processo envolveu uma sequência de camadas totalmente conectadas e camadas deconvolucionais, que expandem o vetor de características até alcançar a dimensão da imagem de saída. O discriminador, por sua vez, é composto por camadas convolucionais seguidas de normalização, que extraem mapas de características das imagens, e uma camada totalmente conectada no final para classificá-las como reais ou sintéticas.

O treinamento das redes, geradora e discriminadora, ocorreu de forma adversária. Cada imagem criada pelo gerador foi avaliada pelo discriminador, que avaliou as características em relação aos dados reais ou sintéticos. Durante o treinamento, os pesos de ambas as redes foram ajustados e, com isso, o gerador aprendeu a produzir imagens mais realistas, enquanto o discriminador melhorou sua capacidade de distinção. Esse processo iterativo resultou em imagens sintéticas cada vez mais próximas das reais, utilizadas para aumentar o conjunto de dados de treinamento. A etapa de treinamento do modelo HistoGAN foi realizado com 100 épocas, 50 iterações em cada época, *batch size* de tamanho 8 e taxa de aprendizado de 0,0001. Esses valores foram escolhidos com base em testes empíricos realizados com os dados reais.

4.2.2 RCAug

A segunda abordagem foi empregada com uso do método RCAug, proposto por Santos et al. (2023). Esse método aplica uma composição aleatória de transformações às imagens de entrada e seus respectivos rótulos. Em cada etapa, há uma chance de 50% (determinada empiricamente) de cada transformação ser aplicada, garantindo flexibilidade e diversidade no conjunto de treinamento sem comprometer o espaço de armazenamento.

As transformações geométricas foram aplicadas para reduzir erros de posicionamento e orientação em imagens histológicas. Essas transformações incluem rotação, transposição e *flips* em ambos os eixos, garantindo que o modelo seja robusto a variações espaciais presentes nos dados. Adicionalmente, para mitigar problemas relacionados a artefatos indesejados, como dobras no tecido, distorções ópticas, borrões e variações de coloração decorrentes de processos de coloração e condições de iluminação, foram utilizadas transformações específicas que minimizam esses erros e melhoram a qualidade das imagens. Finalmente, a transferência de cores foi aplicada, adaptando a aparência cromática de uma imagem de treinamento à de outra imagem-alvo, mantendo a estrutura original da imagem base. Essa abordagem amplia a diversidade cromática do conjunto de treinamento, permitindo que o modelo lide melhor com variações de coloração comuns em imagens histológicas. Essas transformações foram implementadas usando a biblioteca open-source Albumentations. Os detalhes estão disponíveis em (SANTOS et al., 2023).

4.3 Modelos de redes para segmentação de núcleos celulares

Neste estudo, foram utilizadas sete arquiteturas de CNNs para a segmentação dos núcleos celulares, divididas em dois grupos: (1) modelos baseados em semântica, compreendendo os modelos U-Net, FPN e PSPNet, e (2) modelos baseados em segmentação por instâncias, compreendendo as redes Mask R-CNN, StarDist, SOLOv2 e HoVer-Net. As CNNs foram escolhidas com base em resultados relevantes de segmentação encontrados na literatura (JADER et al., 2018; WANG et al., 2020; JI; WU, 2022; WEIGERT; SCHMIDT, 2022). Para ambos os grupos, os dados gerados pelos *backbones* foram utilizados para aprender as características dos núcleos celulares e realizar a segmentação dessas estruturas.

4.3.1 Arquiteturas de redes para extração de características

Neste trabalho, duas arquiteturas de CNN foram utilizadas para a extração das características dos núcleos celulares (chamada de *backbone*) a serem utilizadas pelos modelos segmentadores: a ResNet50 e a MobileNetV3. Essas arquiteturas foram escolhidas devido a características promissoras em tarefas de segmentação, conforme descrito na literatura (ARLOVA et al., 2022; KASSANI et al., 2022).

A ResNet50 é composta por 49 camadas convolucionais, divididas entre a camada de entrada e quatro blocos de operações de convolução, e a camada de classificação (HE et al., 2016). A camada de entrada contém 64 filtros de tamanho de 7×7 e deslocamento de 2 *pixels*. As próximas 48 camadas são divididas em blocos compostos por três camadas cada e que se repetem sucessivamente. Essas estruturas são nomeadas de Blocos 1, 2, 3 e

4. O Bloco 1 é composto por uma camada com 64 filtros de tamanho de 1×1 *pixel*, uma segunda camada com 64 filtros de tamanho de 3×3 e a terceira camada com 256 filtros de tamanho de 1×1 *pixel*. A primeira camada do Bloco 2 possui 128 filtros com tamanho de 1×1 , a segunda possui 128 com tamanho de 3×3 e a terceira é composta por 512 filtros com tamanho de 1×1 *pixel*.

No Bloco 3, a primeira camada possui 256 filtros com tamanho de 1×1 , a segunda possui 256 filtros de tamanho 3×3 e a terceira possui 1024 filtros de tamanho 1×1 . O Bloco 4 possui uma camada com 512 filtros de tamanho 1×1 , uma camada com 512 filtros de tamanho 3×3 e a terceira camada com 2048 filtros com tamanho de 1×1 *pixel*. A última etapa é uma camada totalmente conectada com 1000 neurônios com a função *softmax* para classificar os objetos entre as classes de núcleo ou região de fundo da imagem.

O Bloco 1, repete três vezes, o Bloco 2, repete quatro vezes, e os Blocos 3 e 4, repetem seis e três vezes, respectivamente. Um filtro de *max pooling* é aplicado entre cada bloco de convoluções e, antes da camada de classificação, um filtro de *average pooling* é aplicado para transformar os dados em um vetor de características. A arquitetura da ResNet50 pode ser vista na Figura 13.

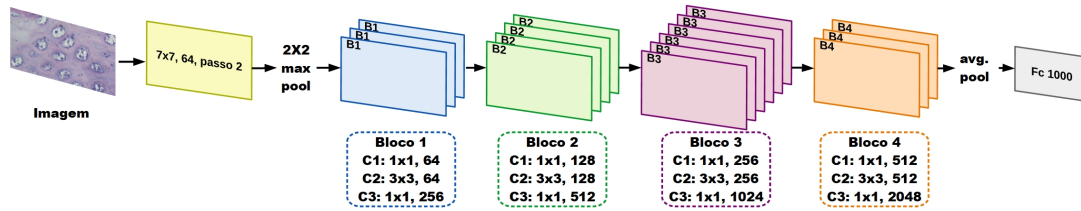


Figura 13 – Arquitetura da CNN *ResNet50* empregada como *backbone* para a segmentação de núcleos. Fonte: (SILVA, 2019).

O modelo MobileNet é composto por blocos de camadas convolucionais que geram características de maneira robusta e com baixo custo computacional (HOWARD et al., 2019). Cada camada é composta por um filtro convolucional de tamanho 1×1 , um filtro de tamanho 3×3 *pixels*, um filtro de *average-pooling* seguido de uma operação de deconvolução e, por fim, um filtro de tamanho 1×1 *pixel*. As operações de *pooling* e deconvolução geram características robustas enquanto que o último filtro 1×1 aumenta a quantidade de camadas geradas para os mapas de características. Um exemplo desse bloco é apresentado na Figura 14. Essa rede é composta pela camada de entrada, 15 blocos convolucionais, uma camada de *average-pooling* de tamanho 7×7 e convoluções de tamanho 1×1 .

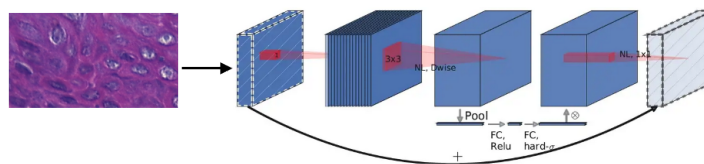


Figura 14 – Arquitetura da CNN *MobileNetV3* empregada como *backbone* para a segmentação de núcleos. Imagem adaptada de (HOWARD et al., 2019).

4.3.2 Arquiteturas de rede para segmentação semântica

As arquiteturas U-Net, FPN e PSPNet realizam segmentação semântica, rotulando cada *pixel* como pertencente ou não a região de núcleo. Nesta seção são apresentados os detalhes dos modelos dessas arquiteturas.

4.3.2.1 Modelo U-Net

O modelo U-Net é composto por um caminho de convolução (*encoder*) e um caminho de deconvolução (*decoder*). No *encoder*, as características das imagens são extraídas através de sucessivas camadas convolucionais (RONNEBERGER; FISCHER; BROX, 2015). Já o *decoder* amplia os mapas de características para o tamanho original da imagem, permitindo a localização precisa de objetos. Para isso, cada etapa desse caminho consiste na expansão dos mapas de características, concatenação com o mapa de características correspondente ao do caminho de convolução e aplicação de duas convoluções, cada uma seguida por uma ReLU. Conexões de atalho são utilizadas para concatenar os mapas de características correspondentes entre os dois caminhos, preservando informações espaciais detalhadas. Na camada final, uma convolução de tamanho 1×1 mapeia os vetores de características para o as classes desejadas. Um exemplo do caminho de *encoder* e *decoder* desse modelo é apresentado na Figura 15.

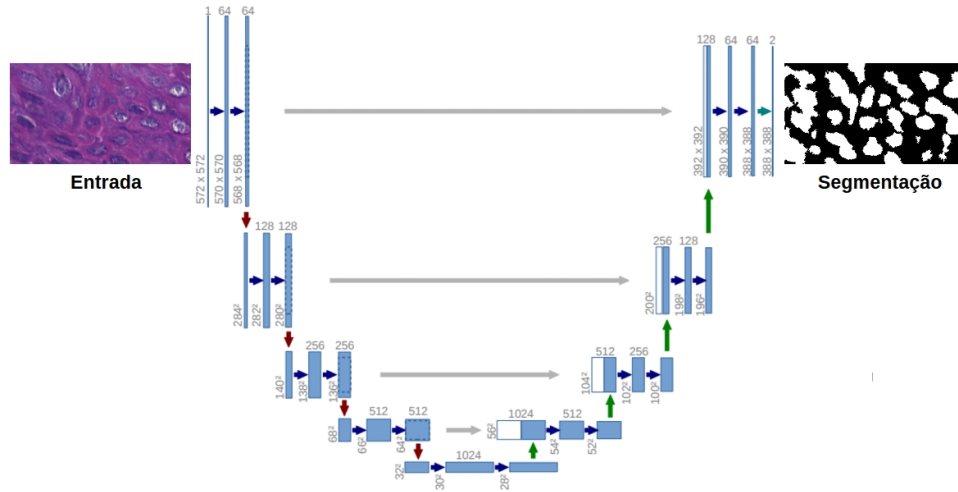


Figura 15 – Arquitetura do modelo U-Net usado para a segmentação de núcleos. Imagem adaptada de (RONNEBERGER; FISCHER; BROX, 2015).

Na combinação da U-Net com a ResNet50, nomeada como U-Net-ResNet50, as camadas convolucionais do *backbone* foram empregadas no *encoder* do modelo, excluindo a camada de classificação da ResNet50. As conexões de atalho foram criadas a partir das saídas intermediárias de cada bloco convolucional, abrangendo as resoluções de entrada progressivamente reduzidas pelo *backbone*. As saídas obtidas, após as convoluções e operações de *pooling* nos blocos da ResNet50, foram conectadas ao caminho *decoder* da U-Net. O *encoder* proposto foi composto por 49 camadas convolucionais provenientes da

ResNet50. O *decoder* foi composto por 49 etapas, cada uma contendo camadas, convolução transposta, para aumentar a resolução dos mapas de descritores (características), seguidas por convoluções duplas de tamanho 3×3 *pixels* com camada de ativação ReLU. A etapa final da segmentação foi obtida por uma convolução de tamanho 1×1 . Assim, o modelo apresentado foi composto por 99 camadas convolucionais.

Com a mesma abordagem, as 15 camadas convolucionais da MobileNetV3 foram aplicadas na etapa de *encoder*, criando a U-Net-MobileNet, em que foi aplicada para capturar as representações das características. Cada bloco do *encoder* foi composto por duas camadas convolucionais seguidas de uma operação de *pooling*, com as dimensões sendo reduzidas progressivamente. O *decoder* seguiu uma abordagem similar à da U-Net-ResNet50, com cada camada aumentando a resolução dos mapas e concatenando com as camadas do *encoder* correspondentes. Os blocos do *decoder* incluem camadas de *upsampling*, seguidas por duas convoluções para refinar os mapas. A última camada foi aplicada com uma convolução de tamanho 1×1 para produzir a saída binária. Este modelo integrado foi composto por 15 camadas da MobileNetV3 para o *encoder*, 15 para o *decoder* e a camada final de segmentação, totalizando 31 camadas.

4.3.2.2 Modelo FPN

O modelo *Feature Pyramid Network* (FPN) recebe uma imagem como entrada e gera mapas de características em vários níveis de forma totalmente convolucional (LIN et al., 2017). A construção da pirâmide é composta por um caminho *bottom-up*, um caminho *top-down* e conexões laterais. O caminho *bottom-up* utiliza os mapas de características gerados pelo *backbone* para criar mapas hierárquicos em diferentes escalas. O caminho *top-down* expande as características geradas nos níveis superiores da pirâmide. Essas características são refinadas por meio de conexões laterais, que mesclam os mapas de mesmo tamanho espacial dos caminhos *bottom-up* e *top-down*. Devido à sua natureza hierárquica e multiescala, a FPN gera características para segmentação de objetos em diversas escalas (KISANTAL et al., 2019). A arquitetura do modelo FPN está ilustrada na Figura 16.

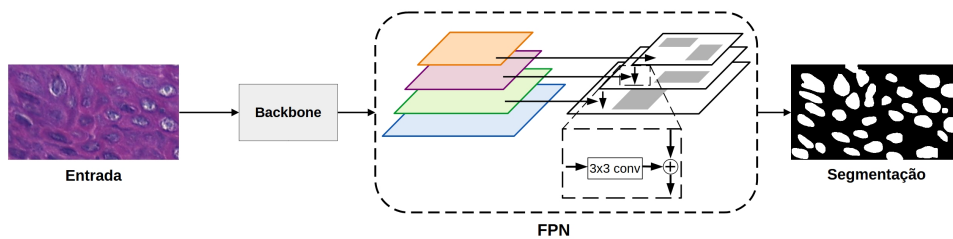


Figura 16 – Arquitetura do modelo FPN usado para a segmentação de núcleos. Imagem adaptada de (SILVA, 2019).

A combinação da FPN com a ResNet50, denominada FPN-ResNet50, utilizou as camadas intermediárias da ResNet50 como entradas para a construção da pirâmide de ca-

racterísticas. No caminho *bottom-up*, as saídas das camadas convolucionais da ResNet50 foram usadas para criar mapas hierárquicos em diferentes escalas, permitindo a extração de características em várias resoluções. O caminho *top-down* expandiu as características da camada superior da pirâmide e por meio de ajuste permitiu combinar com as dimensões das camadas inferiores através de operações de *upsampling* seguida de convoluções de tamanho 1×1 . Filtros de tamanho 3×3 foram empregados em cada nível da pirâmide para refinar as informações extraídas. Em seguida, os mapas de características foram aumentados por operações *upsampling* para recuperar a resolução original da imagem. As saídas dos diferentes níveis da pirâmide foram somados e uma camada convolucional de tamanho 1×1 foi empregada para gerar a segmentação final da imagem. Esse modelo consistiu em 62 camadas convolucionais, sendo 49 camadas da ResNet50, 12 camadas adicionais no *decoder* e uma camada final de segmentação.

Na integração com a MobileNetV3, chamada de FPN-MobileNet, a pirâmide de características foi construída a partir das camadas intermediárias do *backbone*. No caminho *bottom-up*, as saídas das camadas foram utilizadas para gerar características em diferentes resoluções, gerando a pirâmide. As conexões laterais entre as diferentes escalas foram realizadas por meio de somas, seguidas de convoluções de tamanho 3×3 para ajustar as características de cada nível. Após esse ajuste das características, as saídas do caminho *top-down* e as características extraídas pelo *backbone* foram aumentadas por operações de *upsampling* para se ajustarem à resolução original da imagem. A pirâmide gerada foi então concatenada e refinada por uma convolução com filtro 1×1 para gerar as máscaras de segmentação. O modelo final foi composto por 27 camadas convolucionais, sendo 15 camadas do *backbone*, e 12 para a construção da pirâmide e refinamento de saídas.

4.3.2.3 Modelo PSPNet

A *Pyramid Scene Parsing Network* (PSPNet) é composta por processos de *encoder* e *decoder*. O *encoder* é responsável por extrair recursos da imagem em diferentes níveis, enquanto o caminho *decoder* gera a saída de segmentação a partir da agregação desses recursos em várias escalas (ZHAO et al., 2017). No *encoder*, as camadas do modelo de *backbone* foram empregadas para criar uma pirâmide de descritores (características) em múltiplas escalas. No *decoder*, os mapas de características foram expandidos, concatenados com os mapas correspondentes do *encoder* e refinados por meio de duas convoluções. Essa concatenação permitiu a combinação de informações de várias escalas, aprimorando a segmentação semântica. Por fim, uma camada de classificação produziu as máscaras binárias. Na Figura 17, é apresentado a arquitetura deste modelo de rede.

Na versão combinada com a ResNet50, chamada PSPNet-ResNet50, as camadas convolucionais do *backbone* foram adotadas até o último bloco, excluindo a camada de classificação. A partir das saídas intermediárias, o módulo de pirâmide de *pooling* foi implementado com quatro níveis de redução. Cada saída do módulo de *pooling* foi processada

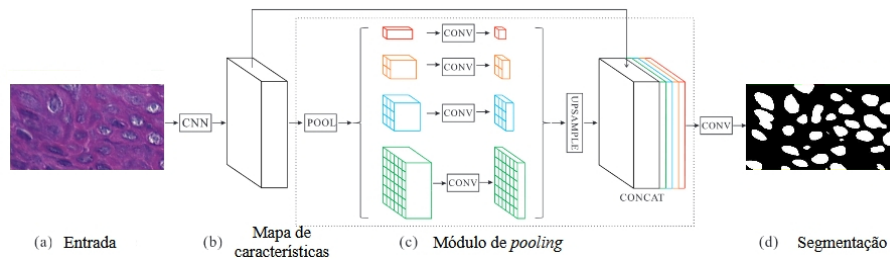


Figura 17 – Arquitetura do modelo PSPNet usado para a segmentação de núcleos. Imagem adaptada de (ZHAO et al., 2017).

por uma convolução de tamanho 1×1 seguida de uma ativação ReLU, garantindo uniformidade na dimensão dos mapas de características. Essas saídas foram então concatenadas e processadas por uma convolução que gerou o mapa integrado para o *decoder*. Nele, as resoluções dos mapas foram ampliadas progressivamente por meio de operações de convolução transposta e ativações ReLU para refinar os mapas. Esse processo foi repetido em quatro etapas, seguido de uma camada totalmente conectada que gerou a máscara de segmentação. O modelo proposto foi composto de 115 camadas convolucionais, sendo 49 camadas da ResNet50 e 66 camadas adicionais na PSPNet.

Na associação com a MobileNetV3, denominada PSPNet-MobileNet, as 15 camadas convolucionais do *backbone* foram utilizadas para extrair características da imagem. O módulo de pirâmide de *pooling* também foi implementado com quatro níveis de redução e as saídas do módulo de *pooling* foram processadas por uma convolução de tamanho 1×1 seguida de uma ativação ReLU. Após a concatenação das saídas do módulo de pirâmide de *pooling*, uma convolução foi empregada para refinar o mapa agregado. No *decoder*, as resoluções dos mapas foram aumentadas por operações de *upsampling* bilinear em duas etapas principais, seguidas por convoluções de tamanho 3×3 e ativações ReLU. Uma convolução de tamanho 1×1 foi responsável por gerar as máscaras de segmentação. Esse modelo foi composto por 56 camadas, sendo 15 da MobileNetV3 e 41 adicionais da PSPNet.

4.3.3 Arquiteturas de rede para segmentação via instâncias

As arquiteturas Mask R-CNN, StarDist, SOLOv2 e HoVer-Net realizam segmentação de instâncias dos núcleos celulares, identificando cada núcleo como um objeto individual.

4.3.3.1 Modelo Mask R-CNN

A rede Mask R-CNN foi desenvolvida para detecção de objetos e segmentação de instâncias (HE et al., 2017). Este *framework* é construído em uma arquitetura de CNN para extração de características, que emprega caixas delimitadoras para realizar a segmentação de instâncias, apresentando contribuições relevantes em imagens médicas (WANG et al., 2019; LONG et al., 2021). O modelo Mask R-CNN utilizado neste estudo consiste na

ResNet50, uma FPN e uma rede para proposta de regiões (do inglês, *Region Proposal Network* (RPN)).

No primeiro estágio, as camadas convolucionais do *backbone* ResNet50 foram usadas para construir uma estrutura FPN, de forma similar à descrita na Seção 4.3.2.2. As características obtidas pela FPN foram usadas como entrada para a RPN, cujo objetivo é identificar objetos candidatos. Uma camada convolucional de tamanho 3×3 foi empregada em cada camada do FPN e os valores resultantes foram empregados em duas camadas totalmente conectadas, gerando as caixas delimitadoras para cada região. Em seguida, uma camada totalmente conectada foi aplicada nos mapas de características dos objetos para definir as máscaras binárias para cada núcleo da imagem histológica. No total, esse modelo consistiu de 61 camadas, considerando as 50 do *backbone* ResNet50, 6 camadas adicionais da RPN, e 5 para gerar as caixas delimitadores e a máscara de segmentação. A Figura 18 apresenta a arquitetura desse modelo.

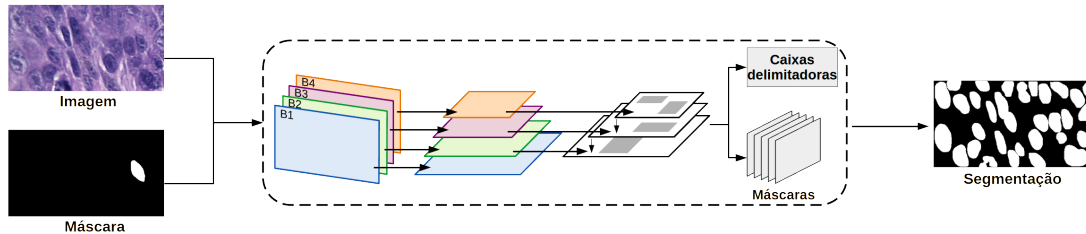


Figura 18 – Arquitetura da rede Mask R-CNN usada na segmentação de núcleos. Imagem adaptada de (SILVA, 2019).

4.3.3.2 Modelo StarDist

A arquitetura StarDist consistiu em duas etapas principais: a detecção de objetos e a segmentação das instâncias (WEIGERT; SCHMIDT, 2022). Na primeira etapa, o *backbone* U-Net foi empregado para extrair características de diferentes escalas da imagem de entrada, utilizando seus caminhos de *encoder* e *decoder*. Após essa etapa, o modelo utilizou uma abordagem baseada em distâncias radiais para identificar os objetos presentes na imagem. Para cada objeto detectado, foi gerada uma estrela, que é uma representação geométrica composta por um conjunto de distâncias poligonais do centroide do objeto até sua borda, em várias direções radiais. O número de direções radiais n selecionado foi de 64, de acordo com testes empíricos e, para cada direção, a distância do centroide até o limite do objeto foi calculada, gerando um mapa de distâncias correspondentes, conforme ilustrado na Figura 19.

Na segunda etapa, as características geradas pelo *encoder* e *decoder* da U-Net, juntamente com os mapas de distâncias radiais, foram passadas por três camadas convolucionais. A primeira camada transformou os mapas de características em vetores, enquanto que a segunda camada foi empregada para realizar o cálculo das distâncias radiais. Fi-

nalmente, na terceira camada obteve as máscaras binárias para cada núcleo da imagem, definindo, assim, a segmentação das instâncias. A arquitetura StarDist foi composta por 22 camadas convolucionais da U-Net, excluindo a FCN e a camada de segmentação final, além das três camadas adicionais para refinar as características, identificar objetos e calcular as distâncias em polígonos, e realizar a segmentação. Isso resultou em um total de 25 camadas no modelo.

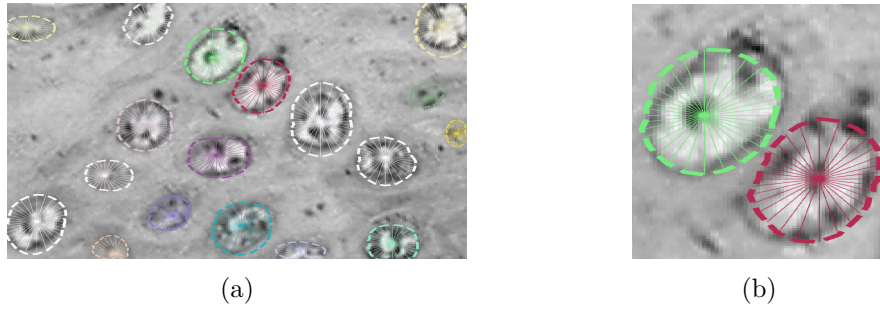


Figura 19 – Exemplos de polígonos gerados pelo modelo StarDist para determinar a área dos núcleos celulares. Imagem elaborada pelo autor.

4.3.3.3 Modelo HoVer-Net

A arquitetura HoVer-Net (GRAHAM et al., 2019) realizou a segmentação dos núcleos celulares utilizando uma arquitetura baseada em um caminho *encoder* e três caminhos *decoder*. O *encoder* integrou as 49 camadas convolucionais da ResNet50 como *backbone*. Os mapas de características de diferentes resoluções foram processados por camadas convolucionais adicionais nos *decoders*, que refinaram as características extraídas e restauraram a resolução espacial, desempenhando um papel crucial para a segmentação.

A arquitetura ainda possui três caminhos distintos para *decoder*, cada um com uma função específica. O primeiro foi responsável por identificar e segmentar regiões nucleares, gerando máscaras de padrão semântico, o qual foi responsável por destacar as bordas dos núcleos. O segundo calculou as distâncias horizontal e vertical de cada pixel do núcleo ao centroide, produzindo os mapas de distâncias horizontais e verticais (Ho-Ver) que foram usados para separar núcleos em contato. O terceiro foi responsável pela classificação dos núcleos, diferenciando entre diferentes tipos celulares. Essa etapa não foi empregada no presente estudo, pois na versão original o objetivo era classificação e este estudo explorou apenas a tarefa de segmentação. Cada *decoder* empregou seis camadas convolucionais, somando 12 camadas para os dois *decoders* utilizados. Ao combinar as 49 camadas convolucionais da ResNet50 e as camadas dos *decoders*, a arquitetura HoVer-Net apresentou um total de 61 camadas convolucionais, utilizadas para realizar a segmentação via instâncias, conforme ilustrado na Figura 20.

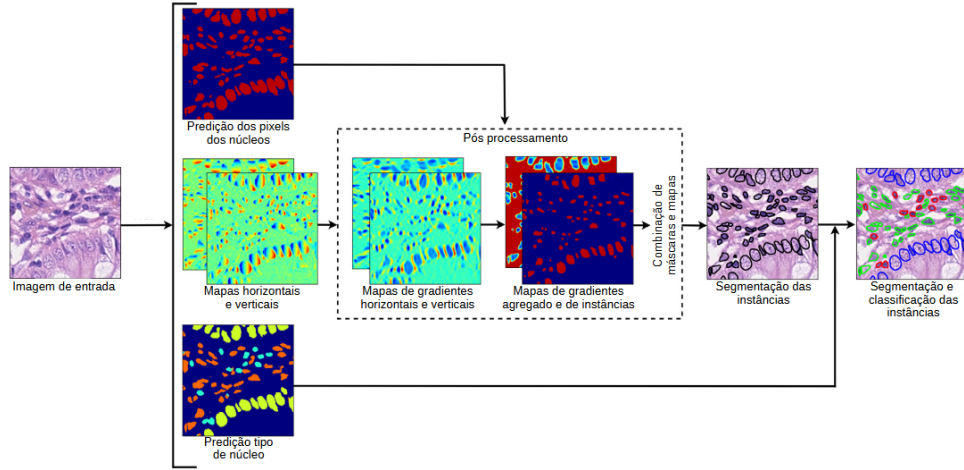


Figura 20 – Estrutura da arquitetura HoVer-Net usado para segmentação via instâncias. Imagem adaptada de (GRAHAM et al., 2019).

4.3.3.4 Modelo SOLO v2

A arquitetura SOLOv2 é uma evolução do modelo *Segmenting Objects by Locations* (SOLO), composto por dois estágios: o primeiro para a identificação semântica dos objetos baseada em regiões e o segundo para o refinamento de máscaras binárias (WANG et al., 2020). O modelo integrou a FPN com uma ResNet50 como *backbone* para a extração de características multi-escala das imagens, proporcionando uma representação das características em várias resoluções. Um total de 49 camadas convolucionais da ResNet50 foram utilizadas para extrair características em múltiplos níveis, que foram empregadas nas etapas subsequentes de segmentação. A combinação das redes permitiu a construção de um modelo que possibilita capturar características globais e locais da imagem.

No primeiro estágio, a imagem foi dividida em uma grade uniforme de tamanho $S \times S$, e cada célula da grade foi classificada como pertencente a um núcleo ou à região de fundo. Em seguida, no segundo estágio, o modelo extraiu as informações necessárias para a criação das máscaras de instâncias a partir dos mapas de características gerados pela FPN. Essas informações foram compostas por dois conjuntos principais: os filtros de máscaras G e as características F . Para a geração dos filtros G , a rede aplicou operações convolucionais em diferentes níveis, sendo um filtro de tamanho 3×3 empregado para reduzir a dimensionalidade dos mapas de características e gerar as máscaras das instâncias. Em paralelo, os descritores (*features*) F foram obtidas por meio da concatenação dos mapas de características provenientes da FPN, seguidos da aplicação de filtros, tamanho 3×3 , e a função de ativação ReLU. A máscara final para cada instância foi gerada pela combinação de G e F por meio de uma camada adicional de tamanho 1×1 , e mais uma camada ReLU. Essas etapas permitiram obter uma segmentação das instâncias dos núcleos presentes na imagem. A arquitetura SOLOv2 foi composta por 62 camadas convolucionais no total, incluindo 49 camadas da ResNet50, 10 camadas da FPN, além das três camadas convolucionais adicionadas para gerar as máscaras de instâncias. A

ilustração desse modelo é apresentada na Figura 21.

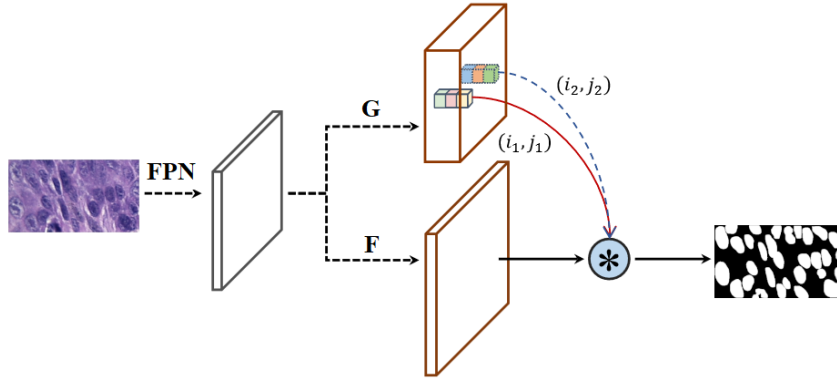


Figura 21 – Arquitetura do modelo SOLOv2 usado para a segmentação de núcleos. Imagem adaptada de (WANG et al., 2020).

4.3.4 Treinamento dos modelos de segmentação

Todos os modelos investigados foram pré-treinados na base ImageNet e a etapa de ajuste foi realizada utilizando três configurações de dados: a base de imagens sem aumento, denominada de (*baseline*), a base aumentada pelo método da HistoGAN e a base aumentada pela técnica RCAug. Para cada modelo, um conjunto específico de hiperparâmetros foi definido e utilizado para as três configurações de dados. Esses valores são apresentados na Tabela 2.

Tabela 2 – Configurações de treinamento dos modelos analisados.

Modelo	<i>Backbone</i>	Iterações	<i>Batch Size</i>	Aprendizado	Otimizador
FPN-ResNet50	ResNet 50	100	4	0,0001	Nadam
FPN-MobileNet	MobileNet V3	100	4	0,0001	Nadam
PSP-ResNet50	ResNet 50	100	4	0,0001	Nadam
PSP-MobileNet	MobileNet V3	100	4	0,0001	Nadam
UNet-ResNet50	ResNet 50	100	4	0,0001	Nadam
UNet-MobileNet	MobileNet V3	100	4	0,0001	Nadam
Mask R-CNN	ResNet 50	150	1	0,001	Nadam
SOLOv2	FPN / ResNet 50	100	1	0,001	Adam
HoVer-Net	ResNet 50	100	1	0,001	SGD
StarDist	U-Net	200	8	0,001	Adam

Os modelos de segmentação semântica utilizaram o mesmo número de iterações por época, enquanto os modelos de segmentação via instâncias Mask R-CNN e StarDist exigiram ajustes nesse parâmetro devido à sua configuração de processar apenas um núcleo por iteração. Além disso, os modelos semânticos empregaram valores idênticos de *batch size* e taxa de aprendizado. Para os modelos de instâncias, o *batch size* foi fixado em um, devido a limitações de hardware, mas esse número foi fixado em oito para o StarDist. Os modelos de segmentação de instâncias utilizaram taxa de aprendizado maior do que os modelos semânticos por tratarem cada núcleo como um objeto de treinamento,

o que pode contribuir para uma convergência mais eficiente. O otimizador *Nadam* foi empregado para a maioria dos modelos com base em estudos prévios (SILVA et al., 2020; SILVA et al., 2022). No entanto, os modelos SOLOv2, HoVer-Net e StarDist empregaram outros otimizadores, selecionados por apresentarem melhor desempenho durante os testes empíricos. Independente das diferenças nos hiperparâmetros, todos os modelos foram treinados durante 400 épocas, com *momentum* fixado em 0,9 e utilizando a função de perda *binary cross entropy*.

Para os treinamentos, a base OralEpitheliumDB foi dividida em 70% para treinamento, 10% para validação e 20% para teste. As técnicas de aumento de dados foram aplicadas exclusivamente ao conjunto de treinamento, preservando a integridade dos conjuntos de validação e teste, garantindo que esses dados não interferissem no processo de ajuste dos modelos. Com base no desempenho dos modelos de aumento de dados, as bases públicas foram treinadas apenas com aumento de dados via RCAug, mantendo as proporções de divisão das imagens.

A capacidade de generalização dos modelos também foi avaliada utilizando validação cruzada (*k-fold*). Esse método dividiu o conjunto de dados em k partes iguais (chamadas de *folds*), permitindo que o modelo fosse treinado k vezes. Em cada iteração, um *fold* foi utilizado como conjunto de validação, enquanto os $k - 1$ restantes foram usados para treinamento. Os resultados das k iterações foram então agregados, proporcionando uma estimativa mais robusta do desempenho do modelo e minimizando os vieses associados a uma única divisão de dados. Neste estudo, foi utilizado $k = 5$, o que resultou na divisão dos dados em cinco partes iguais, com cada parte sendo usada como validação uma vez.

4.4 Pós-processamento em modelos de rede para segmentação

Após a etapa de segmentação, as máscaras geradas pelos modelos de CNN podem apresentar falhas nos contornos, regiões não completamente preenchidas e artefatos, que podem comprometer análises dessas imagens resultantes. Para aprimorar essas máscaras e garantir maior precisão para a subsequente composição dos comitês, foi realizado um processo de refinamento e limpeza, envolvendo operações morfológicas de exclusão.

O primeiro passo desse processo foi aplicado uma operação de dilatação para preencher falhas nos contornos das estruturas celulares. Essa técnica expandiu os limites das regiões segmentadas utilizando um elemento estruturante em forma de cruz, com tamanho de 3×3 *pixels*. O uso desse elemento foi determinado com base em testes empíricos e devido a capacidade de preservar a morfologia geral dos objetos ao mesmo tempo que conecta regiões fragmentadas.

Para tratar regiões com falsos negativos de núcleos parcialmente preenchidos, foi empregada uma função de preenchimento de regiões. Essa função identifica regiões fechadas

por contornos e preenche todos os *pixels* internos com o mesmo valor de intensidade do contorno, garantindo uma representação mais completa das estruturas celulares segmentadas.

Para tratar ruídos e restaurar o tamanho inicial dos núcleos após a dilatação e o preenchimento, uma operação de erosão foi aplicada com um elemento estruturante em forma de cruz de 3×3 . Essa etapa elimina artefatos residuais ao mesmo tempo que restaura o tamanho original das estruturas segmentadas. Com base em análises empíricas, se verificou que a aplicação do filtro de erosão, após o preenchimento de regiões, produz menos falsos negativos do que aplicado numa etapa anterior, otimizando assim o equilíbrio entre remoção de ruídos e preservação das estruturas biológicas.

Por fim, para eliminar artefatos que poderiam ser erroneamente interpretados como núcleos, todos os objetos na imagem com área inferior a 30 *pixels* foram rotulados como região de fundo. Esse tamanho foi determinado experimentalmente, com base em análises do tamanho dos núcleos em relação aos artefatos menores. Essa exclusão sistemática contribui para reduzir os falsos positivos e melhorar a qualidade das máscaras segmentadas. As etapas empregadas no pós processamento são ilustradas na Figura 22.

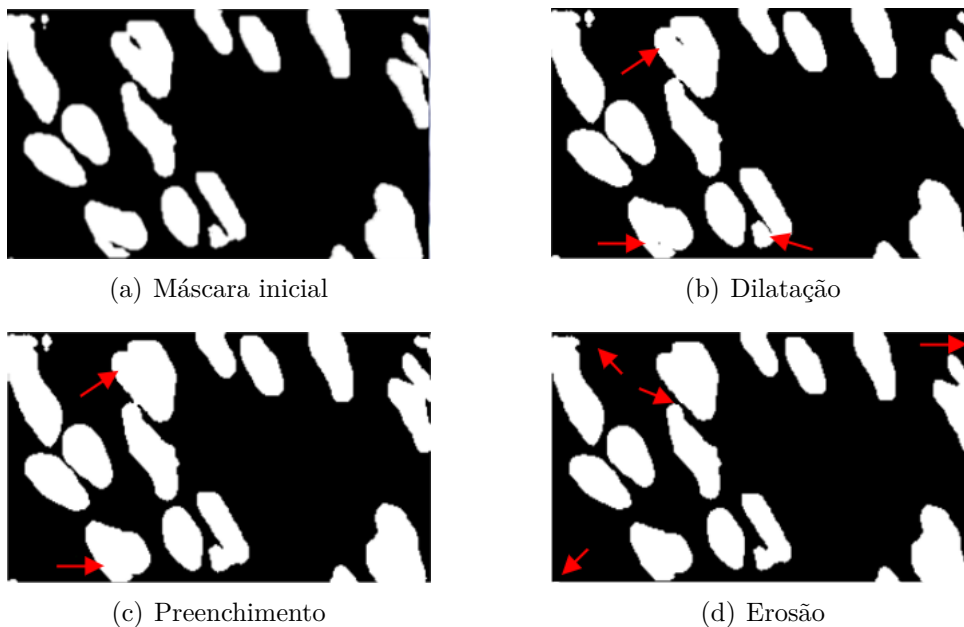


Figura 22 – Etapas empregadas para o pós processamento: (a) máscara inicial, (b) máscara após a operação de dilatação, (c) máscara após a operação de preenchimento de regiões e (d) máscara resultante após as operações de erosão e remoção de artefatos. As setas em vermelho indicam as alterações resultantes das operações.

4.5 Sistemas de comitês de segmentadores

A exploração de comitês se mostra como uma abordagem promissora para melhorar a precisão e a robustez em modelos de DL (LOPES et al., 2024). A combinação de diferentes modelos de CNNs pode gerar resultados mais confiáveis, aproveitando os pontos fortes de cada modelo. As regras de votação são estratégias utilizadas para combinar os resultados dos modelos que compõem um comitê. Neste estudo, foram aplicadas as regras da soma e da soma ponderada.

A regra da soma é uma técnica de combinação de modelos amplamente utilizada em problemas de rotulação com múltiplos modelos de DL. Nessa regra, os resultados dos modelos são somados e a decisão final é baseada no resultado da soma (SAGI; ROKACH, 2018). Se o resultado da soma for maior que um certo limiar T , o *pixel* é classificado como região de objeto; caso contrário, é classificado como região de fundo. Essa abordagem é útil para dados binários. A equação que define a regra é dada por:

$$S(x, y) = \sum_{i=1}^N m_i(x, y), \quad (6)$$

em que $S(x, y)$ é o resultado da soma para o *pixel* na coordenada (x, y) , N é o número de modelos e $m_i(x, y)$ é a máscara gerada pelo modelo i .

A regra da soma ponderada é uma extensão da regra da soma, em que os resultados dos modelos são ponderados por um peso (NANNI; LUMINI; GHIDONI, 2018). Os pesos podem ser atribuídos com base no desempenho do modelo em um conjunto de validação ou em outras métricas específicas. Essa regra é definida pela equação:

$$S_P(x, y) = \sum_{i=1}^N w_i m_i(x, y), \quad (7)$$

em que $S_P(x, y)$ é o resultado ponderado para o *pixel* na coordenada (x, y) , w_i é o peso atribuído ao modelo i e $m_i(x, y)$ é a máscara gerada por ele.

Para ambas as regras, os valores dos *pixels* são comparados a um limiar T . Caso o valor seja maior ou igual ao limiar, o *pixel* é considerado como região de núcleo, caso contrário, é considerado como região de fundo. Essa lógica é expressa na Equação 8:

$$M(x, y) = \begin{cases} 1 & \text{se } R_V(x, y) > T \\ 0 & \text{caso contrário,} \end{cases} \quad (8)$$

em que $M(x, y)$ é a máscara final, R_V representa a regra de votação empregada e T é o limiar.

Essas regras foram exploradas utilizando um valor n variando de dois até o número máximo de modelos, de forma a avaliar o desempenho de diferentes composições de comitês. A soma ponderada foi avaliada utilizando dois esquemas de pesos baseados no D_C . No primeiro, os $n/2$ modelos com maior valor para a métrica receberam um peso

igual a 2, enquanto os demais receberam peso 1. Caso $n/2$ fosse ímpar, o valor foi arredondado para o menor valor. Já no segundo esquema, o próprio valor de D_C obtido por cada modelo foi utilizado como seu peso, eliminando a categorização de pesos gerado pelo primeiro esquema e permitindo que a influência de cada modelo no comitê fosse proporcional ao seu desempenho. A métrica D_C foi empregada como peso para a soma ponderada por ser amplamente utilizada em problemas de segmentação (KIRAN et al., 2022; LIAO; ZHANG, 2023; FIRAT et al., 2024).

Para a regra da soma, o limiar T foi definido como 50% do número total de modelos no comitê, arredondado para baixo caso o número fosse ímpar. Por exemplo, em um comitê com três modelos, o valor de T é 1, enquanto em um comitê com quatro modelos, T é 2. No caso das somas ponderadas, o valor T foi definido como 50% do somatório dos pesos dos modelos compondo o comitê. Esse valor de 50% foi escolhido com base em testes emíricos. Esse calculo foi feito realizado seguindo a equação:

$$T = 0,5 \times \left(\sum_{i=1}^N w_i \right), \quad (9)$$

em que w_i representa o peso atribuído ao modelo i .

Essas regras de comitês foram exploradas em três abordagens: comitês de modelos de segmentação semântica, comitês de segmentação via instâncias e comitês híbridos. Na abordagem de comitês semânticos, apenas os modelos de segmentação semântica foram considerados para compor os comitês, conforme exibido no Algoritmo 1. As máscaras preditas por essas redes foram combinadas utilizando as regras da soma, soma ponderada e soma ponderada pelo D_C e variando o número n de modelos no comitê de dois até o máximo de seis modelos disponíveis. Apenas os modelos U-Net-ResNet50, U-Net-MobileNet, FPN-ResNet50, FPN-MobileNet, PSPNet-ResNet50 e PSPNet-MobileNet foram investigados nessa etapa. Essa abordagem teve como objetivo explorar a combinação entre diferentes arquiteturas semânticas, buscando aprimorar a da segmentação de regiões de interesse. Por combinar imagens binárias, as regras puderam ser aplicadas como operações aritméticas de soma, multiplicação e divisão diretamente sobre os *pixels* das máscaras.

Na abordagem de comitês de segmentação via instâncias, os modelos Mask R-CNN, StarDist, HoVer-Net e SOLOv2 foram combinados para produzir segmentações no formato de instâncias. Semelhante à abordagem semântica, as máscaras geradas por esses modelos foram combinadas utilizando as regras de soma, soma ponderada e soma ponderada pelo D_C , com o n variando de dois até o máximo de quatro. As máscaras de instâncias geradas por esses modelos atribuem em valores distintos a cada núcleo e os modelos podem identificar núcleos em ordens diferentes, resultando em valores diferentes para o mesmo núcleo em cada modelo. Por essa razão as regras de votação não podem ser empregadas de maneira direta sobre os *pixels*. Assim, foi necessário identificar os núcleos correspondentes em diferentes modelos e determinar quais deles representavam a mesma região. Como descrito no Algoritmo 2, esse processo foi realizado calculando as caixas

Algoritmo 1 Comitê de segmentadores semânticos

Entrada: Regras de votação $R \in \{\text{Soma}, \text{SomaPonderada}, \text{SomaPonderadaPorDC}\}$, Máscaras semânticas $\{M_1, M_2, \dots, M_N\}$, Lista de pesos $\{w_1, w_2, \dots, w_N\}$, Lista de desempenhos $\{DC_1, DC_2, \dots, DC_N\}$, Limiar T
Saída: Máscara final segmentada M_{final}

```

1: se  $R = \text{Soma}$  então
2:    $M_{\text{comite}} \leftarrow \sum_{i=1}^N M_i$ 
3: senão
4:   se  $R = \text{SomaPonderada}$  então
5:     Ordenar  $\{DC_i\}$  em ordem decrescente
6:     Atribuir peso  $w_i = 2$  para os  $\frac{N}{2}$  modelos com maior  $DC$ 
7:     Atribuir peso  $w_i = 1$  para os demais modelos
8:   senão se  $R = \text{SomaPonderadaPorDC}$  então
9:      $w_i \leftarrow DC_i, \forall i$ 
10:   fim se
11:    $M_{\text{comite}} \leftarrow \sum_{i=1}^N w_i \cdot M_i$ 
12: fim se
13:
14: para cada pixel  $(x, y)$  em  $M_{\text{comite}}$  faça
15:   se  $M_{\text{comite}}(x, y) \geq T$  então
16:      $M_{\text{final}}(x, y) \leftarrow 1$ 
17:   senão
18:      $M_{\text{final}}(x, y) \leftarrow 0$ 
19:   fim se
20: fim para
21: retorne  $M_{\text{final}}$ 

```

delimitadoras de cada núcleo e determinando a sobreposição entre os modelos. Testes empíricos mostraram que uma sobreposição de 30% proporcionou os melhores resultados, com valores superiores gerando muitas regiões de FN e valores inferiores gerando regiões de FP. Uma vez identificados os núcleos correspondentes em todos os modelos, o valor de 1 foi atribuído a eles e as regras de votação foram aplicadas de maneira semelhantes ao dos comitês semânticos. Após essa etapa, cada instância da máscara final recebeu um valor inteiro único e crescente, mantendo a individualidade de cada instância identificada.

Algoritmo 2 Comitê de segmentadores via instâncias

Entrada: Regras de votação $R \in \{\text{Soma}, \text{SomaPonderada}, \text{SomaPonderadaPorDC}\}$, Máscaras de instâncias $\{M_1, M_2, \dots, M_N\}$, Lista de pesos $\{w_1, w_2, \dots, w_N\}$, Lista de desempenhos $\{DC_1, DC_2, \dots, DC_N\}$, Limiar de sobreposição S , Limiar T
Saída: Máscara de instâncias final M_{final}

```

1:  $ID \leftarrow 1$  {Inicializa um identificador único para os núcleos}
2: para cada máscara  $M_i$  faça
3:   para cada objeto  $O_{i,j}$  em  $M_i$  faça
4:     Calcular a caixa delimitadora  $B_{i,j}$ 
5:   fim para
6: fim para
7: para cada par  $(O_{i,j}, O_{k,l})$  com  $i \neq k$  faça
8:   Calcular a sobreposição  $(B_{i,j}, B_{k,l})$ 
9:   se sobreposição  $> S$  então
10:    Considerar  $O_{i,j}$  e  $O_{k,l}$  como a mesma instância
11:   fim se
12: fim para
13:
14: para cada grupo de instâncias correspondentes faça
15:   Atribuir o valor 1 a todas as instâncias do grupo
16:   Aplicar a regra  $R$  às instâncias conforme Algoritmo 1
17:   Atribuir  $ID$  ao objeto segmentado {Cada núcleo recebe um valor distinto}
18:    $ID \leftarrow ID + 1$ 
19: fim para
20: retorne  $M_{\text{final}}$ 

```

Já a abordagem híbrida combinou modelos semânticos e por instâncias em um único comitê, adotando dois métodos para a integração das máscaras, como descrito no Algoritmo 3. No primeiro método, denominado comitê híbrido semântico, as máscaras geradas pelos modelos de instâncias foram binarizadas, transformando qualquer valor maior do que zero em um. Isso permitiu que essas máscaras fossem combinadas com as máscaras semânticas utilizando regras de votação aplicadas diretamente sobre os *pixels* das imagens. No segundo método, denominado comitê híbrido via instâncias, as máscaras semânticas foram integradas às máscaras de instâncias empregando o cálculo de sobreposição de objetos. Nesse método, cada núcleo identificado pelos modelos via instâncias foi comparado às máscaras semânticas para determinar sobreposições. Após essa etapa, o valor de um foi atribuído a cada núcleos e as regras de votação foram aplicadas, com valores crescentes sendo atribuídos aos núcleos resultantes, gerando um mapa final de segmentação de instâncias. Esses métodos híbridos permitiram avaliar os comitês combinando modelos que utilizam diferentes abordagens para identificação e segmentação de objetos.

Algoritmo 3 Comitê híbrido de segmentação

Entrada: Regra de votação $R \in \{\text{Soma}, \text{SomaPonderada}, \text{SomaPonderadaPorDC}\}$, Máscaras semânticas $\{M_1^{sem}, M_2^{sem}, \dots, M_M^{sem}\}$, Máscaras de instâncias $\{M_1^{inst}, M_2^{inst}, \dots, M_N^{inst}\}$, Lista de desempenhos $\{DC_1, DC_2, \dots, DC_N\}$, Lista de pesos $\{w_1, w_2, \dots, w_N\}$, Limiar de sobreposição S , Limiar T

Saída: Máscara híbrida final M_{final}

```

1: Método 1: Comitê híbrido semântico
2: para cada máscara  $M_j^{inst}$  faça
3:    $M_j^{inst} \leftarrow \text{Binarizar}(M_j^{inst})$  {Transforma valores maiores que zero em 1}
4: fim para
5: Aplicar a regra  $R$  às máscaras  $\{M_i^{sem}\} \cup \{M_j^{inst}\}$  conforme Algoritmo 1
6: retorne  $M_{\text{final}}$ 

7: Método 2: Comitê híbrido via instâncias
8:  $ID \leftarrow 1$  {Inicializa um identificador único para os núcleos}
9: para cada objeto  $O_j$  em  $\{M_i^{sem}\} \cup \{M_j^{inst}\}$  faça
10:   Calcular a caixa delimitadora  $B_{i,j}$ 
11: fim para
12: para cada par  $(O_{i,j}, O_{k,l})$  com  $i \neq k$  faça
13:   Calcular a sobreposição  $(B_{i,j}, B_{k,l})$ 
14:   se sobreposição  $> S$  então
15:     Considerar  $O_{i,j}$  e  $O_{k,l}$  como a mesma instância
16:   fim se
17: fim para
18:
19: para cada grupo de instâncias correspondentes faça
20:   Atribuir o valor 1 a todas as instâncias do grupo
21:   Aplicar a regra  $R$  às instâncias conforme Algoritmo 1
22:   Atribuir  $ID$  ao objeto segmentado {Cada núcleo recebe um valor distinto}
23:    $ID \leftarrow ID + 1$ 
24: fim para
25: retorne  $M_{\text{final}}$ 

```

Experimentos e Análise dos Resultados

Neste capítulo, são apresentados os resultados experimentais obtidos a partir da avaliação dos métodos propostos para a segmentação de núcleos em imagens histológicas da cavidade oral. A qualidade da segmentação foi avaliada de maneira quantitativa e qualitativa, comparando as predições dos modelos com as marcações manuais realizadas pelo especialista. Também são apresentados os resultados relacionados ao desempenho de métodos propostos em estudos da literatura. Nos experimentos, foram realizadas análises utilizando na base de imagens descrita a Seção 4.1. Por fim, os resultados obtidos neste estudo são destacados, com uma discussão sobre os fatores que contribuíram para o desempenho observado, bem como as limitações dos métodos analisados.

5.1 Análise da segmentação com os modelos baseados em aprendizagem profunda

A segmentação foi avaliada utilizando três configurações distintas de dados: (i) *baseline* (sem aumento de dados), (ii) aumento de dados pela técnica HistoGAN e (iii) aumento através do método RCAug. Cada uma dessas abordagens foi testada com os diferentes modelos de CNNs, buscando analisar os efeitos do aumento de dados sobre a capacidade de aprendizado dos modelos. Os resultados obtidos pelos modelos de segmentação semântica são apresentados na Tabela 3. Os melhores resultados obtidos com as métricas são destacados em negrito.

Tabela 3 – Avaliação dos modelos de CNN para segmentação semântica de núcleos em tecidos histológicos.

Rede	<i>Baseline</i>				HistoGAN				RCAug			
	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_C
PSP-ResNet50	84,72%	81,47%	86,64%	0,7717	85,32%	82,31%	87,90%	0,7829	86,42%	83,34%	89,66%	0,8029
PSP-MobileNet	80,83%	72,84%	84,76%	0,7161	83,11%	73,68%	85,88%	0,7602	84,36%	74,62%	86,90%	0,7845
FPN-ResNet50	91,95%	87,99%	92,76%	0,8772	92,34%	88,88%	94,31%	0,8857	93,13%	89,99%	95,29%	0,8879
FPN-MobileNet	89,32%	82,79%	92,55%	0,8384	91,08%	83,62%	94,31%	0,8513	92,07%	85,02%	95,77%	0,8736
U-Net-ResNet50	89,31%	78,50%	94,67%	0,8352	91,34%	79,54%	95,99%	0,8461	92,25%	80,96%	97,78%	0,8771
U-Net-MobileNet	85,48%	87,88%	82,97%	0,8029	86,22%	88,82%	83,88%	0,8181	87,44%	90,57%	84,99%	0,8259

Em relação ao *baseline*, os modelos baseados na FPN apresentaram os maiores valores de A_{CC} , seguidos pelos modelos baseados na U-Net. Os modelos baseados na PSP apresentaram os valores mais baixos. Foi observado que o *backbone* ResNet50 permitiu valores mais altos do que a MobileNet para todos os modelos, com um destaque para a FPN-ResNet50, o qual exibiu um aumento de 2,63% em comparação ao modelo FPN-MobileNet. Com o uso do aumento de dados pela HistoGAN, é possível notar que os valores de A_{CC} foram mais altos do que o *baseline*, principalmente para o modelo PSP-ResNet50, que mostrou um aprimoramento da métrica A_{CC} de 3,0%. Com o aumento de dados realizado pelo método RCAug, os valores foram melhores em comparação com o *baseline* e HistoGAN, representando um aumento de 1,18% para FPN-ResNet50 (*baseline* com $A_{CC} = 91,95\%$ e RCAug com $A_{CC} = 93,13\%$).

Para a sensibilidade (S_E), no *baseline*, o modelo FPN-ResNet50 apresentou o maior valor, seguido pela U-Net-MobileNet. Isso indica que esses modelos são mais robustos em identificar as regiões de núcleos celulares, apresentando menor taxa de FN em relação aos demais modelos. De maneira similar ao apresentado para a métrica A_{CC} , os modelos baseados na ResNet50 apresentaram valores altos de S_E . O aumento pela HistoGAN apresentou valores ligeiramente mais altos do que o *baseline*, apresentando incremento de aproximadamente 1% para todos os modelos. Ainda, o uso do modelo RCAug, permitiu melhores resultados, principalmente para os modelos baseados na MobileNet. Um destaque pode ser observado no modelo U-Net-MobileNet que apresentou aumento de 2,69% em relação ao *baseline*. Os modelos U-Net-ResNet50, FPN-ResNet50 e FPN-MobileNet mostraram valores relevantes para a métrica especificidade E_S com diferença superior a 5% em relação aos demais métodos. Isso permite afirmar que esses modelos foram mais eficazes na identificação de regiões de fundo ou objetos que não são de interesse, apresentando menor taxa de FP. De forma similar aos dados obtidos com a métrica S_E , o aumento pela HistoGAN contribuiu com melhorias para todos os modelos de rede, com incremento aproximado de 1,5%. Já o RCAug permitiu os melhores valores para a métrica, com incremento de cerca de 1,3% em relação a HistoGAN. O modelo com o maior valor de E_S foi a rede U-Net-ResNet50 para o *baseline* e ambos os aumentos de dados.

Ainda é importante destacar que o modelo FPN-ResNet50 apresentou os maiores valores para a métrica de D_C para todas as configurações dos conjuntos de imagens (*baseline* e com aumento pela HistoGAN e RCAug). No cenário *baseline*, esse modelo apresentou diferença de cerca de 4% em relação a FPN-MobileNet, que resultou no segundo maior valor. O aumento de dados permitiu a melhora nos valores, principalmente para os modelos baseados na PSPNet, que obtiveram os menores valores para o *baseline*. Com o uso do aumento de dados HistoGAN, a PSP-MobileNet apresentou incremento na aproximado de 4,5% para a métrica de D_C . Com o aumento de dados obtido pelo método RCAug, esse modelo de rede obteve incremento de 2,4% em relação a HistoGAN. O maior valor de D_C foi obtido pela FPN-ResNet50 com uso do aumento de dados RCAug, com valor próximo

a 89%.

O teste de Wilcoxon foi utilizado para avaliar a relevância estatística entre os métodos investigados em relação as métricas, assim como também para avaliar os métodos de aumento de dados. Os valores de A_{CC} e D_C de cada modelo foram investigados separadamente para cada métrica. Por exemplo, para a rede PSP-ResNet50 foi gerado um vetor com os valores de A_{CC} dos conjuntos obtidos por *baseline*, HistoGAN e RCAug, e outro vetor com os valores de D_C para os dados. Esse procedimento foi repetido para todos os modelos de redes. O modelo FPN-ResNet50 apresentou valores médios de ρ de 0,041 com os dados de A_{CC} e 0,026 para os dados de D_C , indicando que os resultados obtidos por essa rede apresentam relevância estatística significativa em relação aos outros modelos. Para avaliar os métodos de aumento de dados, foram gerados vetores contendo os valores de A_{CC} e D_C de todos os modelos semânticos para cada configuração de dados. Assim, tanto para o *baseline* quanto para a HistoGAN e o RCAug, foram gerados dois vetores, um contendo os valores de A_{CC} e outro contendo os valores de D_C obtidos pelos modelos baseados em FPN, U-Net e PSPNet. A comparação HistoGAN *versus baseline* apresentou valores ρ de 0,039 e 0,036, respectivamente, para os dados de A_{CC} e D_C , representando uma diferença estatisticamente relevante entre as duas configurações de dados. Para RCAug *versus baseline*, os valores foram de 0,029 e 0,027, enquanto na comparação RCAug *versus* HistoGAN, foram observados valores ρ de 0,034 e 0,031 para os dados de A_{CC} e D_C , mostrando que essa técnica de aumento de dados contribui para uma melhoria significativa nos resultados da segmentação de núcleos.

Em uma segunda etapa dos experimentos, foram avaliados os modelos de segmentação por instâncias. Os resultados são apresentados na Tabela 4 com as mesmas estratégias de representação dos dados e com os melhores resultados obtidos destacados em negrito.

Tabela 4 – Avaliação dos modelos de CNN para segmentação via instâncias de núcleos celulares.

Rede	<i>Baseline</i>				HistoGAN				RCAug			
	A_{CC}	S_E	E_S	D_{C2}	A_{CC}	S_E	E_S	D_{C2}	A_{CC}	S_E	E_S	D_{C2}
Mask R-CNN	91,33%	89,22%	92,84%	0,8841	92,04%	90,13%	93,41%	0,9023	92,33%	91,26%	94,15%	0,9065
StarDist	86,72%	84,36%	88,32%	0,8680	87,68%	86,27%	89,34%	0,8893	89,31%	88,13%	90,30%	0,8983
SOLOv2	89,41%	88,25%	90,67%	0,8679	90,25%	88,72%	90,06%	0,8944	91,22%	90,84%	92,38%	0,9048
HoVer-Net	88,97%	88,10%	89,77%	0,8757	91,46%	90,65%	92,33%	0,8972	92,08%	91,61%	92,74%	0,9028

O modelo Mask R-CNN apresentou, para os dados *baseline*, os maiores valores de A_{CC} , seguido pelo modelo SOLOv2. O modelo StarDist exibiu os menores valores para essa métrica. De forma similar ao apresentado na Tabela 3, quando aplicado as estratégias de aumento de dados, foi observado que o método RCAug permitiu um incremento de 1% em relação ao *baseline* (A_{CC} de 91,33% para 92,33%). Para a métrica S_E , os maiores valores foram alcançados pela Mask R-CNN em todas as configurações de dados. O aumento de dados obtido pela HistoGAN proporcionou incremento próximo a 1% para todos os modelos nessa métrica e, com o uso do método RCAug, esse melhoramento chegou a 2,28% para o modelo Mask R-CNN. Foi observado que, no *baseline*, dois modelos de segmentação

por instâncias apresentaram valores de E_S superiores a 90%. Com o aumento de dados HistoGAN, esse número foi elevado para três modelos e com o uso do RCAug, os valores com essa métrica, alcançaram valores superiores a 90% para todos os modelos. Avaliando os resultados com a métrica D_{C2} , a rede Mask R-CNN apresentou os maiores valores em todas as configurações de dados, alcançando 90,65% com o aumento de dados RCAug. Foi observado que a HistoGAN incrementou os valores de D_{C2} de todos os modelos em cerca de 2% em comparação ao *baseline* e que, para o método RCAug, esse incremento foi próximo a 3%.

Similar à análise realizada para os modelos de segmentação semântica, o teste de Wilcoxon foi empregado para avaliar os métodos via instância, utilizando a proposta de organização de dados. Ao avaliar os modelos das redes, a Mask R-CNN apresentou valores médios de ρ de 0,037 e 0,034 para as métricas de A_{CC} e D_{C2} , respectivamente, evidenciando que os resultados obtidos por esse modelo são estatisticamente relevantes em comparação às outras redes. Na avaliação do aumento de dados, foram observados valores ρ de 0,053 e 0,038, para a A_{CC} e o D_{C2} , respectivamente, na comparação HistoGAN *versus baseline*. Para RCAug *versus baseline*, os valores foram de 0,051 e 0,031, enquanto a comparação RCAug *versus* HistoGAN apresentou valores ρ de 0,063 e 0,053, respectivamente, para A_{CC} e D_{C2} . Esses valores ρ indicam que tanto a HistoGAN quanto o RCAug proporcionaram melhorias estatisticamente relevantes para a segmentação via instâncias. No entanto, a melhoria obtida pelo RCAug em relação a HistoGAN não apresentou relevância estatística.

Com os resultados é possível observar que os métodos FPN-ResNet50 e Mask R-CNN se destacaram nas análises. Então, uma análise do treinamento de um modelo de segmentação semântica e para segmentação via instâncias é apresentada na Figura 23. Essas curvas demonstram os impactos das técnicas de aumento de dados nos processos de treinamento e validação. Quando nenhum aumento de dados foi empregado, observou-se que a taxa de perda na validação aumentava significativamente ao longo do tempo, com a FPN-ResNet50 apresentando alta variabilidade e valores de perda atingindo médios de 0,7. Por outro lado, o aumento pela HistoGAN resultou em curvas de treinamento mais estáveis, reduzindo picos e minimizando a lacuna entre as taxas de perda para treinamento e validação. Nesse cenário, a Mask R-CNN apresentou valores de perda na validação próximos aos de treinamento, frequentemente sobrepostos ao longo do tempo. No entanto, o método RCAug demonstrou o melhor desempenho geral, com as curvas de aprendizado tanto para a FPN-ResNet50 quanto para a Mask R-CNN convergindo em torno de 100 épocas, com variações mínimas nas taxas de perda. A Mask R-CNN destacou-se por apresentar as menores variações e as curvas mais estáveis, refletindo em uma melhor convergência e maior eficiência no aprendizado. No geral, os resultados evidenciaram que a técnica RCAug proporcionou condições superiores para o treinamento dos modelos e a Mask R-CNN apresentou maior estabilidade entre as taxas de treinamento e validação.

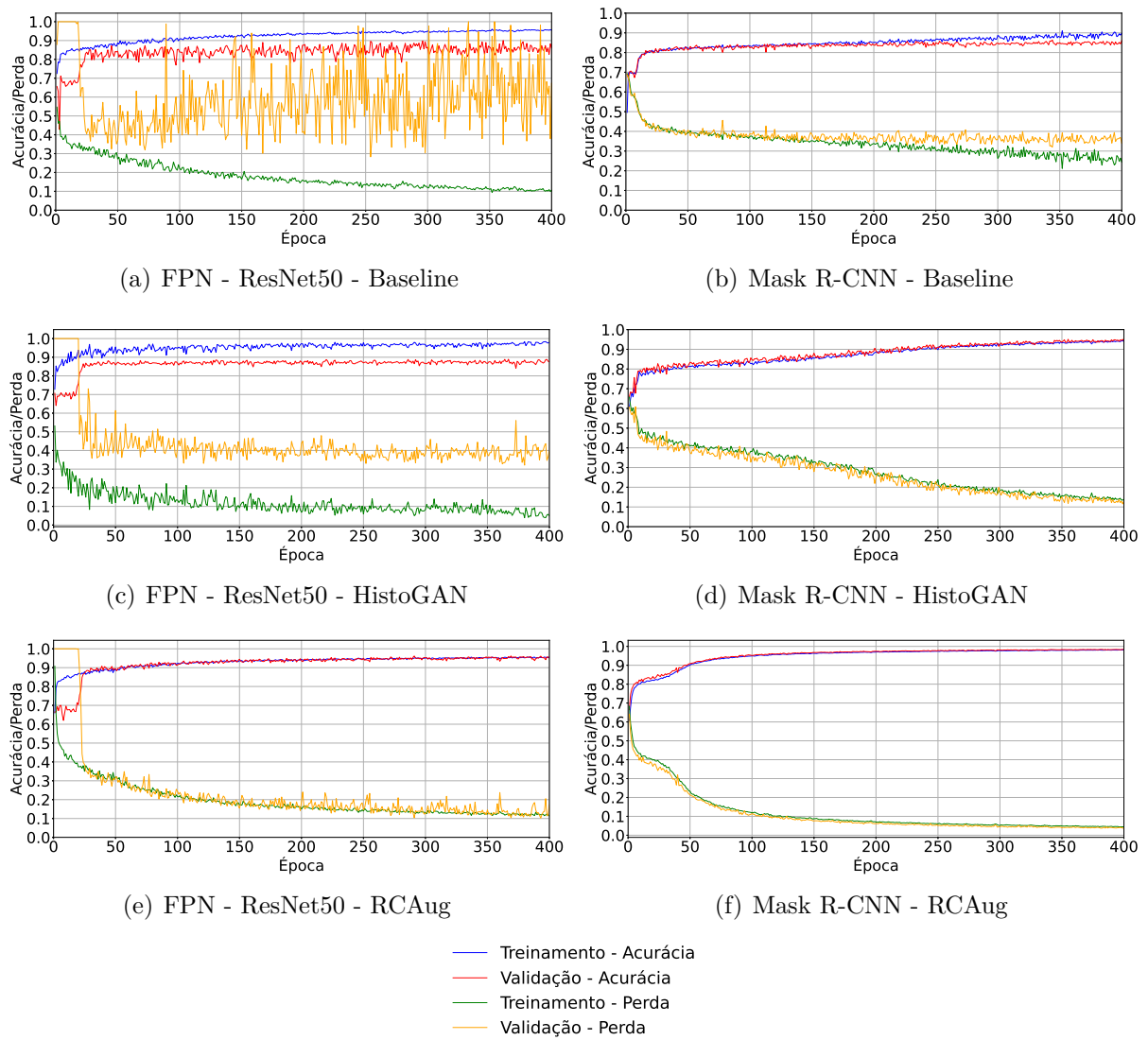


Figura 23 – Curvas de treinamento e validação para as métricas de acurácia e taxa de perda em diferentes configurações de dados: (a) e (b) *baseline*, (c) e (d) HistoGAN, e (e) e (f) RCAug. A coluna da esquerda apresenta os resultados para o modelo de segmentação FPN-ResNet50 e a coluna da direita exibe as curvas correspondentes ao modelo Mask R-CNN.

Pelos dados apresentados nas Tabelas 3 e 4, notou-se que aumento de dados permitiu uma melhora no desempenho dos modelos semânticos e por instâncias. As redes por instâncias, em especial, a Mask R-CNN e a HoVer-Net, apresentaram aumento nas métricas ACC e D_{C2} . O método de aumento RCAug proporcionou os melhores valores para as métricas em ambas as abordagens de segmentação, indicando uma melhor eficácia para a segmentação de estruturas nucleares.

Uma análise qualitativa foi realizada para os modelos FPN-ResNet50 e Mask R-CNN treinados na base aumentada pelo RCAug, por serem os que apresentaram os melhores resultados para segmentação semântica e via instâncias, respectivamente. Para a avaliação do modelo FPN-ResNet50, a Figura 24 apresenta a comparação da máscara gerada para

um tecido com DEO leve e a comparação com o padrão ouro. As regiões de FN e FP foram destacadas, respectivamente, com as cores vermelho e verde. Foi observado que, embora o modelo tenha segmentado os núcleos celulares, pequenas regiões de FP e FN foram geradas em suas bordas, com algumas áreas maiores comprometendo o contorno dos núcleos, como indicado pelas setas na cor ciano. Além disso, um núcleo marcado pelo especialista não foi segmentado, conforme indicado pela seta em amarelo. Destaca-se que, apesar dessas falhas, o modelo apresentou os melhores valores nas métricas de avaliação.

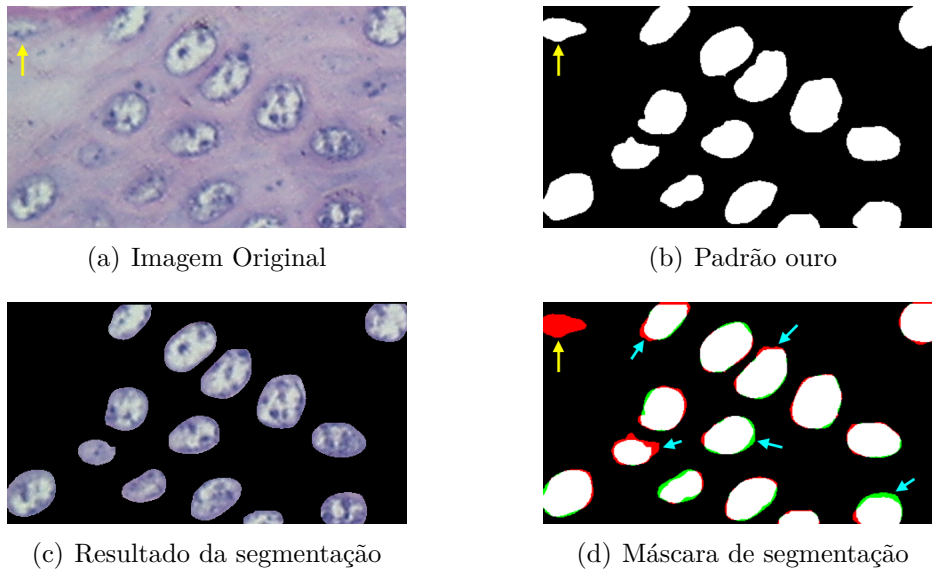


Figura 24 – Comparação visual entre o padrão ouro e o modelo FPN-ResNet50 para um tecido com DEO leve: (a) imagem original, (b) padrão ouro, (c) resultado da segmentação, (d) máscara gerada pelo modelo. Regiões na cor vermelha representam FN e regiões em verde indicam FP. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.

A análise visual do modelo Mask R-CNN é exibida na Figura 25, utilizando um tecido de DEO leve. O modelo foi segmentou os núcleos celulares marcados no padrão ouro, mas deixou de identificar um objeto, indicado pela seta amarela. Adicionalmente, todos os núcleos segmentados apresentaram falhas nos contornos, com as maiores discrepâncias em relação ao padrão ouro apontadas pelas setas em ciano. Embora essas limitações tenham sido identificadas, os resultados da Tabela 4 confirmam que o modelo atingiu o melhor desempenho na segmentação via instâncias.

Ainda com objetivo e avaliar a robustez dos modelos, o treinamento dos modelos CNNs foi avaliada utilizando o método de amostragem validação cruzada, com valor $k = 5$. Essa estratégia permitiu avaliar a consistência dos modelos ao expô-los a diferentes subdivisões dos dados de treinamento e teste. Nesta etapa, a validação cruzada foi empregada sobre os modelos FPN-ResNet50 e Mask R-CNN, selecionados devido aos resultados significativos que apresentaram para a segmentação semântica e via instâncias,

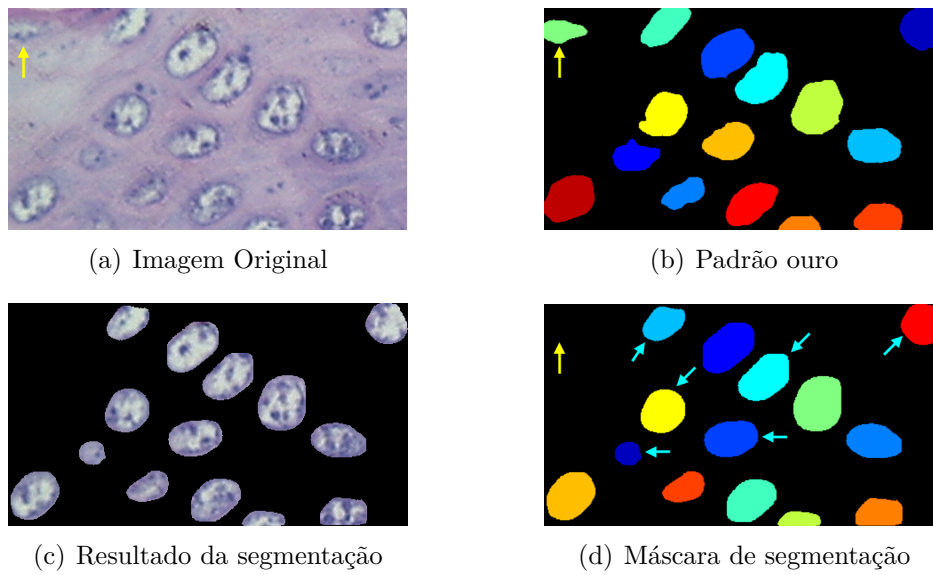


Figura 25 – Análise visual entre o modelo Mask R-CNN para um tecido com DEO leve: (a) imagem original, (b) padrão ouro, (c) resultado da segmentação, (d) máscara gerada pelo modelo. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.

respectivamente. Em cada iteração da validação cruzada, quatro subconjuntos foram utilizados para treinamento e um para validação. As técnicas de aumento de dados foram aplicadas exclusivamente nos dados dos quatro subconjuntos de treinamento em cada *fold*, enquanto os dados de validação permaneceram inalterados. Os valores obtidos para cada *fold* estão apresentados nas Tabelas 5 e 6. Os maiores valores obtidos para as métricas são destacados em negrito.

Para o modelo FPN-ResNet50, no *baseline*, foram observados valores consistentes ao longo dos *folds*, com uma variação próxima a 1% em todas as métricas avaliadas. Além disso, os valores médios foram próximos aos observados nos experimentos realizados com *holdout*, como mostrado na Tabela 3. Essa proximidade nos resultados reforça que os modelos são generalizáveis para diferentes conjuntos de dados. Em relação ao *baseline*, o uso da técnica HistoGAN proporcionou um incremento moderado nos resultados, apresentando aumento aproximado de 0,5%. Já o método RCAug apresentou uma melhoria mais significativa em relação ao *baseline*, com ganhos de cerca de 1,5%, sugerindo que esse método é eficaz para aumentar a diversidade dos dados de treinamento e aumentar a capacidade de generalização do modelo também com abordagem da validação cruzada.

O modelo Mask R-CNN apresentou comportamento semelhante, mas com resultados ligeiramente superiores em relação a FPN-ResNet50. Os valores apresentados foram próximos aos observados em relação aos resultados nos experimentos com *holdout* (ver Tabela 4). Nota-se que também houve pouca variação entre os *folds* para todas as métricas. Destaca-se que o método HistoGAN permitiu uma melhora em relação ao *baseline*, mas o RCAug ofereceu os melhores resultados médios em todas as métricas.

Com base nos resultados, foi possível confirmar a consistência dos modelos avaliados. A proximidade dos valores obtidos com a validação cruzada em relação aos testes *holdout* confirma a robustez dos modelos em ambos os métodos de amostragem. No entanto, deve-se considerar que para avaliação dos modelos usando a técnicas de amostragem *holdout*, o tempo médio necessário para os experimentos foi de 180 minutos. Quando isso foi realizado usando a validação cruzada, esse tempo foi de 900 minutos. Isso mostra que a diferença de 720 minutos torna inviável o uso das duas abordagens uma vez que a maior diferença entre as principais métricas para as técnicas de amostragem foi de cerca de 2% para a segmentação semântica e inferior a 1% para a segmentação via instâncias. Com isso, os experimentos realizados com os modelos de comitês ocorreram apenas com a amostragem dos dados usando *holdout*.

Tabela 5 – Avaliação dos resultados com o modelo FPN-ResNet50 e a técnica de validação cruzada.

<i>Fold</i>	<i>Baseline</i>				HistoGAN				RCAug			
	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_{C2}
1	89,37%	85,53%	90,07%	0,8515	90,06%	85,90%	90,43%	0,8629	91,06%	86,52%	91,61%	0,8630
2	90,30%	85,73%	90,26%	0,8517	90,26%	86,04%	90,64%	0,8630	91,07%	86,68%	91,95%	0,8631
3	90,31%	86,28%	90,88%	0,8520	90,60%	86,29%	91,42%	0,8631	91,20%	86,84%	92,13%	0,8632
4	90,86%	86,33%	91,05%	0,8521	90,74%	86,30%	91,70%	0,8632	91,07%	87,11%	92,28%	0,8633
5	91,05%	86,84%	91,12%	0,8522	91,27%	87,14%	92,19%	0,8635	92,30%	87,91%	92,71%	0,8635
Média	90,38%	86,14%	90,68%	0,8519	90,59%	86,41%	91,28%	0,8631	91,56%	87,49%	92,14%	0,8632

Tabela 6 – Avaliação do modelo Mask R-CNN com a técnica de validação cruzada.

<i>Fold</i>	<i>Baseline</i>				HistoGAN				RCAug			
	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_{C2}
1	90,87%	88,71%	92,34%	0,8797	91,33%	89,45%	92,66%	0,8955	91,74%	90,02%	93,52%	0,9002
2	91,03%	88,83%	92,41%	0,8801	91,50%	89,58%	92,72%	0,8962	91,81%	90,13%	93,56%	0,9013
3	91,24%	88,89%	92,55%	0,8802	91,69%	89,62%	92,85%	0,8975	91,99%	90,25%	93,67%	0,9025
4	91,31%	88,96%	92,67%	0,8823	91,74%	89,73%	92,92%	0,8987	92,03%	90,61%	93,79%	0,9031
5	91,43%	89,22%	92,72%	0,8834	91,82%	89,91%	93,03%	0,8993	92,12%	90,72%	93,81%	0,9037
Média	91,18%	88,92%	92,54%	0,8811	91,62%	89,66%	92,92%	0,8974	91,94%	90,83%	93,67%	0,9022

5.2 Estudo de desempenho entre modelos de comitês

A Tabela 7 apresenta os valores das métricas obtidos para os sistemas de votação de comitês de segmentação semântica com diferentes números de CNNs. Os melhores resultados obtidos com as métricas são destacados em negrito. Nos dados *baseline*, a regra da soma ponderada pelo D_C apresentou o melhor desempenho em termos das métricas A_{CC} e D_C . O comitê formado por quatro modelos— FPN-ResNet50, U-Net-ResNet50, FPN-MobileNet e U-Net-MobileNet — obteve os maiores valores, alcançando 93,82% e 0,8833, respectivamente, para as métricas de A_{CC} e D_C . Esse desempenho indica que a composição com quatro modelos foi a mais eficiente entre as exploradas.

Com o aumento de dados pela HistoGAN, os resultados apresentaram comportamento semelhante. Os resultados obtidos com a regra da soma ponderada com a métrica D_C , com

quatro modelos, proporcionou os melhores valores para as métricas A_{CC} e D_C , indicando que o aumento de dados HistoGAN permite melhora no desempenho em comparação à avaliação do *baseline*. O aumento de dados obtido pelo RCAug permitiu elevar esses valores, com a estratégia da soma ponderada pela D_C , com quatro modelos, os valores de A_{CC} e D_C foram de 95,07% e 0,9116, respectivamente. Estes resultados indicam que a seleção da regra de votação e o método para aumento de dados pode também impactar significativamente no desempenho dos comitês de segmentação. A soma ponderada pela D_C mostrou-se relevante, especialmente quando combinada com o aumento de dados RCAug. De forma semelhante à associação da RCAug, os modelos compostos por quatro CNNs se mostraram como os melhores entre as abordagens investigadas.

O teste de Wilcoxon foi aplicado para avaliar a relevância estatística entre os diferentes sistemas de comitês e sobre a quantidade de modelos que compuseram os comitês. Assim, de forma semelhante às análises anteriores, os valores de A_{CC} e D_C de cada sistema foram combinados em vetores individuais para todas as configurações de modelos. Assim, para a soma ponderada, foi gerado um vetor com os valores de A_{CC} com 2, 3, 4, 5 e 6 modelos e outro vetor com os valores de D_C nas mesmas configurações. Ao comparar os sistemas de quatro modelos com os compostos por três, o teste apresentou valores ρ igual a 0,056 e 0,053 para a A_{CC} e a D_C , respectivamente. Isso indica que o incremento nos valores apresentados pelos sistemas com quatro modelos não são significativos em relação aos comitês com três. Em relação aos comitês compostos 2, 5 e 6 modelos, foram observados valores ρ inferiores a 0,05. Isso sugere que a configuração com quatro modelos permite uma melhoria na segmentação de núcleos. Na avaliação das técnicas de aumento de dados com a métrica D_C , a estratégia RCAug apresentou valores ρ de 0,028 e 0,031 quando comparada com o *baseline* e o aumento de dados pela HistoGAN, respectivamente, confirmando a melhoria proporcionada por essa abordagem para a segmentação semântica.

A Figura 26 mostra os resultados dos modelos em um tecido com DEO leve e a comparação das regras de votação com o padrão ouro. Essa avaliação foi feita com a base aumentada pelo RCAug, devido ao seu desempenho em relação as métricas de avaliação. Nota-se que todas as regras apresentaram regiões (bordas) de FP, identificadas em cor verde, e FN, marcadas em cor vermelho. A regra da soma apresentou muitas regiões com degradação de núcleos, indicadas pelas setas na cor ciano. A soma ponderada apresentou menos regiões de degradação do que a soma, mas não identificou um núcleo marcado pelo especialista, conforme indicado pela seta em amarelo. Com a soma ponderada dada pelo D_C , observa-se que o modelo identificou esse núcleo, conforme indicado pela seta amarela, e apresentou poucas degradações na segmentação, como apontado pelas setas na cor ciano.

Os valores obtidos pelos comitês de segmentação via instâncias são apresentados na Tabela 8. Os maiores valores obtidos para as métricas são destacados em negrito. Os resultados indicam que, para os dados *baseline*, o desempenho das métricas melhoram com

Tabela 7 – Avaliação dos sistemas de votação de segmentadores semânticos.

<i>Baseline</i>					
Qtd. de modelos	Regra de votação	A_{CC}	S_E	E_S	D_C
2	Soma	91,95%	91,12%	92,43%	0,8772
	Soma ponderada	90,56%	89,36%	91,39%	0,8677
	Soma ponderada pela D_C	92,94%	92,41%	93,75%	0,8828
3	Soma	92,50%	91,61%	92,92%	0,8784
	Soma ponderada	91,24%	89,94%	92,01%	0,8730
	Soma ponderada pela D_C	93,58%	92,69%	94,32%	0,8817
4	Soma	92,49%	91,64%	93,08%	0,8738
	Soma ponderada	92,04%	90,86%	92,77%	0,8563
	Soma ponderada pela D_C	93,82%	93,09%	94,28%	0,8833
5	Soma	91,01%	90,67%	91,57%	0,8639
	Soma ponderada	90,99%	89,86%	91,15%	0,8401
	Soma ponderada pela D_C	92,68%	91,54%	92,59%	0,8639
6	Soma	89,95%	89,64%	89,76%	0,8484
	Soma ponderada	89,64%	88,09%	89,70%	0,8315
	Soma ponderada pela D_C	90,96%	90,45%	90,91%	0,8520
HistoGAN					
2	Soma	92,34%	91,64%	92,85%	0,8857
	Soma ponderada	90,99%	89,99%	91,68%	0,8719
	Soma ponderada pela D_C	93,28%	92,76%	93,73%	0,8946
3	Soma	92,93%	92,31%	93,53%	0,8906
	Soma ponderada	91,46%	90,46%	92,26%	0,8776
	Soma ponderada pela D_C	93,37%	93,20%	94,12%	0,8934
4	Soma	93,41%	92,23%	93,92%	0,8892
	Soma ponderada	92,97%	92,10%	93,46%	0,8827
	Soma ponderada pela D_C	93,66%	92,76%	94,89%	0,8984
5	Soma	92,29%	91,18%	92,57%	0,8767
	Soma ponderada	91,20%	90,45%	91,94%	0,8736
	Soma ponderada pela D_C	91,52%	91,71%	93,65%	0,8845
6	Soma	90,96%	89,54%	90,87%	0,8642
	Soma ponderada	89,56%	88,97%	90,34%	0,8578
	Soma ponderada pela D_C	90,25%	90,83%	92,06%	0,8724
RCAug					
2	Soma	93,13%	92,33%	93,87%	0,8879
	Soma ponderada	91,34%	90,97%	92,86%	0,8785
	Soma ponderada pela D_C	94,26%	93,33%	95,17%	0,8969
3	Soma	93,81%	92,87%	94,57%	0,8936
	Soma ponderada	91,81%	91,57%	93,53%	0,8848
	Soma ponderada pela D_C	94,66%	93,73%	95,57%	0,8998
4	Soma	94,72%	93,21%	95,21%	0,9116
	Soma ponderada	94,29%	92,79%	94,85%	0,9085
	Soma ponderada pela D_C	95,07%	93,61%	95,63%	0,9138
5	Soma	93,16%	92,60%	94,11%	0,9030
	Soma ponderada	92,56%	91,56%	93,31%	0,8974
	Soma ponderada pela D_C	93,46%	91,50%	94,61%	0,9002
6	Soma	91,50%	90,84%	92,29%	0,8854
	Soma ponderada	91,02%	90,18%	91,96%	0,8868
	Soma ponderada pela D_C	92,45%	89,71%	93,40%	0,8887

o aumento do número de modelos presente no comitê, especialmente na estratégia da regra da soma simples. Entretanto, a soma ponderada, dado pela D_{C2} , apresentou os maiores valores para as métricas de A_{CC} , E_S e D_{C2} . O aumento de dados HistoGAN permitiu elevar os valores das métricas de avaliação da segmentação. A soma ponderada, dado pela D_{C2} , alcançou os maiores valores para todas as métricas no comitê com quatro modelos,

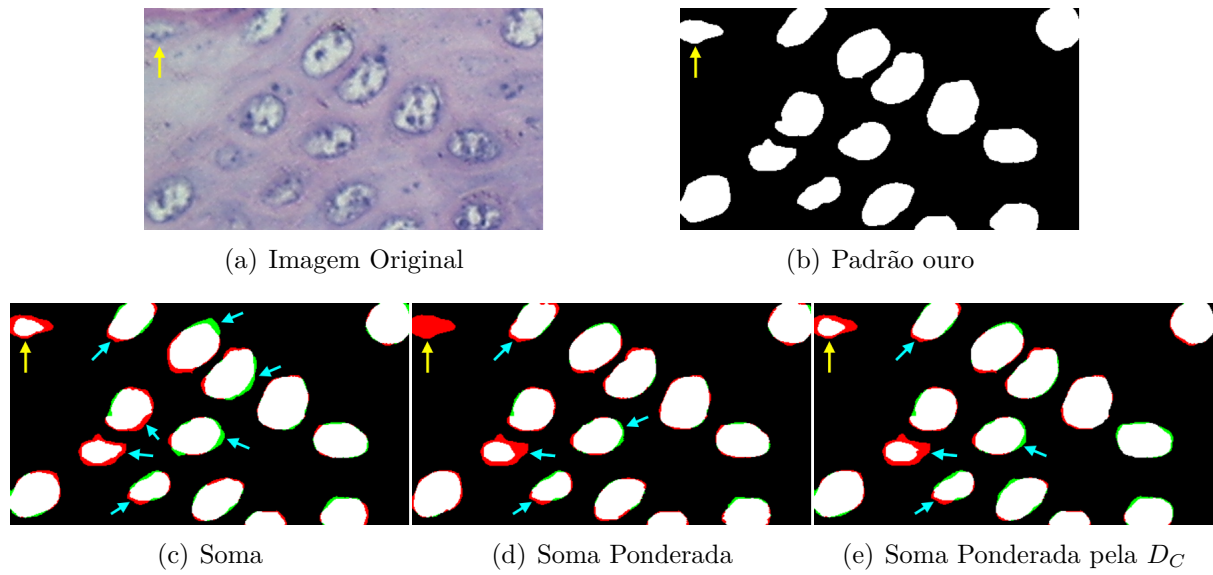


Figura 26 – Comparação visual entre o padrão ouro e as abordagens de segmentação semântica em um tecido com DEO leve: (a) imagem original, (b) padrão ouro marcado pelo especialista, (c-e) regras de votação. Regiões na cor vermelha representam FNs e regiões em verde indicam FPs. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.

apresentando incremento de 0,3% e 0,5% para as métricas de A_{CC} e D_{C2} , respectivamente. Com o aumento de dados RCAug, semelhante a tendência observada nas investigações, esses valores foram ainda melhores. Com o comitê com quatro modelos de CNNs, a soma ponderada, dada pela D_{C2} , apresentando A_{CC} e D_{C2} de 94,09% e 0,9461, indicando um incremento de 0,75% e 1,4%, respectivamente.

De forma semelhante aos comitês semânticos, o teste de Wilcoxon destacou que o sistema com quatro modelos apresentou resultados estatisticamente relevantes em comparação com os demais, com valores de ρ iguais a 0,027, 0,028 para as métricas A_{CC} e D_{C2} , respectivamente. O teste também destacou que o aumento de dados RCAug resultou em melhorias estatisticamente significativas, com valores ρ de 0,032 e 0,029 para a A_{CC} e a D_{C2} , respectivamente, em relação às demais configurações de dados. Os resultados apresentados indicam que a combinação, método de aumento de dados e as regras de votação, é um fator significativo para o desempenho dos comitês de segmentação. O aumento de dados RCAug permitiu os maiores valores para a maioria das métricas, especialmente para a soma ponderada dada pela D_{C2} . Ainda, os comitês compostos por quatro modelos via instância mostraram-se como os mais adequados dentre as abordagens estudadas.

A avaliação qualitativa dos sistemas de comitê via instâncias são apresentados na Figura 27, que mostra os resultados obtidos para um tecido com DEO leve. Assim como na Figura 26, os resultados exibem redes treinadas com a base de dados aumentada pelo RCAug. Como os valores são atribuídos às instâncias conforme a ordem de detecção do modelo, variações nos resultados, como FP e FN, podem alterar a sequência de identifi-

Tabela 8 – Avaliação dos sistemas de votação de segmentadores via instâncias.

<i>Baseline</i>					
Qtd. de modelos	Regra de votação	A_{CC}	S_E	E_S	D_{C2}
2	Soma	90,06%	88,92%	90,73%	0,8491
	Soma ponderada	85,93%	84,18%	86,43%	0,8849
	Soma ponderada pela D_{C2}	89,74%	88,92%	90,21%	0,8870
3	Soma	92,26%	91,76%	92,97%	0,9051
	Soma ponderada	88,55%	88,09%	86,23%	0,8972
	Soma ponderada pela D_{C2}	92,41%	91,92%	92,73%	0,9083
4	Soma	93,21%	92,64%	94,88%	0,9303
	Soma ponderada	91,09%	89,78%	92,11%	0,8936
	Soma ponderada pela D_{C2}	93,34%	92,52%	94,30%	0,9320
<i>HistoGAN</i>					
2	Soma	88,89%	88,08%	89,34%	0,8945
	Soma ponderada	89,17%	87,92%	91,21%	0,8918
	Soma ponderada pela D_{C2}	89,67%	88,51%	90,36%	0,8967
3	Soma	91,09%	89,23%	92,14%	0,9158
	Soma ponderada	89,76%	89,08%	90,32%	0,9113
	Soma ponderada pela D_{C2}	91,34%	88,73%	92,79%	0,9181
4	Soma	92,41%	91,44%	93,98%	0,9364
	Soma ponderada	93,64%	92,11%	94,02%	0,9356
	Soma ponderada pela D_{C2}	93,63%	93,16%	94,26%	0,9371
<i>RCAug</i>					
2	Soma	88,99%	88,13%	90,06%	0,8984
	Soma ponderada	89,55%	88,68%	89,72%	0,8936
	Soma ponderada pela D_{C2}	89,93%	89,17%	90,74%	0,8996
3	Soma	91,40%	89,33%	91,92%	0,9208
	Soma ponderada	91,73%	91,09%	92,40%	0,9174
	Soma ponderada pela D_{C2}	92,17%	90,48%	92,83%	0,9233
4	Soma	94,08%	93,73%	94,68%	0,9453
	Soma ponderada	93,91%	93,26%	94,15%	0,9452
	Soma ponderada pela D_{C2}	94,09%	93,64%	94,81%	0,9461

cação dos núcleos, resultando em diferenças nos valores de núcleos em diferentes modelos. Dessa forma, os comitês mantiveram essa característica, apresentando cores distintas para cada núcleo celular. A regra da soma identificou todos os núcleos marcados pelo padrão ouro, mas apresentou objetos com definição de contornos diferentes da marcação do especialista e região de FP, como indicado pelas setas em cor ciano. Já as regras da soma ponderada e soma ponderada pelo D_{C2} não identificaram um núcleo, como indicado pela seta na cor amarela. É destacado que, embora a regra da soma tenha permitido a segmentação de um núcleo que as outras regras não segmentaram, ela apresentou uma região de FP que não foi anotada no padrão ouro, conforme apontado pelas setas em cor amarelo na Figura 27 (c). Embora todas as regras tenham apresentado falhas nos contornos nucleares, eles apresentaram bom desempenho em relação as métricas de avaliação, como mostrado na Tabela 8. Os resultados obtidos pela soma ponderada pelo D_{C2} apresentou indicam que a atribuição de pesos baseada no desempenho das redes contribui para melhorias na segmentação.

Para os modelos de comitê que geraram os melhores resultados, uma análise foi realizada sobre cada nível de DEO e tecido saudável. Então, as redes FPN-ResNet50 (semântica) e Mask R-CNN (instâncias), bem como os comitês com as regras da soma ponderada

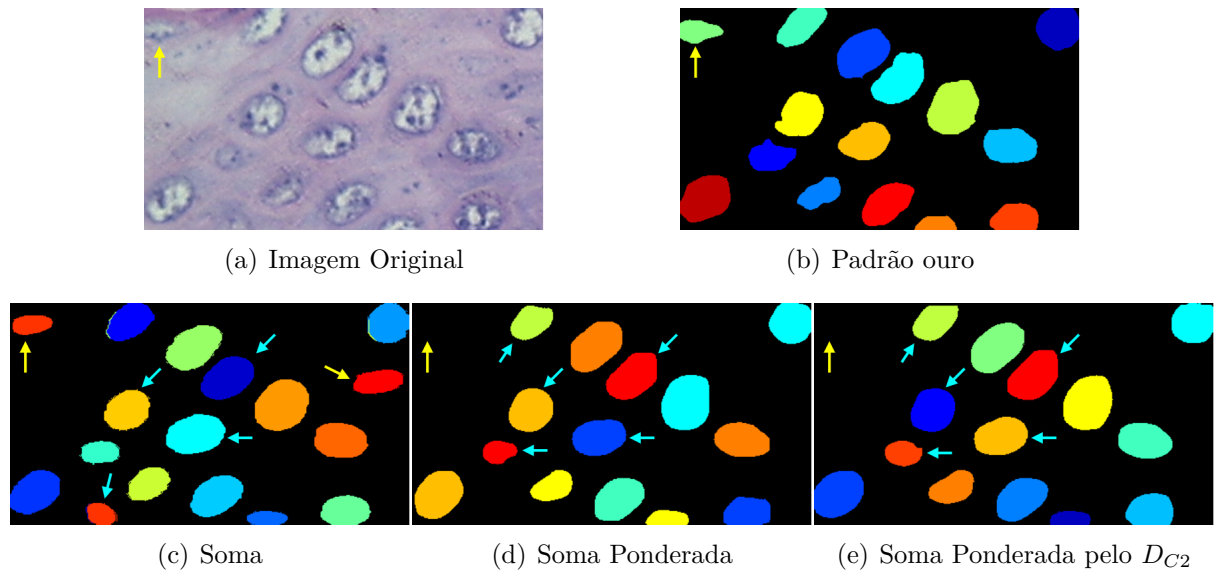


Figura 27 – Comparação visual entre o padrão ouro e as abordagens de comitês para segmentação de instâncias em um tecido com DEO leve: (a) imagem original, (b) padrão ouro marcado pelo especialista, (c-e) regras de votação. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.

DC (semântico) e D_{C2} (instância) e os dados do RCAug foram avaliados. A Tabela 9 traz os resultados obtidos com cada nível de OED e tecidos saudáveis para as imagens investigadas, com os melhores valores destacados em negrito.

Tabela 9 – Desempenho das estratégias de segmentação em diferentes classes de OED.

	FPN-ResNet50				Soma Ponderada - D_C				Mask R-CNN				Soma Ponderada - D_{C2}			
Tecido	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_{C2}	A_{CC}	S_E	E_S	D_{C2}
Saudável	94,31%	91,04%	96,47%	0,8965	95,98%	94,72%	96,92%	0,9223	93,51%	92,35%	95,41%	0,9185	95,36%	94,82%	96,05%	0,9557
Leve	93,81%	90,51%	95,72%	0,8923	95,52%	93,99%	96,07%	0,9176	93,03%	91,82%	94,76%	0,9127	94,61%	94,61%	95,41%	0,9510
Moderado	92,42%	89,07%	94,71%	0,8801	94,53%	92,94%	95,01%	0,9061	91,52%	90,47%	93,24%	0,8971	93,22%	92,63%	94,03%	0,9424
Severo	91,98%	89,33%	94,26%	0,8827	94,25%	92,79%	94,53%	0,9004	91,25%	90,40%	93,19%	0,8976	93,17%	92,51%	93,76%	0,9352
Média	93,13%	89,99%	95,29%	0,8879	95,07%	93,61%	95,63%	0,9138	92,33%	91,26%	94,15%	0,9065	94,09%	93,64%	94,81%	0,9461

Os resultados indicam uma variação nos valores das métricas de acordo com o nível de OED. A FPN-ResNet50 apresentou variação de 2,3% e 1,4 para as métricas de A_{CC} e D_C , indicando uma redução nos valores de acordo com o grau de intensidade da lesão. De forma similar, a soma ponderada pela D_C também apresentou variação nos valores para os diferentes níveis de OED, sendo 1,7% para a A_{CC} e 2,2% para a D_C . Em todos os graus, essa abordagem exibiu D_C superior a 90%. Para a segmentação via instâncias, a Mask R-CNN apresentou cerca de 2,3% de variação para a A_{CC} . Para todas as métricas investigadas, essa rede obteve valores superiores a 90%. Já a soma ponderada pela D_{C2} apresentou aumentos na segmentação via instâncias, com valores superiores a 0,92 para todas métricas, indicando um incremento geral de cerca de 2% em relação a Mask R-CNN. Para todas as abordagens apresentadas, o grau de DEO influenciou no resultado da segmentação, sendo que tecidos saudáveis e com DEO leve apresentaram melhor desempenho em comparação aos níveis moderado e severo.

Finalmente, uma terceira abordagem de comitê foi investigada, integrando os resultados dos modelos de segmentação semântica e via instâncias em sistemas híbridos. Essa estratégia foi empregada combinando os modelos utilizados nas somas ponderadas D_C e D_{C2} . Assim, foram criados dois modelos de comitês híbridos: um semântico, no qual as máscaras geradas pelos modelos via instâncias foram binarizadas para transformar todas as regiões segmentadas em categorias binárias; e outro via instâncias, no qual núcleos sobrepostos foram identificados e separados, transformando os resultados semânticos em instâncias e permitindo a aplicação das regras de votação. Para ambas as abordagens, foram empregadas as melhores composições apresentadas nas Tabelas 7 e 8, ou seja, os comitês compostos por quatro modelos. Dessa forma, os comitês híbridos foram compostos pelos modelos FPN-ResNet50, FPN-MobileNet, U-Net-ResNet50, U-Net-MobileNet, Mask R-CNN, StarDist, SOLOv2 e HoVer-Net treinados nas imagens aumentadas pelo RCAug, devido ao seu desempenho apresentado nos resultados anteriores. Os resultados desses comitês híbridos são apresentados na Tabela 10, sendo os melhores resultados destacados em negrito.

Tabela 10 – Avaliação dos sistemas de comitês híbridos.

Tecido	Comitê Semântico				Comitê via Instâncias			
	A_{CC}	S_E	E_S	D_C	A_{CC}	S_E	E_S	D_{C2}
Saudável	94,98%	89,54%	95,95%	0,9024	94,07%	91,46%	93,51%	0,8933
Leve	94,85%	89,31%	95,73%	0,9003	93,94%	91,27%	93,46%	0,8913
Moderada	94,73%	89,12%	95,51%	0,8958	93,80%	91,09%	93,42%	0,8873
Severa	94,60%	88,93%	95,29%	0,8946	93,67%	90,90%	93,38%	0,8861
Média	94,79%	89,23%	95,62%	0,8983	93,87%	91,18%	93,44%	0,8895

O comitê semântico demonstrou resultados ligeiramente superiores nas métricas de A_{CC} , E_S e D_C , indicando maior precisão geral e melhores valores na segmentação, enquanto o comitê via instâncias obteve valores superiores para S_E , evidenciando menos regiões de FN. Esses valores indicam que o comitê semântico mostrou-se mais eficaz na segmentação, enquanto o comitê via instâncias apresentou maior sensibilidade na identificação de núcleos individuais. Apesar de ter alcançado valores inferiores aos exibidos nas Tabelas 7 e 8, os comitês híbridos apresentaram consistência na segmentação, com variações médias inferiores a 1% entre os diferentes tecidos, reforçando sua robustez e potencial para garantir uma segmentação uniforme para os níveis de DEO.

Adicionalmente, os testes estatísticos foram realizados por meio Wilcoxon e confirmaram diferenças significativas na comparação entre os comitês, semântico e via instâncias, com valores p de 0,042 e 0,037 para A_{CC} e D_C , respectivamente. Esses resultados sugerem que o comitê semântico possui maior eficácia em termos de segmentação global. Entretanto, é importante destacar que esse comitê apresentou diferença média de cerca de 5% entre os valores de S_E e E_S , enquanto que, para o comitê via instâncias, essa diferença foi de aproximadamente 2%.

Os resultados visuais dessas abordagens são apresentados na Figura 28, avaliando os sistemas híbridos em comparação ao padrão ouro e aos comitês semântico e via instâncias. O comitê híbrido semântico identificou todos núcleos celulares, exibindo menos regiões de FN em comparação à soma ponderada pela D_C . No entanto, apresentou maior degradação nos contornos nucleares, com áreas de FP, próximas aos objetos, indicadas pelas setas em cor ciano. Foi observado que o sistema híbrido baseado em instâncias permitiu a segmentação de todos os núcleos, enquanto a soma ponderada pelo D_{C2} não segmentou um objeto, como indicado pela seta na cor amarelo. Entretanto, esse sistema também apresentou falhas nos contornos nucleares (ver setas em ciano). Embora essas abordagens exibam desempenho visual próximo aos demais sistemas, ambas apresentaram limitações na segmentação, reforçando a necessidade de investigações adicionais.

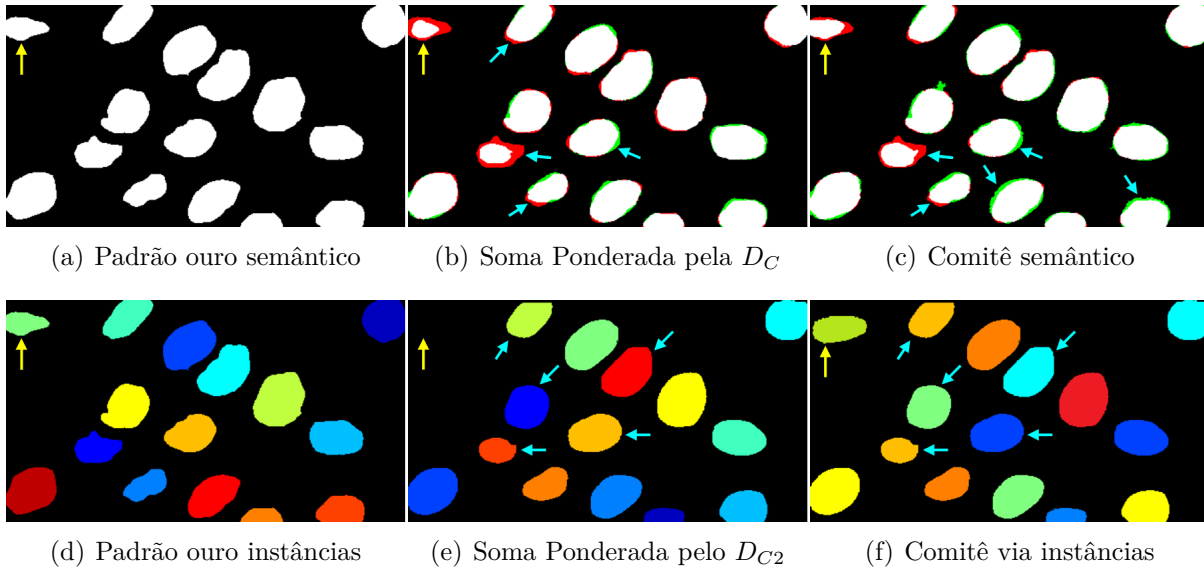


Figura 28 – Comparação visual entre o padrão ouro e as abordagens de comitês híbridos em um tecido com DEO leve: (a) padrão ouro semântico, (b) comitê semântico com soma ponderada pelo D_C , (c) Comitê híbrido semântico, (d) padrão ouro de instâncias, (e) comitê via instâncias com soma ponderada pelo D_C , (f) comitê híbrido via instâncias. Regiões na cor vermelha representam FN e regiões em verde indicam FP. As setas em ciano indicam degradação no contorno nuclear e a seta em amarelo mostra o objeto não identificado.

5.3 Comparação entre os modelos propostos e métodos da literatura

Para analisar os métodos propostos em relação a boa capacidade de generalização, as abordagens com os resultados mais promissores foram avaliada em relação às três bases de dados públicas disponíveis em: CryoNuSeg (MAHBOD et al., 2021), MoNuSeg 2018 (KUMAR et al., 2019) e GlaS (SIRINUKUNWATTANA et al., 2017). Para garantir uma avaliação mais completa da capacidade dos modelos em diferentes contextos, foram selecionados estudos da literatura que utilizaram ao menos duas dessas bases. Os resultados estão apresentados na Tabela 11, onde as somas ponderadas pelo D_C e o comitê híbrido semântico, ambos utilizando aumento de dados via RCAug, foram comparados aos métodos da literatura. Os valores apresentados correspondem à métrica D_C .

Na base CryoNuSeg, os comitês das somas ponderadas apresentaram desempenho próximo aos métodos recentes, sendo inferiores aos métodos de Kiran et al. (2022), demonstrando a relevância da abordagem para segmentação nuclear. O comitê híbrido apresentou resultado superior aos métodos propostos por Liao, Xu e Fan (2022) e por Sharma e Mishra (2023). Para a base MoNuSeg 2018, as somas ponderadas também apresentaram resultados promissores, superando a maioria dos trabalhos avaliados, enquanto mantiveram-se próximos aos trabalhos de Hassan et al. (2021) e de Kiran et al. (2022). Para essa base, o comitê híbrido apresentou resultados superiores a metade dos

estudos investigados. Por fim, na base GlaS, os resultados dos sistemas propostos foram promissores. O comitê via instâncias da soma ponderada obteve o melhor desempenho, alcançando valor superior ao trabalho de Liao, Xu e Fan (2022) e próximo ao método de proposto por Du et al. (2024).

Esses resultados mostram que a metodologia proposta é adaptável às características de diferentes bases de dados, mantendo um desempenho consistente entre conjuntos com tecidos variados. Embora nem sempre apresente os maiores valores em relação à literatura, os sistemas desenvolvidos demonstraram resultados sólidos, evidenciando sua robustez em diferentes cenários de segmentação nuclear.

Tabela 11 – Comparação entre os sistemas propostos e métodos de segmentação na literatura (D_C).

Modelo	CryoNuSeg	MoNuSeg 2018	GlaS
	D_C	D_C	D_C
(FIRAT et al., 2024)	0,88	0,88	-
(HASSAN et al., 2021)	0,86	0,89	-
(KIRAN et al., 2022)	0,92	0,90	-
(LIAO; ZHANG, 2023)	0,82	0,83	-
(SHARMA; MISHRA, 2023)	0,83	0,78	-
(DAS et al., 2023)	0,91	-	0,94
(DU et al., 2024)	-	0,82	0,90
(LIAO; XU; FAN, 2022)	-	0,69	0,85
Soma ponderada pelo D_C	0,87	0,88	0,87
Soma ponderada pelo D_{C2}	0,88	0,89	0,89
Híbrido semântico	0,85	0,86	0,86

Conclusões

Este trabalho apresentou a investigação de métodos para segmentação de núcleos celulares em tecidos histológicos de DEOs, empregando sistemas de comitês baseados em modelos de CNN. A proposta central envolveu o desenvolvimento, avaliação e comparação de abordagens de comitês baseados em modelos semânticos, via instâncias e híbridos. Ainda traz a investigação do uso de métodos para aumento de dados aplicados aos modelos em uma base de dados construída durante o estudo e disponibilizada publicamente.

A criação da base de dados OralEpitheliumDB representa uma importante contribuição deste trabalho. Ela inclui anotações semânticas e de instâncias de núcleos, em tecidos saudáveis e com os diferentes graus de DEO. A disponibilização pública dessa base possibilita à comunidade científica realizar mais estudos e experimentos sobre lesões de DEO, fomentando a replicação e evolução das técnicas de segmentação na área de histopatologia. Além disso, como essa base foi separada por cada grau de DEO, ela pode ser utilizada em estudos nas demais etapas de um sistema CAD, como a classificação.

O uso de técnicas de aumento de dados desempenhou um papel importante na metodologia desenvolvida, permitindo que a etapa de treinamento fosse implementada para todos os modelos de CNN sem introduzir *overfitting*. A técnica RCAug destacou-se como a mais eficaz, superando o desempenho do HistoGAN e do *baseline*. Isso indica que, para o escopo do trabalho, as operações morfológicas e de textura foram mais adequadas do que as imagens sintéticas geradas pelo modelo HistoGAN, reforçando a relevância do uso de estratégias para aumento de dados em cenários onde a escassez de imagens pode influenciar no resultado final do sistema.

Durante a etapa de treinamento, experimentos mostraram que a diferença média entre as métricas obtidas para validação cruzada e *holdout* foi inferior a 2% para a segmentação semântica e inferior a 1% para a segmentação via instâncias. No entanto, o tempo computacional necessário para a validação cruzada foi cerca de cinco vezes maior (180 minutos para o *holdout* e 900 minutos para a validação cruzada), tornando sua aplicação inviável para o presente estudo. Dessa forma, os experimentos dos modelos de comitês foram conduzidos exclusivamente com a amostragem *holdout*.

Os resultados dos experimentos confirmaram que os comitês de segmentadores aprimoram significativamente o desempenho em relação aos modelos individuais. O comitê baseado na regra da soma ponderada pelo D_C obteve os melhores resultados de precisão, com A_{CC} de 95,07% e D_C de 0,9138, superando os valores obtidos pelos modelos individuais. Esses resultados evidenciam a robustez dos comitês na mitigação de falhas na segmentação em tecidos com diferentes graus de DEO.

O comitê híbrido composto por modelos semânticos e via instâncias proposto como uma novidade foi capaz de apresentar um desempenho relevante, embora com valores ligeiramente inferiores aos demais comitês. No entanto, ele demonstrou menor variação nos resultados obtidos para os diferentes níveis de DEO, destacando seu potencial para a segmentação de estruturas nucleares. Essa abordagem, pouco explorada na literatura, reforça a importância de estudos na associação de paradigmas distintos de segmentação para obter resultados robustos. Apesar dos avanços em relação a desempenho dos métodos, os resultados mostraram que esses comitês apresentaram limitações em cenários com alta sobreposição de núcleos. Isso reforça a necessidade de investigações de estratégias para tratar essa limitação em abordagens semânticas e via instâncias.

Na avaliação utilizando as bases públicas, os sistemas de comitês propostos apresentaram valores de D_C similares em todas as bases, demonstrando a capacidade de generalização desses comitês. Além disso, os valores obtidos foram competitivos com métodos de segmentação da literatura que utilizaram essas mesmas bases. Por meio dessa análise comparativa, foi destacada a robustez dos sistemas propostos em diferentes bases, evidenciando seu potencial como uma ferramenta para auxiliar profissionais e pesquisadores da área da saúde.

6.1 Trabalhos Futuros

Uma limitação identificada neste estudo foi o desempenho dos comitês híbridos, que, embora tenham apresentado menor variação entre os níveis de DEO, não alcançaram os melhores resultados em termos das métricas de avaliação. Com base nesse tópico, uma pesquisa a ser explorada no futuro é a investigação de critérios mais robustos para a seleção de modelos a compor o comitê híbrido e a fusão de mapas de segmentação semântica com mapas de identificação de objetos, como caixas delimitadoras ou centroides. A introdução de um identificador de objetos poderia melhorar a precisão da separação de núcleos em casos de alta sobreposição, que foi um dos principais desafios enfrentados pelos comitês híbridos. Além disso, mecanismos de atenção serão para avaliar o impacto dessa abordagem nos comitês de segmentação.

Outra característica a ser investigada é o impacto do pré-processamento em combinação com as técnicas de aumento de dados. Métodos como normalização de cores, alinhamento de histogramas e remoção de artefatos podem ser empregados ao aumento

de dados para melhorar a qualidade das imagens geradas, reduzindo possíveis variações que possam influenciar negativamente o aprendizado dos modelos. Além disso, a integração de técnicas de interpretação, como mapas de ativação e análise de gradiente, pode ajudar a identificar limitações nos modelos utilizados, contribuindo para o critério da composição dos comitês.

6.2 Contribuições em Produção Bibliográfica

Durante o desenvolvimento deste trabalho de doutorado, as seguintes publicações foram produzidas pelo autor da tese:

1. A. B. Silva, D. F. Dos Santos, T. A. Tosta, A. S. Martins, L. A. Neves, B. A. Travençolo, P. R. de Faria, M. Z. do Nascimento. Segmentation of oral epithelial dysplasias employing Mask R-CNN and color normalization. **2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**, 2020.
2. A. B. Silva, A. S. Martins, T. A. A. Tosta, L. A. Neves, J. P. S. Servato, M. S. de Araújo, P. R. de Faria, M. Z. do Nascimento. Computational analysis of histological images from Hematoxylin and Eosin-stained oral epithelial dysplasia tissue sections. **Expert Systems with Applications**, 2022.
3. A. B. Silva, C. I. De Oliveira, D. C. Pereira, T. A. Tosta, A. S. Martins, A. M. Loyola, S. V. Cardoso, P. R. De Faria, L. A. Neves, M. Z. Do Nascimento. Assessment of the association of deep features with a polynomial algorithm for automated oral epithelial dysplasia grading. **2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**, 2022.
4. A. B. Silva, G. B. Rozendo, T. A. Tosta, A. S. Martins, A. M. Loyola, S. V. Cardoso, A. Lumini, L. A. Neves, P. R. de Faria, M. Z. do Nascimento. CNN ensembles for nuclei segmentation on histological images of OED. **2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)**, 2023.
5. A. B. Silva, A. S. Martins, T. A. A. Tosta, A. M. Loyola, S. V. Cardoso, L. A. Neves, P. R. de Faria, M. Z. do Nascimento. Oralepitheliumdb: A dataset for oral epithelial dysplasia image segmentation and classification. **Journal of Imaging Informatics in Medicine**, 2024.
6. A. B. Silva, T. A. Tosta, L. A. Neves, A. S. Martins, P. R. De Faria, M. Z. Do Nascimento. Ensemble of semantic segmentation models for oral epithelial dysplasia images. In: IEEE. **2024 37th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**, 2024.

Além dessas publicações, colaborações foram realizadas, em que experimentos auxiliaram na escrita de artigos científicos para publicação do grupo de pesquisa. Os artigos escritos em colaboração foram:

1. D. F. dos Santos, T. A. Tosta, A. B. Silva, P. R. de Faria, B. A. Travençolo, M. Z. do Nascimento. Automated nuclei segmentation on dysplastic oral tissues using CNN. **2020 International Conference on Systems, Signals and Image Processing (IWSSIP)**, 2020.
2. D. dos Santos, A. Silva, P. de Faria, B. Travençolo, M. do Nascimento. Impacts of color space transformations on dysplastic nuclei segmentation using CNN. **Anais do XVI Workshop de Visão Computacional**, 2020.
3. A. S. Martins, L. A. Neves, P. R. de Faria, T. A. Tosta, L. C. Longo, A. B. Silva, G. F. Roberto, M. Z. do Nascimento. A Hermite polynomial algorithm for detection of lesions in lymphoma images. **Pattern Analysis and Applications**, 2021.
4. A. D. Freitas, A. B. Silva, A. S. Martins, L. A. Neves, T. A. Tosta, P. R. de Faria, M. Z. do Nascimento. Evaluation of normalization technique on classification with deep learning features. **Anais do XVII Workshop de Visão Computacional**, 2021.
5. D. C. Pereira, L. C. Longo, T. A. Tosta, A. S. Martins, A. B. Silva, P. R. de Faria, L. A. Neves, M. Z. Do Nascimento. Classification of lymphomas images with polynomial strategy: An application with Ridge regularization. **2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**, 2022.
6. G. B. Rozendo, M. Z. do Nascimento, G. F. Roberto, P. R. de Faria, A. B. Silva, T. A. A. Tosta, L. A. Neves. Classification of non-Hodgkin lymphomas based on sample entropy signatures. **Expert Systems with Applications**, 2022.
7. J. J. Tenguam, L. H. D. C. Longo, A. B. Silva, P. R. De Faria, M. Z. Do Nascimento, L. A. Neves. Classification of H&E images exploring ensemble learning with two-stage feature selection. **2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)**, 2022.
8. G. B. Rozendo, M. Z. do Nascimento, G. F. Roberto, P. R. de Faria, A. B. Silva, T. A. A. Tosta, L. A. Neves. Sample Entropy Signatures: A new way to interpret SampEn values. **Software Impacts**, 2022.
9. L. A. Neves, J. M. C. Martinez, L. H. d. C. Longo, G. F. Roberto, T. A. A. Tosta, P. R. d. Faria, A. M. Loyola, S. V. Cardoso, A. B. Silva, M. Z. d. Nascimento, G. B.

- Rozendo. Classification of H&E images via CNN models with XAI approaches, DeepDream representations and multiple classifiers. **25th International Conference on Enterprise Information Systems**, 2023.
10. L. H. d. C. Longo, G. F. Roberto, T. A. Tosta, P. R. de Faria, A. M. Loyola, S. V. Cardoso, A. B. Silva, M. Z. do Nascimento, L. A. Neves. Classification of multiple H&E images via an ensemble computational scheme. **Entropy**, 2023.
 11. D. C. Pereira, L. C. Longo, T. A. Tosta, A. S. Martins, A. B. Silva, G. B. Rozendo, G. F. Roberto, A. Lumini, L. A. Neves, M. Z. do Nascimento. Handcrafted features vs deep-learned features: Hermite polynomial classification of liver images. **2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)**, 2023.
 12. T. R. Lopes, G. F. Roberto, C. Soares, T. A. Tosta, A. B. Silva, A. M. Loyola, S. V. Cardoso, P. R. de Faria, M. Z. do Nascimento, L. A. Neves. Association of grad-cam, lime and multidimensional fractal techniques for the classification of H&E images. **19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)**, 2024.
 13. T. R. Soares, A. B. Silva, A. M. Loyola, S. V. Cardoso, P. R. de Faria, L. A. Neves, M. Z. Nascimento, H. Razente. Divergências de Bregman aplicadas na recuperação por conteúdo de imagens de displasias orais. **Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)**, 2024.
 14. J. J. Tengum, L. H. d. C. Longo, G. F. Roberto, T. A. Tosta, P. R. de Faria, A. M. Loyola, S. V. Cardoso, A. B. Silva, M. Z. do Nascimento, L. A. Neves. Ensemble learning-based solutions: An approach for evaluating multiple features in the context of H&E histological images. **Applied Sciences**, 2024
 15. J. M. C. Martinez, L. A. Neves, L. H. d. C. Longo, G. B. Rozendo, G. F. Roberto, T. A. A. Tosta, P. R. de Faria, A. M. Loyola, S. V. Cardoso, A. B. Silva, M. Z. do Nascimento. Exploring DeepDream and XAI representations for classifying histological images. **SN Computer Science**, 2024.
 16. J. J. Tengum, L. H. da Costa Longo, G. F. Roberto, T. A. Tosta, A. B. Silva, M. Z. do Nascimento, L. A. Neves. Higuchi fractal dimension with a multidimensional approach for color images. **Software Impacts**, 2024
 17. V. B. Fernandes, A. B. Silva, D. C. Pereira, S. V. Cardoso, P. R. de Faria, A. M. Loyola, T. A. Tosta, L. A. Neves, M. Z. do Nascimento. Investigation of deep neural network compression based on tucker decomposition for the classification

- of lesions in cavity oral. **19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)**, 2024.
18. R. H. Carvalho, A. B. Silva, A. S. Martins, S. V. Cardoso, G. R. Freire, P. R. de Faria, A. M. Loyola, T. A. Tosta, L. A. Neves, M. Z. do Nascimento. Oral dysplasia classification by using fractal representation images and convolutional neural networks. **19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)**, 2024.

Referências

- ADEL, D. et al. Oral epithelial dysplasia computer aided diagnostic approach. In: IEEE. **2018 13th International Conference on Computer Engineering and Systems (ICCES)**. 2018. p. 313–318. Disponível em: <<https://doi.org/10.1109/ICCES.2018.8639452>>.
- AGGARWAL, C. C. et al. Neural networks and deep learning. **Springer**, Springer, v. 10, p. 978–3, 2018. Disponível em: <<https://doi.org/10.1007/978-3-319-94463-0>>.
- AGHERA, M. et al. Segmentation of nuclei in h&e-stained histological images using deep learning framework: A perspective on ensemble approach and nuclei count. In: IEEE. **2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC)**. 2023. p. 462–467. Disponível em: <<https://doi.org/10.1109/R10-HTC57504.2023.10461806>>.
- AHMED, M. R. et al. Doubleu-netplus: a novel attention and context-guided dual u-net with multi-scale residual feature fusion network for semantic segmentation of medical images. **Neural Computing and Applications**, Springer, v. 35, n. 19, p. 14379–14401, 2023. Disponível em: <<https://doi.org/10.1007/s00521-023-08493-1>>.
- 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/ICEngTechnol.2017.8308186>>.
- ALJUAID, A. et al. A multi-task learning network for automated detection of oral epithelial dysplasia. In: IEEE. **2024 IEEE International Conference on Information Reuse and Integration for Data Science (IRI)**. 2024. p. 25–30. Disponível em: <<https://doi.org/10.1109/IRI62200.2024.00018>>.
- ARLOVA, A. et al. Artificial intelligence-based tumor segmentation in mouse models of lung adenocarcinoma. **Journal of Pathology Informatics**, Elsevier, v. 13, p. 100007, 2022. Disponível em: <<https://doi.org/10.1016/j.jpi.2022.100007>>.
- ASLAM, A.; KHAN, E.; BEG, M. S. Improved edge detection algorithm for brain tumor segmentation. **Procedia Computer Science**, Elsevier, v. 58, p. 430–437, 2015. Disponível em: <<https://doi.org/10.1016/j.procs.2015.08.057>>.

- BALI, A.; SINGH, S. N. A review on the strategies and techniques of image segmentation. In: IEEE. **2015 Fifth International Conference on Advanced Computing & Communication Technologies**. 2015. p. 113–120. Disponível em: <<https://doi.org/10.1109/ACCT.2015.63>>.
- BASU, A. et al. A survey on recent trends in deep learning for nucleus segmentation from histopathology images. **Evolving Systems**, Springer, p. 1–46, 2023. Disponível em: <<https://doi.org/10.1007/s12530-023-09491-3>>.
- BEJNORDI, B. E. et al. A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **SPIE Medical Imaging**. 2015. p. 94200H–94200H. Disponível em: <<https://doi.org/10.1117/12.2081768>>.
- BELSARE, A.; MUSHRIF, M. Histopathological image analysis using image processing techniques: An overview. **Signal & Image Processing**, Academy & Industry Research Collaboration Center (AIRCC), v. 3, n. 4, p. 23, 2012. Disponível em: <<https://doi.org/10.5121/sipij.2012.3403>>.
- BOROVEC, J. Fully automatic segmentation of stained histological cuts. In: **International Student Conference on Electrical Engineering**. [S.l.: s.n.], 2013. v. 17, p. 1–7.
- BRENNAN, P. C. et al. Benefits of independent double reading in digital mammography: a theoretical evaluation of all possible pairing methodologies. **Academic radiology**, Elsevier, v. 26, n. 6, p. 717–723, 2019. Disponível em: <<https://doi.org/10.1016/j.acra.2018.06.017>>.
- BUDUMA, N.; LOCASCIO, N. **Fundamentals of deep learning: Designing next-generation machine intelligence algorithms**. [S.l.]: "O'Reilly Media, Inc.", 2017.
- CHAN, H.-P.; HADJIISKI, L. M.; SAMALA, R. K. Computer-aided diagnosis in the era of deep learning. **Medical physics**, Wiley Online Library, v. 47, n. 5, p. e218–e227, 2020. Disponível em: <<https://doi.org/10.1002/mp.13764>>.
- CHANG, V. et al. Gold-standard and improved framework for sperm head segmentation. **Computer methods and programs in biomedicine**, Elsevier, v. 117, n. 2, p. 225–237, 2014. Disponível em: <<https://doi.org/10.1016/j.cmpb.2014.06.018>>.
- DAS, R. et al. Dense dilated multi-scale supervised attention-guided network for histopathology image segmentation. **Computers in Biology and Medicine**, Elsevier, v. 163, p. 107182, 2023. Disponível em: <<https://doi.org/10.1016/j.combiomed.2023.107182>>.
- DENG, S. et al. Deep learning in digital pathology image analysis: A survey. **Frontiers of medicine**, Springer, p. 1–18, 2020.
- DERRAC, J. et al. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. **Swarm and Evolutionary Computation**, Elsevier, v. 1, n. 1, p. 3–18, 2011. Disponível em: <<https://doi.org/10.1016/j.swevo.2011.02.002>>.

- DOLZ, J. et al. Deep cnn ensembles and suggestive annotations for infant brain mri segmentation. **Computerized Medical Imaging and Graphics**, Elsevier, v. 79, p. 101660, 2020. Disponível em: <<https://doi.org/10.1016/j.compmedimag.2019.101660>>.
- DONG, X. et al. A survey on ensemble learning. **Frontiers of Computer Science**, Springer, v. 14, n. 2, p. 241–258, 2020. Disponível em: <<https://doi.org/10.1007/s11704-019-8208-z>>.
- DOST, F. et al. Malignant transformation of oral epithelial dysplasia: a real-world evaluation of histopathologic grading. **Oral surgery, oral medicine, oral pathology and oral radiology**, Elsevier, v. 117, n. 3, p. 343–352, 2014. Disponível em: <<https://doi.org/10.1016/j.oooo.2013.09.017>>.
- DU, G. et al. Dere-net: A dual-encoder residual enhanced u-net for muscle fiber segmentation of h&e images. **Biomedical Signal Processing and Control**, Elsevier, v. 98, p. 106765, 2024. Disponível em: <<https://doi.org/10.1016/j.bspc.2024.106765>>.
- EELBODE, T. et al. Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. **IEEE transactions on medical imaging**, IEEE, v. 39, n. 11, p. 3679–3690, 2020. Disponível em: <<https://doi.org/10.1109/TMI.2020.3002417>>.
- EHSAEYAN, E. An efficient image segmentation method based on expectation maximization and salp swarm algorithm. **Multimedia Tools and Applications**, Springer, v. 82, n. 26, p. 40625–40655, 2023. Disponível em: <<https://doi.org/10.1007/s11042-023-15149-8>>.
- FARIA, P. R. de et al. Absence of galectin-3 does not affect the development of experimental tongue carcinomas in mice. **Experimental and molecular pathology**, Elsevier, v. 90, n. 2, p. 189–193, 2011. Disponível em: <<https://doi.org/10.1016/j.yexmp.2010.12.007>>.
- FIRAT, H. et al. Convnext mixer-based encoder decoder method for nuclei segmentation in histopathology images. **International Journal of Imaging Systems and Technology**, Wiley Online Library, v. 34, n. 5, p. e23181, 2024. Disponível em: <<https://doi.org/10.1002/ima.23181>>.
- FONSECA-SILVA, T. et al. Association between histopathological features of dysplasia in oral leukoplakia and loss of heterozygosity. **Histopathology**, Wiley Online Library, v. 68, n. 3, p. 456–460, 2016.
- GANAI, 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.
- GHOSH, S. et al. Colorectal histology tumor detection using ensemble deep neural network. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 100, p. 104202, 2021.
- 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**, Elsevier, v. 58, p. 101563, 2019. Disponível em: <<https://doi.org/10.1016/j.media.2019.101563>>.
- GU, Y. et al. Glomerulus semantic segmentation using ensemble of deep learning models. **Arabian Journal for Science and Engineering**, Springer, v. 47, n. 11, p. 14013–14024, 2022.
- HASSAN, L. et al. Efficient stain-aware nuclei segmentation deep learning framework for multi-center histopathological images. **Electronics**, MDPI, v. 10, n. 8, p. 954, 2021.
- 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**. [S.l.: s.n.], 2016. p. 770–778.
- HESAMIAN, M. H. et al. Deep learning techniques for medical image segmentation: achievements and challenges. **Journal of digital imaging**, Springer, v. 32, n. 4, p. 582–596, 2019.
- HOWARD, A. et al. Searching for mobilenetv3. In: **Proceedings of the IEEE/CVF international conference on computer vision**. [S.l.: s.n.], 2019. p. 1314–1324.
- HUMEAU-HEURTIER, A. Texture feature extraction methods: A survey. **IEEE Access**, IEEE, v. 7, p. 8975–9000, 2019. Disponível em: <<https://doi.org/10.1109/ACCESS.2018.2890743>>.
- 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.
- JADER, G. et al. Deep instance segmentation of teeth in panoramic x-ray images. In: **IEEE. 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.], 2018. p. 400–407.
- JAYASOORIYA, P. R. et al. Malignant transformation of oral leukoplakia: a follow-up study. **Clinical Oral Investigations**, Springer, v. 24, n. 12, p. 4563–4569, 2020.
- JI, J. et al. Automated cervical cell segmentation using deep ensemble learning. **BMC Medical Imaging**, Springer, v. 23, n. 1, p. 137, 2023.
- JI, M.; WU, Z. Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic. **Computers and Electronics in Agriculture**, Elsevier, v. 193, p. 106718, 2022. Disponível em: <<https://doi.org/10.1016/j.compag.2022.106718>>.
- 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.

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>>.

KASSANI, S. H. et al. Classification of histopathological biopsy images using ensemble of deep learning networks. **arXiv preprint arXiv:1909.11870**, 2019.

_____. Deep transfer learning based model for colorectal cancer histopathology segmentation: A comparative study of deep pre-trained models. **International Journal of Medical Informatics**, Elsevier, v. 159, p. 104669, 2022.

KATO, S.; HOTTA, K. Automatic preprocessing and ensemble learning for low quality cell image segmentation. **arXiv preprint arXiv:2108.13118**, 2021.

KAUSHAL, C. et al. Recent trends in computer assisted diagnosis (cad) system for breast cancer diagnosis using histopathological images. **IRBM**, Elsevier, v. 40, n. 4, p. 211–227, 2019.

KIRAN, I. et al. Denseres-unet: Segmentation of overlapped/clustered nuclei from multi organ histopathology images. **Computers in Biology and Medicine**, Elsevier, v. 143, p. 105267, 2022. Disponível em: <<https://doi.org/10.1016/j.compbimed.2022.105267>>.

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.

KISANTAL, M. et al. Augmentation for small object detection. **arXiv preprint arXiv:1902.07296**, 2019.

KRAWCZYK, B. et al. Ensemble learning for data stream analysis: A survey. **Information Fusion**, Elsevier, v. 37, p. 132–156, 2017.

KRISHNAN, M. M. R. et al. Texture based segmentation of epithelial layer from oral histological images. **Micron**, Elsevier, v. 42, n. 6, p. 632–641, 2011.

_____. Textural characterization of histopathological images for oral sub-mucous fibrosis detection. **Tissue and Cell**, Elsevier, v. 43, n. 5, p. 318–330, 2011. Disponível em: <<https://doi.org/10.1016/j.tice.2011.06.005>>.

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.

LEE, L. K.; LIEW, S. C.; THONG, W. J. A review of image segmentation methodologies in medical image. **Advanced computer and communication engineering technology**, Springer, p. 1069–1080, 2015.

- LEE, S. et al. Class probability-guided ensemble learning-based semantic segmentation to predict cancerous regions on hematoxylin and eosin-stained images. In: IEEE. **2024 IEEE 18th International Conference on Semantic Computing (ICSC)**. [S.l.], 2024. p. 49–56.
- LI, X.; CHEN, D. A survey on deep learning-based panoptic segmentation. **Digital Signal Processing**, Elsevier, p. 103283, 2021.
- LIAO, L.; ZHANG, Z. Nuclear segmentation based on recurrent iteration and fusion attention mechanism. In: IEEE. **2023 3rd International Conference on Electronic Information Engineering and Computer Science (EIECS)**. 2023. p. 365–371. Disponível em: <<https://doi.org/10.1109/EIECS59936.2023.10435439>>.
- LIAO, Z.; XU, K.; FAN, N. Swin transformer assisted prior attention network for medical image segmentation. In: **Proceedings of the 8th International Conference on Computing and Artificial Intelligence**. [S.l.: s.n.], 2022. p. 491–497.
- LIN, T.-Y. et al. Feature pyramid networks for object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 2117–2125.
- LIU, D. et al. Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [S.l.: s.n.], 2020. p. 4243–4252.
- LIU, Z. et al. Robustifying deep networks for medical image segmentation. **Journal of digital imaging**, Springer, v. 34, p. 1279–1293, 2021.
- LONG, K. et al. Probability-based mask r-cnn for pulmonary embolism detection. **Neurocomputing**, Elsevier, v. 422, p. 345–353, 2021.
- LOPES, T. R. et al. Association of grad-cam, lime and multidimensional fractal techniques for the classification of h&e images. **Proceedings Copyright**, v. 441, p. 447, 2024. Disponível em: <<https://doi.org/10.5220/0012358200003660>>.
- LUMERMAN, H.; FREEDMAN, P.; KERPEL, S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. **Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology**, Elsevier, v. 79, n. 3, p. 321–329, 1995.
- LYNCH, C. J.; LISTON, C. New machine-learning technologies for computer-aided diagnosis. **Nature medicine**, Nature Publishing Group, v. 24, n. 9, p. 1304–1305, 2018.
- MA, J. et al. Segment anything in medical images. **Nature Communications**, Nature Publishing Group UK London, v. 15, n. 1, p. 654, 2024.
- MAHBOD, A. et al. 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. Disponível em: <<https://doi.org/10.1016/j.compbiomed.2021.104349>>.

- MARTINO, F. et al. Deep learning-based pixel-wise lesion segmentation on oral squamous cell carcinoma images. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 10, n. 22, p. 8285, 2020.
- MENDONÇA, D. F. et al. The inactive form of glycogen synthase kinase-3 β is associated with the development of carcinomas in galectin-3 wild-type mice, but not in galectin-3-deficient mice. **International journal of clinical and experimental pathology**, e-Century Publishing Corporation, v. 5, n. 6, p. 547, 2012.
- MINAEE, S. et al. Image segmentation using deep learning: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, 2021.
- MOHAMMED, A.; KORA, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. **Journal of King Saud University-Computer and Information Sciences**, Elsevier, v. 35, n. 2, p. 757–774, 2023.
- 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>>.
- NANNI, L.; LUMINI, A.; FANTOZZI, C. Exploring the potential of ensembles of deep learning networks for image segmentation. **Information**, MDPI, v. 14, n. 12, p. 657, 2023. Disponível em: <<https://doi.org/10.3390/info14120657>>.
- NANNI, L.; LUMINI, A.; GHIDONI, S. Ensemble of deep learned features for melanoma classification. **arXiv preprint arXiv:1807.08008**, 2018.
- NANNI, L. et al. An empirical study on ensemble of segmentation approaches. **Signals**, MDPI, v. 3, n. 2, p. 341–358, 2022.
- OBERMEYER, Z.; EMANUEL, E. J. Predicting the future—big data, machine learning, and clinical medicine. **The New England journal of medicine**, NIH Public Access, v. 375, n. 13, p. 1216, 2016.
- PALLUA, J. et al. The future of pathology is digital. **Pathology-Research and Practice**, Elsevier, v. 216, n. 9, p. 153040, 2020.
- PEREIRA, J. dos S. et al. Epidemiology and correlation of the clinicopathological features in oral epithelial dysplasia: analysis of 173 cases. **Annals of diagnostic pathology**, Elsevier, v. 15, n. 2, p. 98–102, 2011.
- POLIKAR, R. Ensemble learning. In: **Ensemble machine learning**. Springer, 2012. p. 1–34. Disponível em: <https://doi.org/10.1007/978-1-4419-9326-7_1>.
- RASOOL, A.; FRAZ, M. M.; JAVED, S. Multiscale unified network for simultaneous segmentation of nerves and micro-vessels in histology images. In: IEEE. **2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)**. [S.l.], 2021. p. 1–6.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. [S.l.], 2015. p. 234–241.

- 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>>.
- SALEH, H. et al. Heterogeneous ensemble deep learning model for enhanced arabic sentiment analysis. **Sensors**, MDPI, v. 22, n. 10, p. 3707, 2022.
- SANT'ANA, J. M. D. A. et al. Activation of the wnt/beta-catenin signaling pathway during oral carcinogenesis process is not influenced by the absence of galectin-3 in mice. **Anticancer research**, International Institute of Anticancer Research, v. 31, n. 9, p. 2805–2811, 2011.
- 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 Digital Imaging**, Springer, p. 1–16, 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>>.
- SHARMA, A.; MISHRA, P. K. Dri-unet: dense residual-inception unet for nuclei identification in microscopy cell images. **Neural Computing and Applications**, Springer, v. 35, n. 26, p. 19187–19220, 2023.
- SHEARSTON, K. et al. Malignant transformation rate of oral leukoplakia in an australian population. **Journal of Oral Pathology & Medicine**, Wiley Online Library, v. 48, n. 7, p. 530–537, 2019.
- SHEPHARD, A. J. et al. A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. **npj Precision Oncology**, Nature Publishing Group UK London, v. 8, n. 1, p. 137, 2024. Disponível em: <<https://doi.org/10.1038/s41698-024-00624-8>>.
- _____. Simultaneous nuclear instance and layer segmentation in oral epithelial dysplasia. In: **Proceedings of the IEEE/CVF International Conference on Computer Vision**. [S.l.: s.n.], 2021. p. 552–561.
- SHIRAZI, A. Z. et al. The application of deep convolutional neural networks to brain cancer images: a survey. **Journal of Personalized Medicine**, Multidisciplinary Digital Publishing Institute, v. 10, n. 4, p. 224, 2020.
- SILVA, A. B. **Métodos Computacionais para Análise e Classificação de Displasias em Imagens da Cavidade Bucal**. Tese (dissertation) — Universidade Federal de Uberlândia, 2019.
- SILVA, A. B. et al. 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>>.

_____. Segmentation of oral epithelial dysplasias employing mask r-cnn and color normalization. In: IEEE. **2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**. 2020. p. 2818–2824. Disponível em: <<https://doi.org/10.1109/BIBM49941.2020.9313101>>.

SIRINUKUNWATTANA, K. et al. Gland segmentation in colon histology images: The glas challenge contest. **Medical image analysis**, Elsevier, v. 35, p. 489–502, 2017.

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.

SOLOMON, C.; BRECKON, T. **Fundamentos de processamento digital de imagens: uma abordagem prática com exemplos em Matlab**. [S.l.]: Grupo Gen-LTC, 2000.

SOUZA, M. V. R. de et al. Expression of apc protein during tongue malignant transformation in galectin-3-deficient mice challenged by the carcinogen 4-nitroquinoline-n-oxide. **International journal of clinical and experimental pathology**, e-Century Publishing Corporation, v. 7, n. 6, p. 3255, 2014.

SPEIGHT, P. M. et al. Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. **Oral surgery, oral medicine, oral pathology and oral radiology**, Elsevier, v. 120, n. 4, p. 474–482, 2015.

SULTANA, F.; SUFIAN, A.; DUTTA, P. Evolution of image segmentation using deep convolutional neural network: a survey. **Knowledge-Based Systems**, Elsevier, v. 201, p. 106062, 2020. Disponível em: <<https://doi.org/10.1016/j.knosys.2020.106062>>.

SUNDARA, S. M.; AARTHI, R. Segmentation and evaluation of white blood cells using segmentation algorithms. In: IEEE. **2019 3rd international conference on trends in electronics and informatics (ICOEI)**. [S.l.], 2019. p. 1143–1146.

TILAKARATNE, W. M. et al. Oral epithelial dysplasia: Causes, quantification, prognosis, and management challenges. **Periodontology 2000**, Wiley Online Library, v. 80, n. 1, p. 126–147, 2019.

VUOLA, A. O.; AKRAM, S. U.; KANNALA, J. Mask-rcnn and u-net ensembled for nuclei segmentation. In: IEEE. **2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)**. [S.l.], 2019. p. 208–212.

WANG, E. K. et al. A deep learning based medical image segmentation technique in internet-of-medical-things domain. **Future Generation Computer Systems**, Elsevier, v. 108, p. 135–144, 2020.

WANG, X. et al. Volumetric attention for 3d medical image segmentation and detection. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. [S.l.], 2019. p. 175–184.

_____. Solov2: Dynamic and fast instance segmentation. **Advances in Neural information processing systems**, v. 33, p. 17721–17732, 2020.

WANG, Z.; WANG, E.; ZHU, Y. Image segmentation evaluation: a survey of methods. **Artificial Intelligence Review**, Springer, v. 53, n. 8, p. 5637–5674, 2020.

- WARIN, K.; SUEBNUKARN, S. Deep learning in oral cancer-a systematic review. **BMC Oral Health**, Springer, v. 24, n. 1, p. 212, 2024. Disponível em: <<https://doi.org/10.1186/s12903-024-03993-5>>.
- 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.
- WEI, L.; GAN, Q.; JI, T. Cervical cancer histology image identification method based on texture and lesion area features. **Computer Assisted Surgery**, Taylor & Francis, v. 22, n. sup1, p. 186–199, 2017.
- WEIGERT, M.; SCHMIDT, U. Nuclei instance segmentation and classification in histopathology images with stardist. In: IEEE. **2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)**. [S.l.], 2022. p. 1–4.
- WILCOXON, F. Individual comparisons by ranking methods. In: **Breakthroughs in statistics: Methodology and distribution**. Springer, 1992. p. 196–202. Disponível em: <https://doi.org/10.1007/978-1-4612-4380-9_16>.
- XIE, X. et al. Robust segmentation of nucleus in histopathology images via mask r-cnn. In: SPRINGER. **International MICCAI Brainlesion Workshop**. [S.l.], 2018. p. 428–436.
- XUE, Y. et al. Selective synthetic augmentation with histogan for improved histopathology image classification. **Medical image analysis**, Elsevier, v. 67, p. 101816, 2021.
- YANASE, J.; TRIANTAPHYLLOU, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. **Expert Systems with Applications**, Elsevier, v. 138, p. 112821, 2019.
- YANG, D. et al. Automatic liver segmentation using an adversarial image-to-image network. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. [S.l.], 2017. p. 507–515.
- YANG, Y.; LV, H.; CHEN, N. **A Survey on Ensemble Learning under the Era of Deep Learning**. 2021.
- YILDIZ, S.; MEMİŞ, A.; VARLI, S. An ensemble learning based nuclei segmentation: Using base deep learning models with different loss functions. In: IEEE. **2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS)**. 2024. p. 1–6. Disponível em: <<https://doi.org/10.1109/ISAS64331.2024.10845426>>.
- YING, X. An overview of overfitting and its solutions. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2019. v. 1168, p. 022022.
- ZELST, J. C. van et al. Validation of radiologists' findings by computer-aided detection (cad) software in breast cancer detection with automated 3d breast ultrasound: a concept study in implementation of artificial intelligence software. **Acta Radiologica**, SAGE Publications Sage UK: London, England, v. 61, n. 3, p. 312–320, 2020.

ZHAO, H. et al. Pyramid scene parsing network. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2017. p. 2881–2890. Disponível em: <<https://doi.org/10.1109/CVPR.2017.660>>.