



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Classificação e Segmentação de Imagens de Lesões
Podocitárias: Contribuições para Cenários
Multicorantes e Desbalanceados Utilizando Modelos
de Aprendizado Profundo**

George Oliveira Barros

Tese apresentada como requisito parcial para
conclusão do Doutorado em Informática

Orientador

Prof. Dr. Flávio de Barros Vidal

Coorientador

Prof. Dr. Angelo Amâncio Duarte

Brasília
2025

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

OB277c Oliveira Barros, George
Classificação e Segmentação de Imagens de Lesões
Podocitárias: Contribuições para Cenários Multicorantes e
Desbalanceados Utilizando Modelos de Aprendizado Profundo /
George Oliveira Barros; orientador Flávio de Barros Vidal;
co-orientador Angelo Amâncio Duarte. Brasília, 2025.
93 p.

Tese(Doutorado em Informática) Universidade de Brasília,
2025.

1. Patologia computacional. 2. Aprendizagem profunda. 3.
Classificação de imagens. 4. Podocitopatia. 5. Segmentação
de imagens. I. de Barros Vidal, Flávio , orient. II. Amâncio
Duarte, Angelo, co-orient. III. Título.

Dedicatória

Dedico esta tese aos meus pais, Higino Barros Meira e Jizonete Oliveira Silva Barros, por todo o amor, incentivo e esforço em me proporcionar a oportunidade de trilhar a carreira acadêmica. Por fim, dedico também esta tese à minha esposa, Raquel Elisama Alcântara Barros, por todo o amor, apoio e compreensão.

Agradecimentos

Gostaria de expressar minha profunda gratidão ao meu orientador, Professor Dr. Flávio de Barros Vidal, e ao meu coorientador, Professor Dr. Angelo Amancio Duarte, por todo o conhecimento compartilhado, pelos valiosos conselhos e pela paciência ao longo deste processo. Ambos são verdadeiras referências como pesquisadores, e seus ensinamentos me acompanharão por toda a minha vida profissional.

Agradeço aos membros do grupo de pesquisa Pathospotter, cuja colaboração, direta ou indireta, foi fundamental para a conclusão deste trabalho. Um agradecimento especial ao Professor Dr. Washington Luis Conrado dos Santos, líder e idealizador deste movimento, que tem contribuído de forma significativa para as áreas de computação e patologia, com impacto internacional.

Sou grato também aos patologistas Dr. José Nathan Andrade Muller da Silva e Dr. David Campos Wanderley, que colaboraram ativamente na anotação e revisão dos conjuntos de dados e artigos científicos.

Agradeço ainda às instituições que apoiaram esta pesquisa, tanto com infraestrutura quanto com suporte financeiro: Universidade de Brasília, Universidade Estadual de Feira de Santana, Instituto Federal Goiano (Campus Posse) e o Centro de Pesquisa Gonçalo Moniz - Fiocruz.

Por fim, sou grato a Deus por permitir a concretização de tamanha conquista.

Soli Deo Gloria.

Resumo

A podocitopatia é uma condição caracterizada por lesões nos podócitos, células essenciais dos glomérulos renais responsáveis pela filtragem do sangue. Essas lesões comprometem a função renal e podem evoluir para glomerulosclerose, proteinúria e nefropatia diabética. A identificação de lesões podocitárias em imagens histológicas é desafiadora, devido à dificuldade de distinção celular e ao caráter laborioso do diagnóstico. Este estudo propôs estratégias para aprimorar algoritmos de classificação binária e segmentação semântica de lesões podocitárias em cenários de dados limitados, desbalanceados e multicoloridos. Foram desenvolvidos um classificador de podocitopatia e uma nova abordagem de segmentação baseada na fusão dos espaços de cor RGB e HED, além da criação de um conjunto de dados multicolorante, anotado em quatro classes de lesões podocitárias. Os experimentos indicaram que as soluções propostas elevaram o desempenho dos modelos de classificação, com aumento do *F1-score* de 87,8% para 90,9%. Também foi observado impacto positivo na prática médica, com o índice de concordância entre patologistas (*Fleiss' Kappa*) aumentando de 0,59 para 0,83. Na segmentação, os métodos propostos proporcionaram ganhos médios de até 9,1% no *F1-score* e 8,5% no *IoU*. Os resultados evidenciam o potencial das técnicas desenvolvidas para apoiar o diagnóstico na nefropatologia e fortalecer a aplicação da inteligência artificial na patologia computacional.

Palavras-chave: Patologia computacional, Aprendizagem profunda, Imagens histológicas, Classificação, Segmentação, Glomérulos, Podocitopatia.

Abstract

Podocytopathy is a pathological condition characterized by injuries to podocytes, key cells in the renal glomeruli responsible for blood filtration. These injuries affect renal function and can lead to glomerulosclerosis, proteinuria, and diabetic nephropathy. Identifying podocyte lesions on histological images is challenging due to the difficulty of cellular differentiation and the labor intensive diagnostic process. This study proposed strategies to enhance binary classification and semantic segmentation algorithms for podocyte lesion detection under conditions of limited, imbalanced, and multi-stained datasets. A podocytopathy classifier and a novel segmentation approach based on the fusion of RGB and HED color spaces were developed, along with the creation of a multistained dataset annotated into four podocyte lesion classes. The experimental results showed that the proposed solutions improved the classification performance, increasing the *F1-score* from 87.8% to 90.9%. A positive impact on medical practice was also observed, with the inter-pathologist agreement (*Fleiss' Kappa*) rising from 0.59 to 0.83. In segmentation tasks, the proposed methods produced average gains of up to 9.1% in *F1-score* and 8.5% in *IoU*. These findings highlight the potential of the techniques developed to support diagnostic practices in nephropathology and to advance the integration of artificial intelligence into computational pathology.

Keywords: Computational pathology, Deep learning, Histological images, Classification, Segmentation, Glomeruli, Podocytopathy.

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 1 |
| 1.1 | Contextualização | 1 |
| 1.2 | Motivação e justificativa | 3 |
| 1.3 | Hipóteses de pesquisa | 4 |
| 1.4 | Objetivos | 4 |
| 1.5 | Organização do documento | 5 |
| 2 | Fundamentação Teórica | 6 |
| 2.1 | A estrutura do glomérulo e a podocitopatia | 6 |
| 2.2 | Definições sobre imagens histológicas | 8 |
| 2.3 | Redes Neurais Convolucionais | 11 |
| 2.3.1 | Hiperparâmetros e Componentes das Redes Neurais Convolucionais . | 13 |
| 2.4 | Classificação e Segmentação de Imagens | 16 |
| 2.5 | Redes de Fusão de Características Profundas | 18 |
| 2.6 | Métricas de Avaliação | 19 |
| 3 | Trabalhos Relacionados | 24 |
| 3.1 | Trabalhos com foco em glomérulos renais | 25 |
| 3.2 | Trabalhos com foco em podócitos | 30 |
| 3.3 | Lacunas e desafios identificados | 33 |
| 4 | Metodologia | 34 |
| 5 | Elaboração dos Conjuntos de Dados | 37 |
| 5.0.1 | Conjunto de dados para classificação: D1v.1 | 38 |
| 5.0.2 | Conjunto de dados para classificação: D1v2 e D2 | 39 |
| 5.0.3 | Conjunto de dados para segmentação: D3 | 40 |
| 6 | Identificação de Podocitopatia: PodNet | 43 |
| 6.0.1 | Avaliação dos modelos de referência (<i>baselines</i>) | 43 |

| | |
|---|-----------|
| 6.0.2 Solução proposta: Modelo PodNet | 44 |
| 7 Implicações do Uso de Classificadores Automatizados na Prática Médica | 47 |
| 7.0.1 Treinamento e validação dos modelos de classificação | 48 |
| 7.0.2 Classificação dos patologistas | 50 |
| 8 Segmentação de Lesões Podocitárias | 52 |
| 8.0.1 Avaliação dos modelos de referência (<i>baselines</i>) | 52 |
| 8.0.2 Solução proposta | 54 |
| 9 Resultados | 58 |
| 9.1 Identificação de podocitopatia: PodNet | 58 |
| 9.2 Impacto do uso de classificadores automatizados na prática dos patologistas | 59 |
| 9.3 Segmentação de lesões podocitárias | 65 |
| 9.3.1 Impacto da Solução Proposta | 66 |
| 10 Considerações Finais | 72 |
| 10.1 Perspectivas Futuras | 73 |
| 10.2 Limitações | 74 |
| 10.3 Publicações | 74 |
| Referências | 76 |

Lista de Figuras

| | | |
|-----|---|----|
| 2.1 | Glomérulo renal: (a) imagem observada em microscópio, (b) ilustração com a localização das estruturas internas.. | 7 |
| 2.2 | Glomérulo com (a) e sem (b) lesão podocitária (degeneração).. | 9 |
| 2.3 | Etapas básicas de aquisição de imagens histológicas.. | 9 |
| 2.4 | Exemplos de corantes utilizados em imagens de glomérulos renais: HE (a), PAM (b), PAS (c) e Tricômio (d).. | 10 |
| 2.5 | Exemplo da operação de convolução 2D, demonstrando como um <i>Kernel</i> desliza sobre uma matriz de entrada (<i>Input</i>). Cada elemento da matriz resultante é obtido pela multiplicação ponto a ponto dos valores correspondentes entre o <i>Kernel</i> e uma submatriz da entrada, seguida da soma desses produtos. Esta operação é repetida ao longo da matriz de entrada, resultando em uma nova matriz (<i>Output</i>) de dimensão reduzida, que destaca características específicas da imagem original.. | 12 |
| 2.6 | Arquitetura padrão da rede <i>U-Net</i> . Cada caixa azul representa um mapa de atributos multicanal. As caixas brancas indicam mapas de recursos copiados. As setas representam as diferentes operações realizadas.. | 17 |
| 2.7 | Arquitetura padrão da <i>DeepLav v3+</i> . O módulo codificador captura informações contextuais em múltiplas escalas por meio da convolução atrous em diferentes níveis, enquanto o decodificador eficiente refina a segmentação ao longo dos contornos dos objetos.. | 19 |
| 2.8 | Exemplo de matriz de confusão para um problema de classificação binária.. | 20 |
| 2.9 | Exemplo de mensuração do <i>IoU</i> | 22 |
| 4.1 | Resumo ilustrativo das fases e experimentos realizados no trabalho.. | 35 |
| 5.1 | Etapas de anotação para obtenção dos conjuntos de dados.. | 38 |
| 5.2 | Exemplos de imagens que compõem o conjunto de dados. Com lesão (a-d) e sem lesão (e-h).. | 39 |

| | |
|---|----|
| 5.3 Exemplos de amostras do conjunto de dados. As condições observadas incluem hiperplasia (a), hipertrofia (b), degeneração (c) e ausência de lesões evidentes (d).. | 42 |
| 6.1 Etapas básicas da proposta de solução (PodNet). | 45 |
| 7.1 Visão geral do protocolo experimental. | 48 |
| 8.1 Visão geral da abordagem proposta: (1) conversão do espaço de cores, (2) extração de características usando blocos encoder-decoder, (3) fusão das saídas dos decoders e (4) bloco de segmentação. | 56 |
| 9.1 Curvas ROC: Modelos treinados do zero em RGB. | 60 |
| 9.2 Curvas ROC: Modelos treinados do zero em nível de cinza. | 60 |
| 9.3 Curvas ROC: Modelos treinados com <i>transfer learning</i> em RGB. | 61 |
| 9.4 Curvas ROC: Top-4 melhores modelos. | 61 |
| 9.5 Área sob a curva (AUC) dos modelos ao classificar imagens no conjunto de dados <i>D2</i> . | 63 |
| 9.6 Diagramas de <i>Venn</i> ilustrando a distribuição dos erros de classificação cometidos pelos patologistas com e sem auxílio do classificador automatizado.. | 64 |
| 9.7 Distribuição das arquiteturas de rede (A, B) e codificadores (C, D) entre os 10% melhores e piores modelos de referência.. | 66 |
| 9.8 <i>Boxplot</i> das diferenças no <i>F1-score</i> entre os modelos de referência (<i>Baselines</i>) e a solução proposta utilizando diferentes estratégias de fusão. | 67 |
| 9.9 Predições do melhor modelo obtido com a adoção da solução proposta (Unet++ com Timm-RegNetX) em três amostras do conjunto de teste.. | 68 |

Lista de Tabelas

| | | |
|------|--|----|
| 3.1 | Resumo dos trabalhos correlatos mais relevantes. | 32 |
| 5.1 | Distribuição dos corantes entre as imagens do conjunto de dados <i>D1v2</i> . . . | 40 |
| 5.2 | Distribuição dos corantes entre as imagens do conjunto de dados <i>D2</i> | 40 |
| 5.3 | Distribuição do número de imagens por técnica de coloração no conjunto de dados <i>D3</i> | 41 |
| 7.1 | Intervalo de valores considerados durante o ajuste de hiperparâmetros. . . | 49 |
| 7.2 | Melhores valores de hiperparâmetros para cada modelo avaliado. LR: <i>Learning rate</i> | 49 |
| 8.1 | Proporção de colorações presentes em cada subconjunto. | 53 |
| 8.2 | Operações de fusão avaliadas no método proposto. | 57 |
| 9.1 | Resultados obtidos nos modelos de classificação. TL: <i>transfer learning</i> . FS: <i>from scratch</i> | 59 |
| 9.2 | Resultados da validação cruzada k-fold para os modelos de CNN. Modelos com a sigla TL (<i>Transfer Learning</i>) foram pré-treinados; modelos com a sigla FS (<i>From Scratch</i>) foram treinados com inicialização aleatória de pesos. | 62 |
| 9.3 | Desempenho de três patologistas com e sem o auxílio de um classificador automatizado na classificação de imagens do conjunto <i>D2</i> | 62 |
| 9.4 | <i>IoU</i> e <i>F1-score</i> do grupo <i>Top-10</i> (%) | 69 |
| 9.5 | <i>IoU</i> e <i>F1-score</i> do grupo <i>Random-10</i> (%) | 70 |
| 9.6 | <i>IoU</i> e <i>F1-score</i> do grupo <i>Bottom-10</i> (%) | 71 |
| 10.1 | Artigos produzidos no decorrer do doutoramento. | 75 |

Lista de Abreviaturas e Siglas

AUC *Area Under Curve.*

CNNs *Convolutional Neural Networks.*

DPA *Digital Pathology Association..*

HE *Hematoxylin and Eosin.*

IoU *Intersection Over Union.*

PAM *Periodic Acid Methenamine Silver.*

PAS *Periodic Acid–Schiff.*

ReLU *Rectified Linear Unit.*

RNN *Recurrent Neural Networks.*

ROC *Receiver Operating Characteristic.*

SBN *Sociedade Brasileira de Nefrologia.*

SGD *Stochastic Gradient Descent.*

WSI *Whole Slide Imaging.*

Lista de Símbolos

| | |
|---------------------|--|
| η | Taxa de aprendizado |
| κ | <i>Fleiss' Kappa</i> |
| \oplus | Operação de fusão |
| ∂ | Derivação |
| θ | Rede de segmentação semântica |
| D | Conjunto de dados |
| G | Domínio das matrizes de entrada de uma rede |
| i | Número de iterações |
| L | Valor retornado por uma função de perda |
| M | Resultante do <i>decoder</i> de uma rede de segmentação |
| N | Resultante da fusão entre as saídas de dois <i>decoders</i> |
| S | Bloco de segmentação |
| T | Matriz de transformação |
| t | Total de pixels |
| W | Pesos da rede |
| x | Valor de entrada |
| Y | Máscara resultante do bloco de segmentação de uma rede e segmentação |
| y | Valor de saída |
| F_{fusion} | Conjunto de atributos resultantes de uma fusão |

Capítulo 1

Introdução

1.1 Contextualização

A evolução dos métodos e técnicas computacionais, observada nas últimas décadas, tem contribuído significativamente para a otimização de tarefas em diversas áreas do conhecimento [1]. Na medicina, a colaboração com a computação tem promovido melhorias no diagnóstico e prognóstico de doenças, além de auxiliar na análise de estruturas biológicas [2].

Entre as áreas da medicina com mais exemplos de aplicações computacionais, a patologia se destaca [3, 4]. Este ramo da medicina e biologia destina-se a investigar a causa de doenças por meio da observação visual de células, órgãos, fluidos corporais e tecidos biológicos [5], tendo como principal objeto de estudo as imagens histológicas, que são obtidas principalmente por meio de exames de biópsia e câmeras acopladas a microscópios [6].

A difusão dos métodos computacionais na patologia levou ao surgimento de um novo campo de estudo, conhecido como patologia computacional [7]. Segundo a Associação de Patologia Digital (*Digital Pathology Association* - DPA), a patologia computacional envolve a análise de doenças por meio de uma variedade de métodos computacionais, como redes neurais e algoritmos clássicos de processamento de imagens [4, 8].

Segundo Barisoni *et al.* [9], os trabalhos em patologia computacional se organizam em três grupos: telepatologia, patologia digital e análise computacional de imagens. A telepatologia refere-se à transmissão de imagens microscópicas para patologistas localizados em diferentes localidades. A patologia digital refere-se à criação de ambientes digitais que possibilitam a análise de lâminas inteiras de tecido (*Whole Slide Imaging* - WSI). Por fim, a análise computacional de imagens está associada ao uso de algoritmos de visão computacional, que é um campo de estudo destinado a construir modelos computacionais baseados em dados visuais (imagens e vídeos) [10]. Esses métodos são utilizados na análise

computacional de imagens para extração de características, classificação ou segmentação de estruturas biológicas lesionadas ou doentes.

Entre os trabalhos de análise computacional de imagens, é possível encontrar estudos genéricos, destinados a imagens histológicas em geral, independentemente do órgão [11, 12]. No entanto, de forma mais comum, há trabalhos especializados em diferentes estruturas biológicas, tecidos ou órgãos do corpo humano [13, 14, 15].

Na patologia renal, especificamente, observa-se um número crescente de estudos voltados à análise computacional de imagens [16]. Muitos desses estudos caracterizam-se pelo uso de métodos automatizados para auxiliar na análise de transplantes [17], segmentação de estruturas [12] ou análise de glomérulos [18, 19, 20], que são as principais estruturas do rim responsáveis pela filtração do sangue [21].

Ao analisar os trabalhos na área renal, focados na análise de glomérulos, observa-se que, apesar do crescente interesse dos pesquisadores, refletido no aumento do número de publicações, ainda há lacunas significativas, tanto metodológicas quanto em relação a doenças pouco exploradas. [22, 23].

Em relação às lacunas metodológicas, destaca-se a necessidade de uma grande massa de dados, com distribuição balanceada de amostras por classe, adquiridos de diferentes laboratórios, utilizando variados corantes e devidamente anotados por especialistas humanos (patologistas) [23]. Esses desafios surgem principalmente porque a maioria dos métodos propostos se baseia em aprendizagem profunda (*deep learning*), que é um sub-campo do aprendizado de máquina (*machine learning*) que se concentra no uso de redes neurais artificiais com múltiplas camadas para modelar e aprender representações de dados complexos [24]. Portanto, estudos que ofereçam novas estratégias de aprendizado para enfrentar essas dificuldades podem promover avanços na elaboração de técnicas, tanto para imagens de glomérulos renais quanto para outros tipos de imagens histológicas.

Em relação às doenças pouco exploradas, é possível destacar o desafio de segmentação de lesões e estruturas glomerulares [25], a predição e o apoio ao estudo de doenças crônicas [22, 26]. Entre essas lacunas, encontra-se o desafio de segmentar os diferentes tipos de lesões em podócitos, células do tecido epitelial visceral do glomérulo [27]. O estudo dessas lesões podocitárias é fundamental, pois os patologistas enfrentam dificuldades para diferenciar podócitos das demais células intra-glomerulares em imagens de microscopia de luz. Além disso, os podócitos são biomarcadores importantes para doenças como glomerulosclerose progressiva, diabetes e proteinúria [28, 29, 30, 31, 32].

Até o momento, foram identificados apenas seis estudos com o objetivo de detectar podocitopatia em imagens de glomérulos renais, e ainda menos propostas de segmentação de podócitos com base no tipo de lesão. Entre os trabalhos encontrados, nenhum dos métodos propostos realiza a tarefa de informar ao usuário quando um glomérulo possui

ou não podocitopatia, concentrando-se principalmente na detecção e quantificação de podócitos, associando-os a outras doenças [28, 33, 34, 31].

Considerando o contexto deste estudo e as lacunas apresentadas, pode-se concluir que o presente trabalho se situa em uma área de interseção entre a patologia computacional e a visão computacional. As contribuições pretendidas visam atacar ambas as lacunas: computacional e de aplicação, fornecendo, assim, (1) novas ferramentas para a patologia e (2) novos métodos e abordagens computacionais para análise automática de imagens histológicas de glomérulos renais em relação à presença de podocitopatia.

1.2 Motivação e justificativa

Este trabalho possui duas motivações principais: A primeira está relacionada à importância de desenvolver novas metodologias e abordagens para a análise de imagens de glomérulos renais, visando superar os desafios já identificados nos estudos recentes de imagens histológicas. A segunda motivação refere-se às possíveis aplicações na patologia, com ênfase na relevância do estudo da podocitopatia.

- **Motivação Computacional.** Considerando que os algoritmos comumente utilizados se baseiam em técnicas de aprendizagem profunda [35, 36, 17], é necessário adquirir conjuntos de dados compostos por uma ampla diversidade de imagens histológicas de glomérulos renais com lesões podocitárias. Isso se deve ao fato de que as imagens histológicas dos glomérulos são obtidas utilizando diferentes corantes, equipamentos de aquisição (como câmeras e escâneres) e protocolos [37]. Portanto, para que os modelos sejam adequadamente treinados e validados, esses aspectos devem ser devidamente considerados.

A construção de conjuntos de dados que satisfaçam esses critérios representa um desafio significativo. A podocitopatia, por exemplo, é uma doença que ainda não possui um conjunto de dados publicado, muito menos um que contenha uma variedade de amostras que satisfaça os requisitos mencionados. Dessa forma, a criação de novos conjuntos de dados pode impulsionar pesquisas no campo da patologia computacional.

No entanto, mesmo com a elaboração de conjuntos de dados que garantam a diversidade de amostras, em casos de algumas doenças, como a podocitopatia, não é possível assegurar uma distribuição proporcional entre as diferentes classes presentes no conjunto de dados, tampouco garantir a mesma distribuição em relação aos corantes utilizados. Assim, é necessário avaliar os métodos tradicionais de aprendi-

zado e desenvolver alternativas que melhorem o desempenho desses métodos diante da complexidade dos dados.

- **Motivação Aplicacional para Patologia.** Em relação à importância do estudo da doença, é relevante considerar que as lesões nos podócitos são biomarcadores de diversas outras doenças, como glomerulosclerose progressiva, tumor de Wilm's e diabetes nefrótica [29, 28, 30, 31]. Portanto, métodos e técnicas de análise automática da podocitopatia em imagens de glomérulos podem se tornar ferramentas extremamente úteis para os nefropatologistas.

Esses métodos têm o potencial de reduzir a subjetividade das análises, acelerar o processo de diagnóstico e aprimorar a precisão das decisões médicas. Além disso, podem ser utilizados como ferramentas de apoio na formação e no treinamento de novos patologistas [38].

Como resultado deste trabalho, espera-se que os dados gerados, baseados em glomérulos com podocitopatia, fortaleçam grupos de pesquisa e possibilitem melhorias nas técnicas de análise de imagens histológicas, tanto por meio de abordagens tradicionais de aprendizagem profunda quanto por meio de novas estratégias analíticas, além de gerar conjuntos de dados relevantes para pesquisadores nas áreas de patologia e computação.

1.3 Hipóteses de pesquisa

As seguintes perguntas norteiam este trabalho e sintetizam o problema de pesquisa:

1. Em um contexto de dados desbalanceados e multicoloração, qual é o desempenho alcançado pelos métodos baseados em aprendizagem profunda, recorrentemente utilizados em trabalhos similares, nas tarefas de classificação de glomérulos quanto à presença de podocitopatia e segmentação de lesões podocitárias em imagens de glomérulos renais?
2. Neste mesmo contexto de dados, é possível elaborar novas abordagens a partir dos modelos já existentes que, no entanto, superem os resultados obtidos por esses métodos convencionais?

1.4 Objetivos

O objetivo geral deste trabalho é propor novas soluções para o problema da análise automatizada de podocitopatia em imagens de glomérulos renais, superando os resultados obtidos por métodos computacionais baseados em aprendizagem profunda, tradicionalmente

utilizados na literatura especializada, no contexto de conjuntos de dados desbalanceados e multicoloração.

Os objetivos específicos deste trabalho são:

1. Fornecer novos conjuntos de dados sobre lesões podocitárias, que simulem cenários de multicoloração e desbalanceamento, contribuindo para o avanço dos estudos na comunidade de pesquisadores de patologia computacional.
2. Aperfeiçoar o desempenho de modelos de classificação binária, de identificação da presença ou ausência de podocitopatia, a partir desses dados por meio de soluções metodológicas inovadoras.
3. Estudar o impacto do uso de classificadores automatizados na prática dos patologistas.
4. Propor modificações de topologia e arquitetura para aprimorar o desempenho de modelos de segmentação semântica, considerando o conjunto de dados desenvolvido.

1.5 Organização do documento

Este documento está organizado da seguinte forma: No Capítulo 2, são apresentadas as definições relacionadas a imagens histológicas, podocitopatia e às técnicas computacionais frequentemente utilizadas para sua análise automática. No Capítulo 3, são descritos os trabalhos correlatos, organizados em seções que gradativamente se aproximam do problema da podocitopatia. No Capítulo 4, é detalhada a metodologia e as contribuições propostas. No capítulo 5, são apresentados os resultados obtidos. No Capítulo 6, são apresentadas as conclusões e os trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Neste capítulo, são apresentadas as definições fundamentais relacionadas ao escopo deste trabalho. Inicialmente, introduzem-se conceitos fundamentais sobre o processo de aquisição de imagens histológicas, seguidos de uma breve explicação sobre a estrutura interna dos glomérulos e informações acerca da podocitopatia. Em seguida, discutem-se as definições associadas às técnicas e métodos computacionais utilizados na análise automática de imagens histológicas, com ênfase em abordagens baseadas em aprendizagem profunda (*deep learning*).

2.1 A estrutura do glomérulo e a podocitopatia

Os rins, órgãos do sistema urinário, são responsáveis pela formação da urina e possuem dimensões de aproximadamente 10 a 13 cm de comprimento, peso entre 120 e 180 g, e estão localizados na região lombar, acima da cintura, um de cada lado da coluna vertebral [39].

A Sociedade Brasileira de Nefrologia (SBN) esclarece que os rins desempenham quatro funções principais no organismo: eliminação de toxinas do sangue por meio de um sistema de filtração; regulação da formação do sangue e dos ossos; controle da pressão arterial; e manutenção do balanço químico e hídrico do corpo [40].

Ainda segundo a SBN, as doenças renais têm um impacto significativo na população, gerando um gasto de 2,2 bilhões de reais por ano apenas com hemodiálise [41], pois o mau funcionamento dos rins frequentemente leva à necessidade de diálise. Na maioria dos casos, o tratamento deve ser mantido por toda a vida, a menos que o paciente possa ser submetido a um transplante renal [40]. Um dado alarmante é que, anualmente, cerca de 21 mil brasileiros precisam iniciar tratamento por hemodiálise¹ [42] [43].

¹Processo artificial que substitui a função dos rins quando estes não conseguem mais filtrar as impurezas e excesso de líquidos do sangue [40].

Anatomicamente, os rins possuem duas zonas principais: córtex e medula renal. É no córtex que se encontram os néfrons, unidades funcionais dos rins, responsáveis pela formação da urina. Estima-se que um rim humano contenha entre 800 mil e 1 milhão de néfrons, cada um composto por numerosos glomérulos que desempenham um papel ativo na filtração sanguínea [39, 44].

Os néfrons são estruturas que não se regeneram [44]. Dessa forma, lesões renais, doenças ou o próprio processo de envelhecimento levam à redução progressiva do número de néfrons. Cada néfron é composto por um grupo de capilares glomerulares denominado glomérulo, responsável por filtrar grandes quantidades de líquido do sangue, e por dois longos túbulos, onde o líquido filtrado é convertido em urina [45]. O glomérulo é constituído por uma rede de capilares glomerulares envolvidos por células epiteliais e circundado pela cápsula de Bowman [45], apresentado na Figura 2.1.

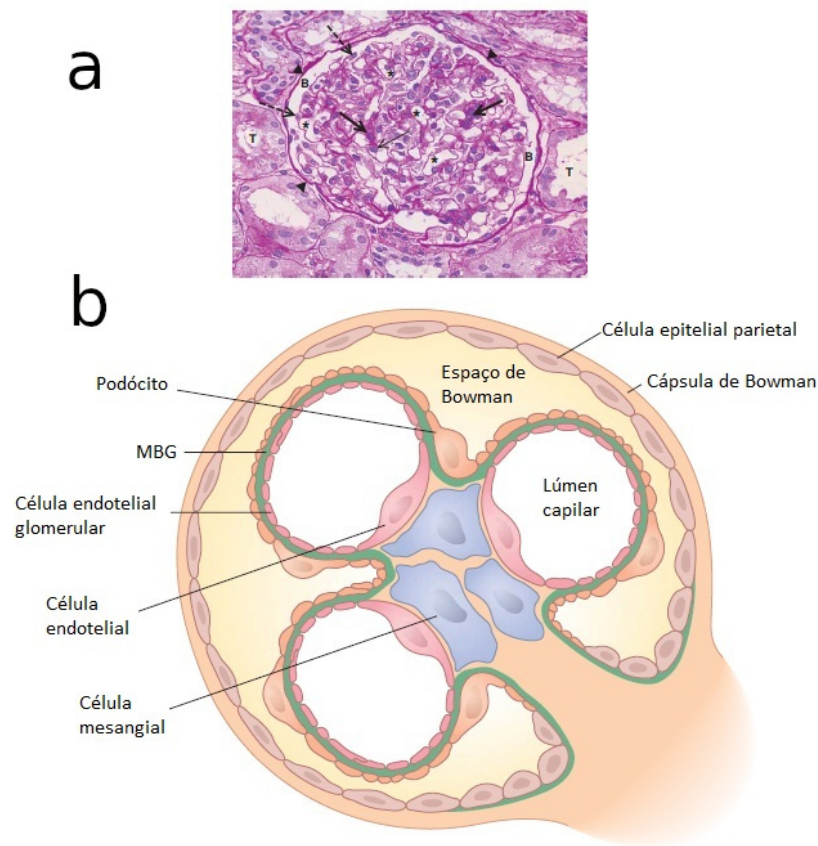


Figura 2.1: Glomérulo renal: (a) imagem observada em microscópio, (b) ilustração com a localização das estruturas internas. (Fonte: [46]).

Os glomérulos possuem, em sua estrutura interna, diferentes tipos de células: mesangiais, endoteliais, epiteliais parietais e epiteliais viscerais (também conhecidas como podócitos) [39]. A principal função dos podócitos é restringir a passagem de proteínas do sangue

para a urina [27]. Assim, lesões podocitárias podem comprometer a capacidade filtrante do glomérulo devido aos danos causados à estrutura glomerular [21]. Adicionalmente, lesões em podócitos são biomarcadores de uma série de doenças, como: glomeruloesclerose progressiva, tumor de Wilm's, diabetes e proteinúria [28, 29, 30, 31, 32].

De acordo com Fuad *et al.* [47], lesões podocitárias também estão associadas ao fator de crescimento endotelial vascular. Além disso, a redução do número e/ou densidade de podócitos, como resultado de apoptose ou descolamento, espessamento da membrana basal com composição de matriz alterada e diminuição da proteína nefrina. Essa condição manifesta-se clinicamente como albuminúria e proteinúria. A proteinúria, por sua vez, pode induzir um programa genético nas células tubulares que leva à inflamação tubulointersticial, fibrose e atrofia tubular.

Em um estudo conduzido por Trimarchi e Coppo [48], destaca-se que a identificação de podocitopatias também está relacionada à nefropatia por imunoglobulina A (IgAN), definida como uma doença mesangiopática. Os autores indicam que a progressão da IgAN deve-se, principalmente, ao desenvolvimento de proteinúria persistente e a função e anatomia dos podócitos desempenham um papel importante nesse processo. Na IgAN, as alterações podocitopáticas são consequência de danos iniciais na região mesangial. Os podócitos são afetados por interações de sinais originados do mesângio e, após os danos, destacam-se da membrana basal glomerular. Essa podocitopatia favorece não apenas o desenvolvimento de esclerose glomerular focal e segmentar, mas também a perda progressiva da função renal.

A podocitopatia também está associada à doença de Lupus [49] e à nefropatia diabética (ND) [50]. A ND é caracterizada pela perda da função podocitária do glomérulo renal. Nos Estados Unidos, os gastos com ND entre 2015 e 2021 avançaram de 748 mil dólares para mais de 2 bilhões e meio de dólares por ano [51].

Por fim, os podócitos contribuem para a integridade da barreira de filtração glomerular, e evidências recentes sugerem que essas células podem ser alvos diretos de hormônios, lipídios e adipocinas circulantes afetados no diabetes [50, 49].

Ao analisar lesões podocitárias [21], pode-se observar 3 tipos básicos de lesões. Este trabalho concentra-se nestes três tipos: degeneração, hiperplasia e hipertrofia. A Figura 2.2 apresenta 2 exemplos de imagens de glomérulos renais, destacando podócitos com lesão degenerativa (a) e podócitos saudáveis (b).

2.2 Definições sobre imagens histológicas

As imagens histológicas são representações visuais de tecidos biológicos, obtidas a partir de técnicas específicas de preparo e coloração [35]. Segundo Caputo *et al.* [37], o processo

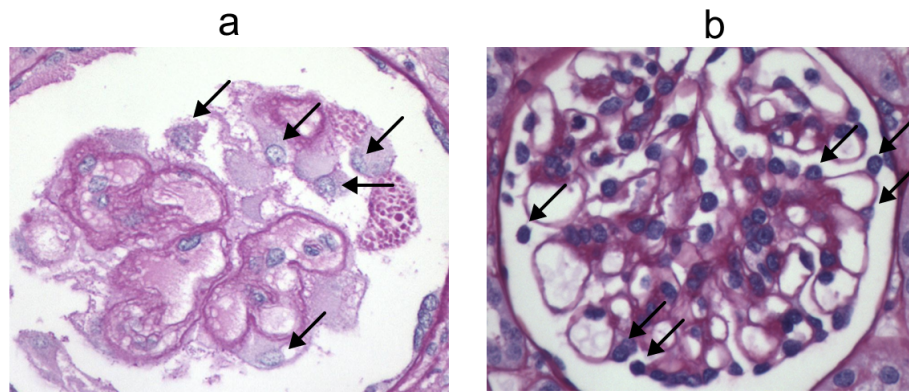


Figura 2.2: Glomérulo com (a) e sem (b) lesão podocitária (degeneração) (Fonte: [52]).

para obtenção de amostras de tecido biológico envolve as seguintes etapas básicas: coleta, fixação, clivagem, processamento, inclusão, microtomia (corte) e coloração. Após a preparação das lâminas de tecido, a captura das imagens digitais é realizada utilizando uma câmera acoplada ao microscópio. A Figura 2.3 ilustra cada etapa do processo de aquisição destas.

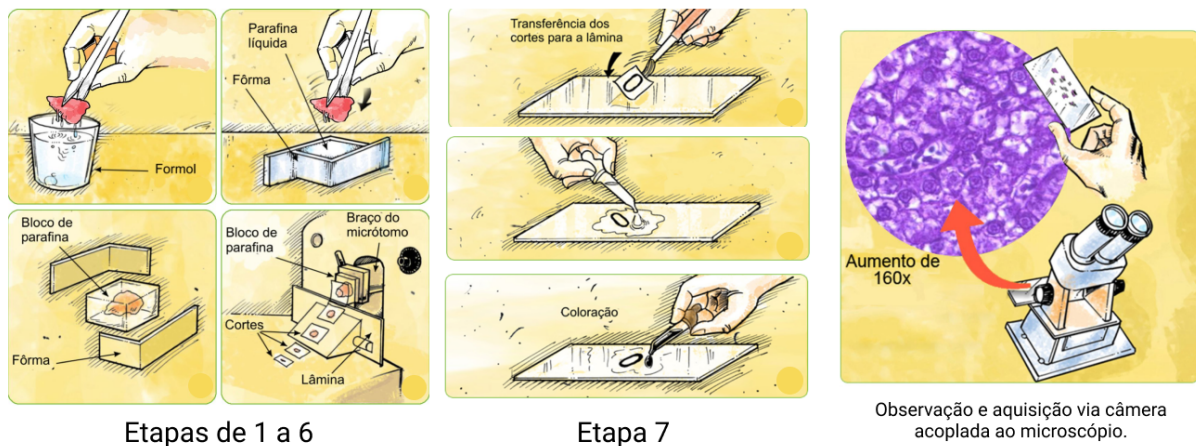


Figura 2.3: Etapas básicas de aquisição de imagens histológicas (Fonte: [53]).

Silvano *et al.* [37], descreve as etapas de aquisição de imagens histológicas: *Coleta* (Etapa 1), consiste em remover amostras de tecido de um organismo por meio de biópsia, que é o procedimento de remoção de uma pequena porção de tecido para análise microscópica. *Fixação* (Etapa 2) é realizada para interromper o processo de degradação do tecido (metabolismo celular) e preservar seus elementos estruturais. Durante a *clivagem* (Etapa 3), o objetivo é reduzir as dimensões dos fragmentos de tecido para tamanhos entre 3

mm e 5 mm. O *processamento* (Etapa 4) promove a difusão de reagentes no interior dos tecidos e a remoção de líquidos teciduais.

A etapa de *inclusão* (Etapa 5) consiste em inserir os tecidos processados em parafina, utilizando uma pinça aquecida. Na microtomia (Etapa 6), os tecidos são cortados em fatias extremamente finas, com espessura entre 4 μm e 6 μm , permitindo sua análise em microscopia de luz. Por fim, na etapa de coloração (Etapa 7), compostos orgânicos são aplicados às lâminas de tecido para atribuir cores específicas às diferentes estruturas. Essa etapa é essencial, pois tanto as células quanto o material extracelular são geralmente transparentes. A diferenciação por cores auxilia patologistas na análise das lâminas [37]. A Figura 2.4 apresenta imagens histológicas de glomérulos renais coradas com diferentes tipos de corantes, cada um destacando estruturas específicas do tecido.

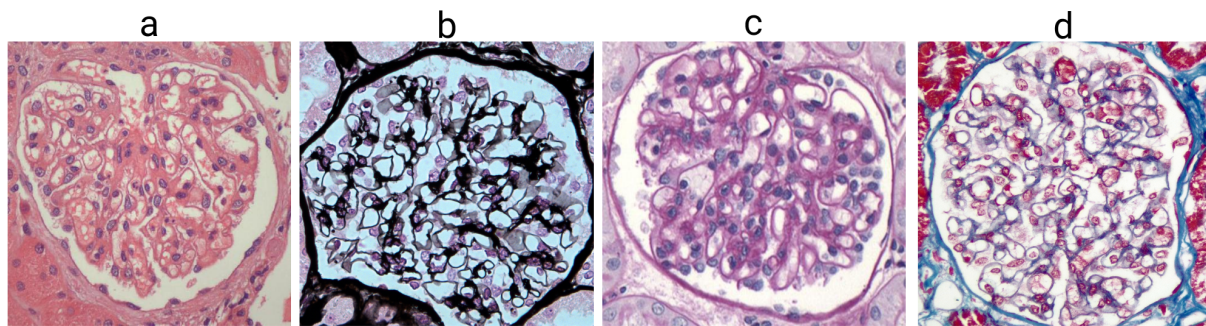


Figura 2.4: Exemplos de corantes utilizados em imagens de glomérulos renais: HE (a), PAM (b), PAS (c) e Tricômio (d) (Fonte: [52]).

Além do uso de diferentes corantes, a escolha de variados espaços de cores pode melhorar significativamente a capacidade de redes neurais profundas para extrair características relevantes de imagens histológicas [54, 55]. O espaço de cores HED (Hematoxilina, Eosina e Diaminobenzidina) [56], específico para imagens histológicas, foi projetado para separar os principais componentes histológicos: núcleo (corado pela Hematoxilina), citoplasma (corado pela Eosina) e regiões coradas por técnicas de imunohistoquímica (Diaminobenzidina) [57].

A transformação de uma imagem do espaço de cores RGB para o espaço HED é realizada através da operação *color deconvolution* [56]. A conversão pode ser expressa matematicamente da seguinte forma [56, 58].

Os vetores de corantes representam as cores características de cada corante usado na coloração histológica. Para o espaço de cor HED, os vetores típicos são:

- **Hematoxilina (H):** (0, 65; 0, 70; 0, 29)
- **Eosina (E):** (0, 07; 0, 99; 0, 11)

- **DAB (D):** (0, 27; 0, 57; 0, 78)

A matriz de transformação T é construída empilhando esses vetores como colunas:

$$T = \begin{bmatrix} 0,65 & 0,07 & 0,27 \\ 0,70 & 0,99 & 0,57 \\ 0,29 & 0,11 & 0,78 \end{bmatrix} \quad (2.1)$$

Para cada pixel na imagem RGB, representado como um vetor $I_{RGB} = (R, G, B)^T$, a conversão para o espaço HED é realizada multiplicando o vetor I_{RGB} pela matriz de transformação T :

$$I_{HED} = T \cdot I_{RGB} \quad (2.2)$$

sendo $I_{HED} = (H, E, D)^T$.

Após a conversão, cada canal H , E e D no vetor I_{HED} representa a intensidade relativa dos corantes Hematoxilina, Eosina e DAB, respectivamente, na imagem original.

2.3 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) surgiram a partir de estudos sobre a percepção visual em organismos biológicos, com experimentos conduzidos por dois pesquisadores: Hubel e Wiesel, na década de 1960 [59]. Eles demonstraram como neurônios no córtex visual de gatos respondem seletivamente a padrões específicos, como bordas e orientações. A partir desse conceito, foi desenvolvido o *Neocognitron*, proposto por Fukushima [60], que se tratava de uma arquitetura baseada em camadas hierárquicas para reconhecer padrões visuais. No entanto, apenas no final da década de 80, que Lecun *et al.* [61] aplicaram o algoritmo de retropropagação para treinar redes convolucionais de forma eficiente, resultando no *LeNet-5*, um modelo pioneiro para reconhecimento de dígitos manuscritos.

Com o avanço do poder computacional e a disponibilidade de grandes conjuntos de dados, as CNNs se tornaram fundamentais para aplicações em visão computacional, o que culminou no sucesso do modelo *AlexNet* em 2012 [62], revolucionando o campo da inteligência artificial ao vencer a competição ImageNet [63] com um desempenho superior às abordagens tradicionais.

Em seu funcionamento, as redes neurais convolucionais são caracterizadas por serem um tipo específico de rede neural voltada para o processamento de dados com topologia em grade [24, 59]. As imagens são um exemplo desses dados. A principal característica

desse tipo de rede é a utilização da operação de convolução em pelo menos uma de suas camadas [24].

Uma camada de convolução realiza, essencialmente, o produto escalar entre duas matrizes: a janela de convolução (*Kernel*) e o campo receptivo, que corresponde a uma porção da entrada (como uma região de uma imagem). Para imagens compostas por três canais (RGB, por exemplo), a altura e a largura do *Kernel* são tipicamente pequenas, mas sua profundidade deve corresponder ao número de canais da entrada [24, 64]. O computo da operação de convolução em uma região da imagem é ilustrada na Figura 2.5, e o resultado desse processamento (*Output*) é denominado mapa de ativação.

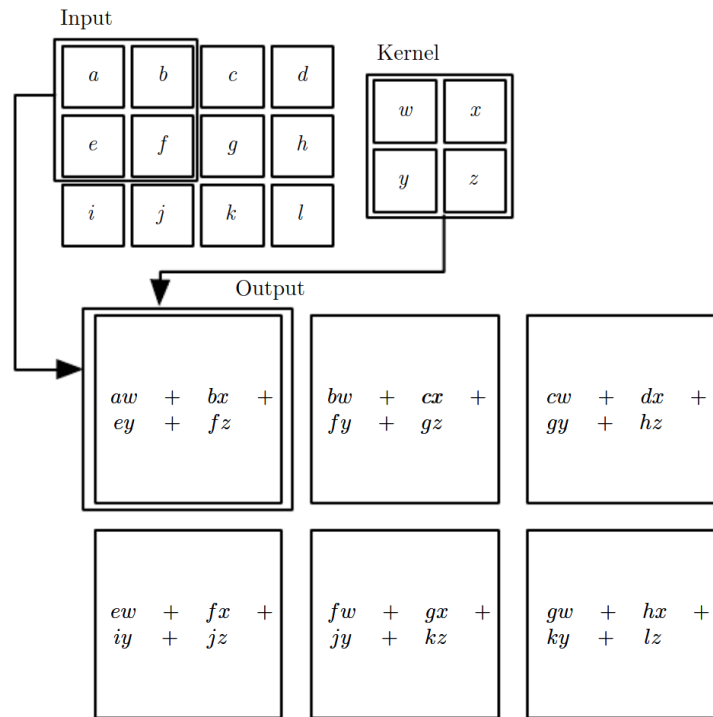


Figura 2.5: Exemplo da operação de convolução 2D, demonstrando como um *Kernel* desliza sobre uma matriz de entrada (*Input*). Cada elemento da matriz resultante é obtido pela multiplicação ponto a ponto dos valores correspondentes entre o *Kernel* e uma submatriz da entrada, seguida da soma desses produtos. Esta operação é repetida ao longo da matriz de entrada, resultando em uma nova matriz (*Output*) de dimensão reduzida, que destaca características específicas da imagem original. (Fonte: [24]).

Após a aplicação da convolução, os mapas de ativação passam por uma função de ativação não linear, como a ReLU (*Rectified Linear Unit*) [65], que introduz não linearidades ao modelo, permitindo a rede aprender representações mais complexas e ajudar a evitar o problema de desaparecimento de gradiente (*vanishing gradient*). Além disso, operações como subamostragem (ou *pooling*) [66] são frequentemente empregadas para reduzir as dimensões espaciais dos mapas de ativação, o que diminui a quantidade de pa-

râmetros e o custo computacional, ao mesmo tempo em que preserva as informações mais relevantes. Essas camadas são empilhadas em profundidade, criando uma hierarquia de características, desde padrões simples, como bordas e texturas, nas camadas iniciais, até representações mais abstratas, como objetos completos, nas camadas mais profundas [24]. Esse processo de extração e hierarquização de características é fundamental para o bom desempenho das redes convolucionais em tarefas como classificação de imagens, detecção de objetos e segmentação semântica [24, 64].

2.3.1 Hiperparâmetros e Componentes das Redes Neurais Convolucionais

Há uma série de hiperparâmetros e componentes necessários no contexto de uma rede neural convolucional. Essas definições são variáveis ou funções escolhidas antes do treinamento do modelo e influenciam diretamente seu desempenho, capacidade de generalização e eficiência computacional [67].

Alguns desses hiperparâmetros são as funções de otimização, perda e ativação, tamanho do lote (*batch size*) e taxa de aprendizado (*learning rate*):

Os otimizadores são métodos utilizados para atualizar os pesos e a taxa de aprendizado durante o treinamento, com o objetivo de reduzir a Função de perda (*Loss function*) do modelo. O otimizador mais simples utilizado em redes neurais é o Gradiente Descendente Estocástico (*Stochastic Gradient Descent* - SGD) [68]. Este é um algoritmo de primeira ordem, pois calcula a derivada de primeira ordem da função de perda. Através do processo de retropropagação (*backpropagation*), a perda é propagada de camada para camada, e os pesos são ajustados [24]. Outros métodos de otimização incluem, por exemplo, *Adaptive Gradient Algorithm* (*Adagrad*) [69], *Root Mean Square Propagation* (*RMSprop*) [70] e (*Adadelta*) [71].

O *Adagrad* é uma modificação do algoritmo SGD, que ajusta a taxa de aprendizado para cada parâmetro do modelo, ao invés de usar uma única taxa de aprendizado, fixa, para todo o modelo [69]. Essa abordagem pode melhorar a convergência, especialmente quando os dados são esparsos [69, 24]. O *RMSprop*, por sua vez, adapta a taxa de aprendizado de cada peso dividindo a taxa de aprendizado por uma média das magnitudes dos gradientes recentes para aquele peso [70].

As funções de perda (*loss functions*) [24] são utilizadas em conjunto com os otimizadores com a finalidade de calcular repetidamente a perda de um dado modelo a cada atualização dos pesos da rede. Há funções adequadas para os diferentes problemas de aprendizado. Em problemas de classificação geralmente utiliza-se as seguintes funções:

Binary cross-entropy [72] (para problemas binários), *Hinge* [73] e *Cross-entropy* (para classificação em problemas multiclasse) [24].

Diferentes funções de ativação podem ser usadas em redes convolucionais, como *Sigmoid* [74], Tangente hiperbólica (*Tanh*) [75] e ReLU (*Rectified Linear Unit*) [76]. Essas funções são aplicadas diretamente após a camada convolucional [77]. Elas se diferenciam em relação às suas fórmulas e ao intervalo dos valores de saída. Por exemplo, a *ReLU* retorna valores no intervalo $[0, \infty)$, enquanto a tangente hiperbólica tem saída no intervalo $(-1, 1)$.

O *batch size* [67] é o número de amostras processadas antes que os pesos do modelo sejam atualizados. O conjunto total de amostras é dividido em lotes de treinamento, os quais têm o seu tamanho definido através deste hiperparâmetro [24]. Esta escolha é bem importante no processo de treinamento de uma rede neural. Uma escolha de valor alto para o *batch size* pode possibilitar acelerações computacionais do paralelismo das GPUs. Adicionalmente, usar um *batch size* igual a todo o conjunto de treinamento pode garantir a convergência para o ótimo global da função [67]. Por outro lado, o uso de tamanhos menores tende a gerar uma convergência mais rápida, que se explica intuitivamente pelo fato de valores de *batch size* menores permitirem que um dado modelo inicie o aprendizado antes de processar todos os dados [78].

A taxa de aprendizado (*learning rate*) é um valor associado à função de otimização e implica em dizer o quão rápido ou lento um dado modelo aprende uma tarefa, ou em outras palavras, o tamanho dos passos que a função de otimização dá até atingir os valores mínimos de perda do modelo [24]. A Equação 2.3 descreve a atualização de pesos da rede e também mostra o papel da taxa de aprendizado (representada como η) no cálculo dos novos pesos. Os demais elementos da Equação 2.3 representam: o novo valor de peso (W_n), o valor antigo do peso (W_o), o valor retornado pela função de perda (L) e a derivada parcial de L em relação ao W_o ($\frac{\partial L}{\partial W_o}$),

$$W_n = W_o - \eta \frac{\partial L}{\partial W_o}. \quad (2.3)$$

Entre os componentes fundamentais para o treinamento de um modelo de CNN, também estão as operações de *pooling* e de regularização.

A operação de *pooling*, por exemplo, reduz o tamanho espacial da representação do resultado da convolução, diminuindo a quantidade necessária de computação e pesos. Entre as operações de *pooling*, o *max pooling* é uma estratégia comum em diversas arquiteturas de rede [79, 80, 81]. Nesse caso, a redução da dimensão é feita selecionando os maiores valores na aplicação do *kernel*, o que, teoricamente, filtra os atributos de maior importância [64]. No que diz respeito às funções de ativação, elas padronizam os valores de entrada

e têm como objetivo inserir não-linearidade na rede. Como a convolução é uma operação linear, a não-linearidade torna possível que a rede resolva problemas não triviais.

As operações de regularização visam reduzir o risco de sobreajuste (*overfitting*), que ocorre quando uma rede neural continua melhorando o desempenho nos dados de treinamento, mas apresenta um desempenho inferior nos dados de teste [82]. Uma das técnicas mais comuns de regularização em redes neurais é o *Dropout* [83], que desativa uma porcentagem dos neurônios durante o treinamento, impedindo que seus pesos sejam atualizados em todas as épocas.

Alguns exemplos de arquiteturas conhecidas de redes neurais convolucionais são as *ResNets* [84], *DenseNets* [85], *Inception* [80], *VGG* [79], *Inception-ResNet* [81], *Xception* [86] e as *EfficientNets* [87]. Cada uma dessas arquiteturas adota diferentes estratégias e mecanismos de aprendizado, inserindo novas operações e aprimoramentos nas redes convolucionais tradicionais, o que resulta em desempenhos ainda mais precisos em tarefas de classificação, regressão e segmentação de imagens.

Ajuste de hiperparâmetros

Para alcançar bons resultados e garantir uma generalização satisfatória nas CNNs, é fundamental escolher adequadamente seus hiperparâmetros. A escolha inadequada de um hiperparâmetro, como a taxa de aprendizado, por exemplo, pode impedir que um modelo encontre os valores mínimos da função de perda. Um modelo é composto não apenas por sua arquitetura de rede e pelos pesos obtidos no treinamento, mas também por uma combinação específica de hiperparâmetros [82].

As estratégias para a escolha de hiperparâmetros podem variar desde abordagens manuais baseadas em um processo intuitivo até métodos automatizados, como a *otimização bayesiana* [88] e *algoritmos genéticos* [89]. A seguir, apresentamos uma descrição intuitiva de algumas dessas estratégias: *Grid search* [90], *Random search* [91] e *Hyperband* [88].

O método *Grid search* (busca em grade) é uma estratégia de busca exaustiva. O ajuste dos hiperparâmetros ocorre da seguinte maneira: inicialmente, define-se um espaço de busca com intervalos fixos para cada hiperparâmetro. Em seguida, o espaço de busca de cada hiperparâmetro é discretizado, assim como o espaço de busca geral, que se torna o produto cartesiano dos espaços de cada hiperparâmetro. No final, o algoritmo treina um modelo para cada configuração de hiperparâmetros e, ao final, seleciona a melhor configuração. Este método requer grande poder computacional, pois é necessário treinar vários modelos (um para cada configuração), além de ser suscetível à *maldição da dimensionalidade* (número exponencial de configurações a serem testadas) [90].

O método *Random search* (busca randômica) é uma variação do método anterior e também pode ser classificado como uma busca exaustiva. A diferença fundamental é

que, em vez de discretizar o espaço de busca com uma grade cartesiana, o algoritmo realiza uma escolha aleatória dentro desse espaço. A busca é finalizada após um tempo previamente definido. Embora também sofra com a maldição da dimensionalidade, este método costuma apresentar melhores resultados do que a busca em grade [91].

Por fim, o método de ajuste *Hyperband*, se baseia na abordagem de otimização de hiperparâmetros proposta em [92]. Sua intuição assemelha-se a uma competição em busca da melhor combinação de hiperparâmetros. Inicialmente, define-se o número (i) de iterações nas quais as combinações de hiperparâmetros serão avaliadas. Em seguida, o algoritmo calcula aleatoriamente uma amostra de 64 conjuntos de hiperparâmetros a partir do espaço de busca. A partir desse momento, a perda no conjunto de validação associada a essas combinações é avaliada consecutivamente após $i/2$ iterações. Após a avaliação, metade das combinações com menor desempenho são descartadas. As combinações restantes são avaliadas novamente por mais $i/2$ iterações. Esse processo de descarte é repetido até que o modelo gerado pela melhor combinação seja o único restante [88].

2.4 Classificação e Segmentação de Imagens

A tarefa de *classificação de imagens* é um problema na área de visão computacional [93, 10], que consiste em associar um rótulo (ou classe) a imagens de um conjunto de teste, sendo que o modelo não avaliou esses rótulos previamente. Nas últimas décadas, muitos dos métodos mais eficazes para a classificação de imagens em diferentes domínios têm sido baseados em redes neurais convolucionais (CNNs) [94, 95]. No campo da Patologia Computacional, essa técnica tornou-se o estado da arte tanto para *classificação* quanto para *detecção e segmentação* de imagens histológicas [96, 97].

A *segmentação de imagens* é uma tarefa do processamento digital de imagens e visão computacional que consiste em dividir uma imagem em regiões ou partes homogêneas, com base em características como cor, textura ou intensidade, por exemplo, com o objetivo de facilitar a análise e interpretação [98, 10]. Dentro deste contexto, a segmentação semântica refere-se à classificação de cada pixel da imagem em uma categoria específica, atribuindo a cada região ou objeto um rótulo correspondente ao seu significado [99]. Diferentemente da segmentação convencional, que apenas separa a imagem em áreas distintas, a segmentação semântica busca entender o conteúdo semântico de cada parte da imagem, tornando-a mais útil para tarefas como reconhecimento de objetos, navegação autônoma e diagnóstico médico [100].

A utilização das CNNs na tarefa de segmentação é viabilizada por arquiteturas específicas para este fim. Ao pesquisar os algoritmos de segmentação semântica baseados em redes neurais convolucionais, comumente utilizados em imagens histológicas, destacam-se

duas principais arquiteturas: a *U-Net* [101] e a *DeepLabv3* [42]. Essas arquiteturas de rede também aparecem na literatura com versões modificadas, que alteram o número de camadas, os *encoders* (redes de classificação utilizadas) e a inserção de mecanismos de atenção [102].

Compreender a *U-Net* [103] ajuda a entender intuitivamente como a maioria das arquiteturas de redes para segmentação semântica se organizam. A *U-Net* é composta por dois blocos principais: o Codificador (*Encoder*) e o Decodificador (*Decoder*). A Figura 2.6 ilustra a arquitetura padrão da *U-Net*. O bloco à esquerda refere-se ao Codificador, enquanto os blocos à direita correspondem ao Decodificador.

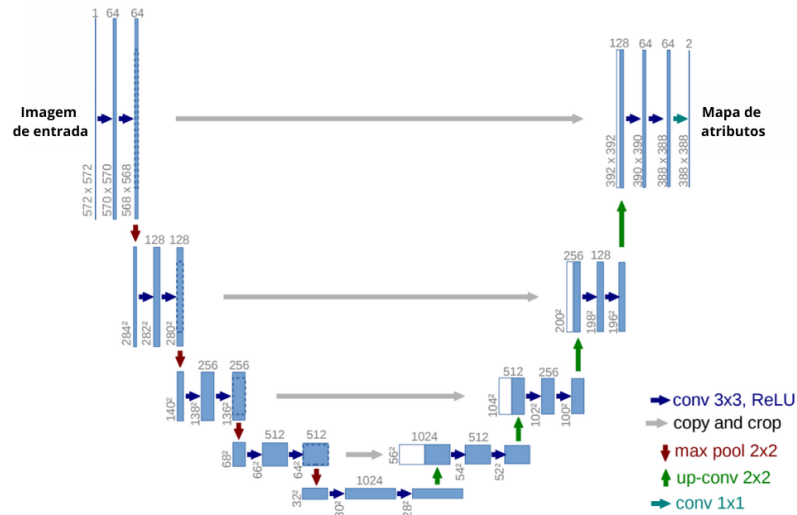


Figura 2.6: Arquitetura padrão da rede *U-Net*. Cada caixa azul representa um mapa de atributos multicanal. As caixas brancas indicam mapas de recursos copiados. As setas representam as diferentes operações realizadas. (Fonte: [101]).

O Codificador é responsável por capturar o contexto da imagem e segue uma arquitetura típica de rede neural convolucional. Esta etapa envolve a aplicação repetida de convoluções 3×3 , seguidas da função ReLU e de uma operação de *max pooling*, realizando uma redução de dimensionalidade (*Downsampling*) da imagem. Já o Decodificador tem como objetivo calcular a localização precisa dos objetos, utilizando convoluções transpostas [103]. Neste estágio, ocorre a reconstrução dimensional (*Upsampling*) dos mapas de atributos, seguido por uma convolução, que reduz pela metade o número de canais de atributos. Além disso, há uma concatenação com os respectivos mapas de atributos gerados pelo Codificador. Esse processo de concatenação é necessário devido à perda de informações nas bordas durante as operações de convolução. Na camada final, uma convolução

1×1 é aplicada para mapear cada vetor de atributos de 64 componentes para o número desejado de classes [103, 104]. No total, a rede possui 23 camadas convolucionais, e, com esses dois blocos principais, a arquitetura da *U-Net* pode ser classificada como uma rede totalmente convolucional (FCN), por não conter camadas densas [101].

A rede *DeepLabv3*² [42] é especializada em segmentação semântica e tem sido amplamente adotada em trabalhos com imagens histológicas e apresenta aprimoramentos em relação às versões anteriores, *DeepLabv1* e *DeepLabv2* [105].

De maneira semelhante à *U-Net*, a *DeepLabv3* [106] e *DeepLabv3+* [107] (apresentada na Figura 2.7) são compostas por duas etapas principais: codificação e decodificação. Na fase de codificação, o objetivo é extrair informações essenciais da imagem utilizando uma rede neural convolucional pré-treinada. Na fase de decodificação, as informações extraídas na codificação são utilizadas para reconstruir a saída com as dimensões apropriadas. Para isso, a rede emprega a técnica de Poolização de Pirâmide Espacial (*Spatial Pyramid Pooling*) [108], um tipo de operação de *pooling* que melhora o desempenho da rede em objetos deformados. Essa técnica resolve uma limitação das redes que utilizam *pooling* tradicional, onde a entrada de imagens de tamanho fixo reduz a precisão no reconhecimento de objetos ou sub-imagens de tamanho ou escala arbitrária [42].

Existem várias arquiteturas de redes para segmentação semântica, entre as mais populares estão: *DeepLabv3* [105], *DeepLabv3+* [106], *U-Net* [103], *U-Net++* [104], *LinkNet* [109], *PSPNet* [110], *FPN* [111], *MAnet* [112] e *PAN* [113]. Cada uma dessas arquiteturas possui diferentes estratégias de aprendizado e predição, sendo mais adequadas a diferentes problemas de segmentação, dependendo da complexidade das imagens e dos recursos computacionais disponíveis.

2.5 Redes de Fusão de Características Profundas

As redes de fusão de características profundas (*Deep Feature Fusion Networks* - DFFNs) representam uma classe de redes neurais projetadas para combinar informações provenientes de diferentes fontes ou camadas de características [114, 115]. Esse processo visa aprimorar o desempenho em diversas tarefas, como classificação de imagens, detecção de objetos, segmentação semântica, entre outras [116, 25, 117].

Matematicamente, a operação de fusão de características em uma DFFN pode ser descrita como segue:

Dado um conjunto de características $F = \{F_1, F_2, \dots, F_n\}$, onde cada F_i representa as características extraídas de uma camada ou fonte específica, a operação de fusão é definida pela Equação 2.4:

²<https://github.com/tensorflow/models/tree/master/research/deeplab>

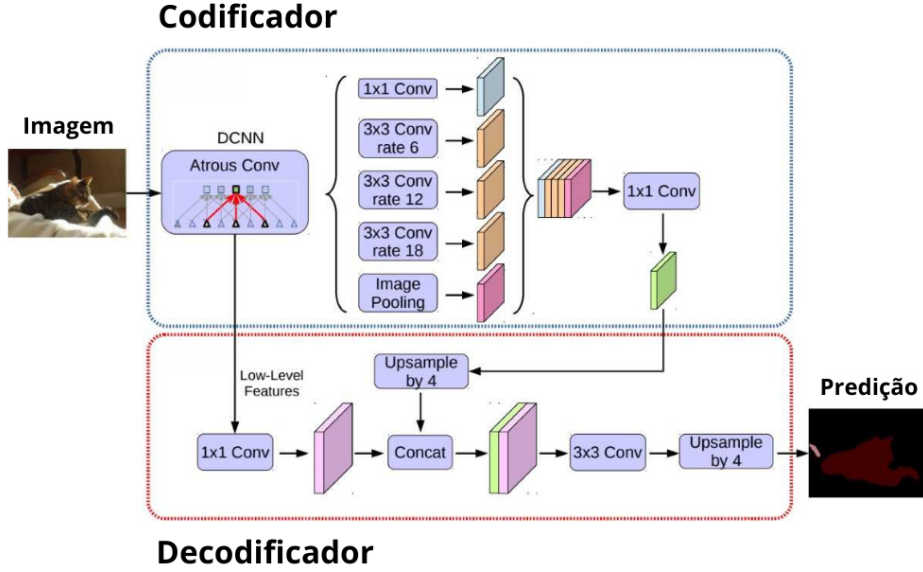


Figura 2.7: Arquitetura padrão da *DeepLab v3+*. O módulo codificador captura informações contextuais em múltiplas escalas por meio da convolução atrous em diferentes níveis, enquanto o decodificador eficiente refina a segmentação ao longo dos contornos dos objetos (Fonte: [107]).

$$F_{\text{fusion}} = \Phi(F_1, F_2, \dots, F_n), \quad (2.4)$$

F_{fusion} denota as características resultantes do processo de fusão, e Φ representa uma função que combina as características segundo uma estratégia específica. Exemplos comuns de estratégias de fusão incluem: concatenação, soma ponderada, concatenação seguida de convolução, entre outras [118].

2.6 Métricas de Avaliação

Para avaliar corretamente os modelos em problemas de classificação, as métricas mais comumente utilizadas são: acurácia, precisão, revocação e *F1-score* [119]. Já em problemas de segmentação de objetos, é comum o uso da Interseção sobre União (*Intersection over Union* - IoU) (ou *Jaccard index*) [120].

Considerando os resultados de classificação de um determinado modelo, estes podem ser apresentados por meio de uma matriz de confusão (veja a Figura 2.8) [119]:

- TP (*True Positives*): refere-se às amostras positivas classificadas corretamente como positivas;
- TN (*True Negatives*): são as amostras negativas classificadas corretamente como negativas;
- FP (*False Positives*): são as amostras negativas classificadas erroneamente como positivas;
- FN (*False Negatives*): são as amostras positivas classificadas erroneamente como negativas.

| | | Predição | |
|------|---|----------|----|
| | | P | N |
| Real | p | TP | FN |
| | n | FP | TN |

Figura 2.8: Exemplo de matriz de confusão para um problema de classificação binária (Fonte: [121]).

James *et al.* [122] definem as métricas que são amplamente utilizadas em problemas de classificação, sobretudo em contextos médicos. A acurácia (Equação 2.5) é definida como o percentual geral de acertos em relação ao total de amostras classificadas, sendo dada pela razão entre o número de amostras corretas e o total de amostras [122]. A precisão (Equação 2.6) é uma métrica que indica a qualidade das previsões positivas, calculando a taxa de acerto em relação aos dados positivos. Essa métrica é útil para avaliar a quantidade de falsos positivos.

A revocação (*recall*), também conhecida como sensibilidade (Equação 2.7), é a taxa de verdadeiros positivos, ou seja, a proporção dos dados que realmente deveriam ser classificados como positivos e que foram corretamente identificados pelo modelo. A especificidade (Equação 2.8), é usada para medir a capacidade do modelo de identificar corretamente os exemplos negativos. Por fim, o $F1$ -score (Equação 2.9) é uma métrica que combina a precisão e a revocação, sendo particularmente útil para avaliar o desempenho de modelos em conjuntos de dados desbalanceados [24, 119].

$$Acurácia = \frac{(TN + TP)}{(TN + FP + TP + FN)} \quad (2.5)$$

$$Precisão = \frac{TP}{(FP + TP)} \quad (2.6)$$

$$Revocação = \frac{TP}{(TP + FN)} \quad (2.7)$$

$$Especificidade = \frac{TN}{(TN + FP)} \quad (2.8)$$

$$F1-score = 2 \times \frac{(Precisão \times Revocação)}{(Precisão + Revocação)} \quad (2.9)$$

Por fim, entre as métricas utilizadas para problemas de classificação de imagens, destaca-se a curva ROC (*Receiver Operating Characteristic*) [123], que é uma representação gráfica que expressa a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos de um modelo de classificação, considerando diferentes valores de limiar de decisão. A curva resultante permite visualizar a capacidade do modelo de distinguir entre as classes positivas e negativas. Modelos ideais aproximam-se do canto superior esquerdo do gráfico, indicando alta taxa de verdadeiros positivos e baixa taxa de falsos positivos. A AUC (Área sob a Curva) corresponde ao valor numérico da área sob a curva ROC. Este valor varia de 0 a 1 e resume, em um único número, o desempenho do classificador:

- **AUC = 1:** Classificação perfeita, separação total entre as classes.
- **AUC = 0,5:** Classificação aleatória, equivalente a uma escolha sem critério.
- **AUC < 0,5:** Indica desempenho inferior ao aleatório, sugerindo que as predições podem estar invertendo as classes.

Assim, quanto maior o valor da AUC, melhor o modelo é em distinguir corretamente entre as classes.

Para problemas de segmentação, a métrica mais comum é a Interseção sobre União (*Intersection over Union - IoU*, também conhecido como *Jaccard Index*) [124].

Seja P a máscara predita por um modelo de detecção ou segmentação de objetos, e G a máscara verdadeira do objeto, o IoU mede a sobreposição entre as áreas das duas máscaras. O objetivo é avaliar o quanto a máscara predita se sobrepõe à máscara verdadeira. O valor do IoU varia entre 0 e 1, sendo que valores mais próximos de 1 indicam melhor desempenho. A Figura 2.9 ilustra o cálculo do IoU , que é dado pela razão entre a área de sobreposição e a área da união das máscaras [125].

Por fim, entre estudos da patologia computacional e medicina em geral, recorrentemente, é necessário mensurar a concordância entre observadores [119]. Para tal, utiliza-se

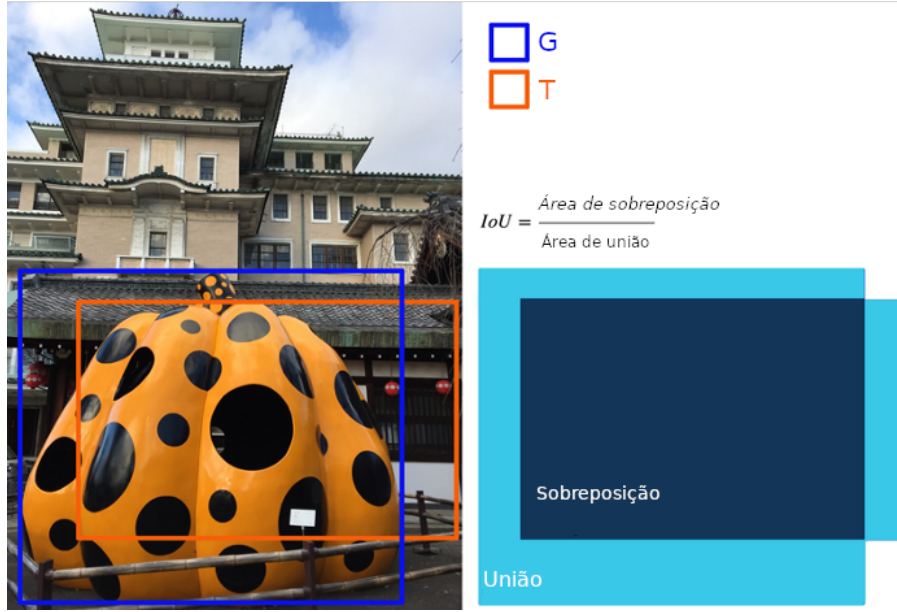


Figura 2.9: Exemplo de mensuração do IoU . (Fonte: [126]).

o *Cohen's Kappa* (κ), que é uma métrica que considera a concordância observada e a concordância esperada pelo acaso. A fórmula do *Cohen's Kappa* é dada por [127]:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

Onde, P_o é a proporção de concordância observada e P_e é a proporção de concordância esperada pelo acaso. Este índice é útil em situações binárias ou multicategoriais, sendo aplicado quando há apenas dois avaliadores.

De forma complementar, a métrica *Fleiss' Kappa* (κ) é uma extensão do Kappa de Cohen, sendo utilizado para medir a concordância entre mais de dois avaliadores. Esta métrica generaliza o cálculo da concordância observada e esperada, levando em consideração múltiplos avaliadores e múltiplas categorias.

Neste capítulo, foram apresentadas algumas definições sobre imagens histológicas, incluindo como esse tipo de imagem médica é obtido, informações gerais sobre a estrutura de um glomérulo e a célula interna do glomérulo, que é o foco deste trabalho: os podócitos. A observação dos podócitos em imagens histológicas de glomérulos revela o quão desafiador é diferenciá-los das demais células intra-glomerulares. Além disso, foram apresentadas definições sobre as técnicas computacionais comumente utilizadas para a análise automática de imagens histológicas, especialmente no contexto de aprendizagem profunda.

No próximo capítulo, serão discutidos os trabalhos relacionados. Inicialmente, será fornecido um panorama geral de estudos envolvendo imagens histológicas de glomérulos

renais, em seguida, serão apresentados trabalhos que abordam imagens histológicas de glomérulos renais com ênfase nos podócitos.

Capítulo 3

Trabalhos Relacionados

Neste capítulo, são apresentados os trabalhos relacionados ao problema da análise automatizada de glomérulos renais por meio de algoritmos de aprendizagem profunda. A revisão da literatura está dividida em duas seções. A primeira oferece um panorama geral sobre estudos que utilizam imagens histológicas de glomérulos renais para diversas finalidades. A segunda, apresenta um levantamento bibliográfico focado na análise de trabalhos que abordam a tarefa de analisar imagens de glomérulos renais, com ênfase na análise dos podócitos.

Os trabalhos encontrados durante o processo de revisão de literatura descrito a seguir, em sua maioria, se encaixaram em dois dos grupos de trabalhos citados em Barisoni *et al.* [9], que são: patologia computacional e análise computacional de imagens.

As fontes de pesquisa utilizadas na revisão de literatura foram: o portal de periódicos da CAPES e o Google Acadêmico. Além disso, foram realizadas buscas por trabalhos em diretórios específicos, com vasto número de trabalhos associados a área de patologia computacional, como: *Springer*, *PLoS*, *EMBS IEEE Library*, *IEEE Library*, *Biblioteca Digital SPIE* e *Elsevier*. O critério de seleção dos artigos considerou o ano de publicação (últimos sete anos) e principalmente, a similaridade do problema de pesquisa enfrentado com o problema estudado neste trabalho.

Para realizar as buscas, utilizou-se uma abordagem baseada em expressões regulares (*regular expressions*, regex) para garantir a recuperação eficiente dos artigos mais relevantes. Foram aplicadas buscas insensíveis a maiúsculas e minúsculas ((?i)) e delimitadas por bordas de palavras () para evitar correspondências parciais indesejadas.

Os termos foram combinados utilizando operações lógicas para refinar os resultados:

- **E (AND)**: retorna artigos que contêm todos os termos especificados.
- **OU (OR)**: retorna artigos que contêm pelo menos um dos termos.

- **XOR**: retorna artigos que contêm apenas um dos termos, mas não ambos simultaneamente.

A seguir, apresenta-se a estrutura hierárquica dos termos e expressões regulares utilizados na pesquisa por trabalhos correlatos:

1. Inteligência artificial aplicada à patologia

- *Artificial intelligence* (?i)(artificial intelligence|AI)_
- *Computational pathology* (?i)(computational pathology)_
- *Deep learning* (?i)(deep learning)_
- *Convolutional neural networks* (?i)(convolutional neural networks|CNN)_

2. Estruturas renais

- *Glomeruli* (?i)(glomeruli|glomerulus)_
- *Podocytopathy* (?i)(podocytopathy)_

3. Lesões e segmentação de podócitos

- *Podocyte detection* (?i)(podocyte detection)_
- *Podocyte segmentation* (?i)(podocyte segmentation)_
- *Podocyte lesions* (?i)(podocyte lesions?)_
- *Podocyte injury* (?i)(podocyte injury)_

4. Técnicas de processamento de imagens

- *Semantic segmentation* (?i)(semantic segmentation)_
- *Glomeruli classification* (?i)(glomeruli classification)_

3.1 Trabalhos com foco em glomérulos renais

Entre os estudos, foram identificadas propostas dedicadas à análise de lâminas inteiras de tecido (*Whole Slide Images* - *WSI*), com *pipelines* completos que englobam a detecção, segmentação e classificação de glomérulos [128, 129, 130, 131, 132, 133, 134].

Paralelamente, também foram encontrados estudos focados na análise de imagens específicas de glomérulos ([135, 136, 19, 18, 137]. Essas abordagens investigam uma ampla gama de lesões e biomarcadores associados a doenças renais, frequentemente buscando classificar os glomérulos de acordo com lesões específicas, como glomeruloesclerose [137], esclerose ([135]), hiper celularidade [19, 18] e fibrose [136].

No estudo de Uchino *et al.* [135], são utilizadas 15.888 imagens de glomérulos renais, distribuídas entre sete classes de lesões histológicas: esclerose global, esclerose segmentar, proliferação endocapilar, acúmulo de matriz mesangial, proliferação de células mesangiais, crescentes e alterações estruturais na membrana basal. O trabalho utilizou 14 modelos de redes neurais profundas, treinados a partir da *Inception v3*, para cada tipo de lesão. Participaram 25 patologistas que classificaram as amostras com e sem o auxílio dos modelos treinados. O objetivo do estudo é avaliar se os modelos de inteligência artificial poderiam melhorar o desempenho dos especialistas. Os resultados mostram que, com o uso dos modelos, há uma melhoria nas métricas de sensibilidade em 10 dos 14 modelos e de especificidade em 8 dos 14.

Já em Kannan *et al.* [128] é proposto um modelo para processar lâminas inteiras de tecido renal (WSI) coradas em tricrômico, segmentar os glomérulos e classificá-los em quatro categorias: (i) não glomérulo, (ii) normal, (iii) parcialmente esclerosado e (iv) globalmente esclerosado (GS). O modelo de CNN alcançou acurácia de $92,67\% \pm 2,02\%$ e coeficiente *Kappa* de $0,8681 \pm 0,0392$ na discriminação entre imagens glomerulares e não glomerulares. Para a segmentação, o classificador multilabel baseado na *Inception v3* obteve resultados precisos na identificação de glomérulos GS nos dados de teste, com coeficiente de correlação de *Matthews* (MCC) de 0,628.

Em Gallego *et al.* [129], os autores também realizam a tarefa de segmentação de glomérulos em lâminas WSI, classificando-os como normais ou esclerosados. A segmentação e a classificação dos glomérulos são realizadas por um modelo baseado na arquitetura *U-Net*. Posteriormente, as classificações glomerulares são refinadas com base na histomorfometria glomerular (análise morfológica com critérios pré-definidos). O conjunto de dados utilizado é composto por 51 lâminas de tecido, coradas em PAS (37 lâminas) e HE (14 lâminas). Nos WSIs corados com PAS, os glomérulos normais e esclerosados foram classificados, respectivamente, com *F1 score* de 0,97 e 0,68. Nos WSIs corados com HE, as pontuações *F1* foram de 90,8% e 78,1%. Considerando ambos os corantes, o desempenho de classificação foi de *F1* de 94,5% e 76,8%. Um dos diferenciais deste trabalho é que os pesquisadores realizaram o treinamento com imagens de um corante e testaram o desempenho em imagens coradas com outro corante.

No trabalho de Mathur *et al.* [136], o objetivo é analisar a presença de fibrose em lâminas de tecido renal. O estudo está dividido em duas tarefas: (i) classificar os glomérulos como normais ou anormais e (ii) classificar regiões de tecido sem glomérulos entre três classes de fibrose: suave, moderada ou severa. Os conjuntos de dados são compostos por *patches* de imagens extraídas das lâminas de tecido, totalizando 935 imagens de glomérulos normais ou anormais (619 anormais e 316 normais) no conjunto de dados 1, e 923 imagens de regiões sem glomérulos no conjunto de dados 2. O trabalho avalia 3

abordagens para classificar as imagens: (a) transferência de aprendizado em CNNs fim a fim, (b) extração de atributos com CNN e classificação com algoritmo supervisionado, e (c) um novo modelo proposto, a *Multi-Gaze Attention Network* (MGANet), que utiliza o mecanismo de *self-attention* com diferentes saídas e conexões residuais paralelas em uma arquitetura CNN. O melhor desempenho foi obtido com o modelo proposto, alcançando 87,25% de precisão na classificação de glomérulos e 81,47% na classificação de fibrose.

Em mais um trabalho de segmentação de glomérulos, Ginley *et al.* [130] analisam imagens de glomérulos em relação à nefropatia diabética. O conjunto de dados utilizado consiste em 25 lâminas de tecido renal de camundongo e 54 lâminas de tecido renal humano. Os métodos empregados no estudo incluem a rede *DeepLab v2* (*ResNet* no *backbone*) para a detecção dos glomérulos, além de algoritmos tradicionais de processamento de imagens para a identificação das bordas dos glomérulos. Adicionalmente, os pesquisadores utilizam uma rede neural recorrente (RNN) para quantificar a estrutura intra-glomerular em três componentes: núcleos, lumina capilar e espaços de *Bowman*. Os resultados obtidos são de 93% de acurácia na detecção dos glomérulos, 94% de sensibilidade [122] e 94% de especificidade [122] para os núcleos, e 95% de sensibilidade e 99% de especificidade para as estruturas internas dos glomérulos.

Entre os trabalhos de classificação de imagens de glomérulos sem a realização prévia de segmentação ou detecção em lâminas inteiras WSI, Barros *et al.* [18] destaca-se como um dos últimos a utilizar métodos tradicionais de aprendizado de máquina. Na metodologia adotada, as imagens são pré-processadas com técnicas clássicas de processamento de imagens, como conversão de espaços de cor, limiarização por Otsu e operações morfológicas [10]. A extração de atributos é realizada com engenharia de atributos e a classificação é realizada com o algoritmo *kNN* [10]. O resultado obtido foi uma acurácia de 88,3% na classificação. O conjunto de dados utilizado conta com 811 imagens coradas em PAS e HE.

No trabalho de Chagas *et al.* [19], os autores realizam a mesma tarefa que Barros *et al.* [18], utilizando o mesmo conjunto de dados, mas empregaram redes neurais convolucionais como extratores de atributos e o algoritmo SVM para a classificação. Além da classificação binária, os autores realizam a classificação das sublesões de hiperplasia: mesangial, endocapilares e ambas. Os resultados obtidos são de 82% de acurácia média na classificação dos tipos de lesão e 100% de acurácia na classificação binária. Tanto na tarefa binária quanto na multiclassificação, o método proposto superou o desempenho das redes *Xception*, *ResNet50* e *InceptionV3*.

No trabalho de Shubham *et al.* [131], é realizada a segmentação de glomérulos utilizando o conjunto de dados público HuBMAP¹. As imagens do conjunto de dados são

¹<https://hubmapconsortium.org/>

lâminas WSI, coradas em PAS. O método utilizado para segmentação é a rede *U-Net*, e o resultado obtido foi de 99,68% de acurácia e 0,90 de coeficiente Dice.

Em Jiang *et al.* [133], é utilizada a rede *Mask R-CNN* para segmentar e classificar glomérulos em lâminas (WSI) de tecido. O modelo é treinado com *snapshots*, que são porções menores das lâminas contendo um ou mais glomérulos. Os glomérulos no conjunto de dados são rotulados em 3 classes: normal, com esclerose ou com outras lesões. O conjunto de treino é formado por 296 lâminas (1.018 *snapshots*), e o conjunto de teste, por 52 lâminas (105 *snapshots*). Os resultados obtidos na classificação dos glomérulos alcançam os seguintes *F1 scores*: 0,914 para glomérulos normais, 0,896 para glomérulos com esclerose, 0,681 para esclerose global e 0,756 para outras lesões.

Em uma abordagem ligeiramente diferente, Yang *et al.* [134] realizam também a segmentação e classificação de glomérulos. A rede *Mask R-CNN* [138] (com a *ResNet 101* [84] pré-treinada com *ImageNet* [139] como *backbone*) é utilizada para segmentar os glomérulos, e uma abordagem combinando uma CNN (DenseNet) com uma LSTM foi empregada para classificar as doenças entre 5 diferentes classes de lesões. O conjunto de dados utilizado é composto por 1379 lâminas (WSI) de tecido, coradas em PAS, PAM, HE e Tricrômico. Os resultados obtidos foram um *F1 score* de 0,94 para detecção dos glomérulos e uma acurácia de até 0,94 para a classificação de diferentes lesões em imagens coradas em HE.

No trabalho de Ginley *et al.* [137], é realizada a segmentação de glomeruloesclerose e fibrose intersticial com atrofia tubular em lâminas inteiras de tecido. A rede utilizada é a *DeepLab v2* [105]. Adicionalmente, os autores correlacionam os resultados da rede com os diagnósticos de quatro patologistas, que utilizaram ferramentas estatísticas clássicas. O conjunto de dados é composto por 79 lâminas para treinamento e 20 lâminas para teste.

Além dos estudos já citados, que aplicam técnicas de aprendizado profundo em lâminas de tecido, a literatura também apresenta propostas focadas no problema de obtenção de dados anotados e devidamente validados por especialistas humanos. Exemplos disso incluem os trabalhos de DeHaan *et al.* [132], Aron *et al.* [140], Minamiguchi *et al.* [141] e Lutnick *et al.* [20].

No trabalho de Aron *et al.* [140], os autores definem nove classes de padrões morfológicos glomerulares e treinam doze modelos de redes neurais convolucionais (CNNs). O treinamento é realizado em duas etapas: o primeiro conjunto de dados foi definido por um nefropatologista especialista (12253 imagens) e o segundo, por um consenso de três especialistas na área (11142 imagens). A validação é feita utilizando um terceiro conjunto de dados composto por 180 imagens, comparando os resultados com o consenso entre os patologistas (valores *kappa* entre 0,838 e 0,938). Adicionalmente, os autores destacam as áreas da imagem decisivas para a tomada de decisão baseada na CNN, utilizando mapas

de ativação para cada classe. Uma das vantagens do método proposto é a capacidade de identificar a presença de diferentes padrões de doença glomerular que podem coexistir em um único glomérulo (por exemplo, necrose combinada com hiper celularidade mesangial e endocapilar), contrastando com abordagens anteriores que, geralmente, reconhecem apenas um padrão específico. Os autores concluem sugerindo que os resultados fornecem estímulos para projetos em andamento que integram outros tipos de dados, como imunohistoquímica, microscopia eletrônica e informações clínicas, visando o desenvolvimento de ferramentas aplicáveis à nefropatologia diagnóstica de rotina.

DeHaan *et al.* [132] propõem um modelo de rede adversária generativa (GAN), baseado no modelo *CycleGAN* [142], para realizar transferência de estilo (*style transfer*) de lâminas de tecido renal (contendo glomérulos renais) coradas em HE para PAS e Tricrômico. Os resultados são avaliados por quatro patologistas, que consideraram as imagens geradas como semelhantes às reais. A vantagem desse trabalho reside no uso de *data augmentation*, que pode ampliar o conjunto de dados de um *dataset*, além de operações tradicionais como rotação, inversão e outras.

No trabalho de Kumar *et al.* [143], é realizada uma investigação sobre modelos de aprendizado profundo para a detecção e classificação de câncer ovariano e doenças glomerulares renais utilizando imagens histopatológicas. Os autores analisam o impacto de diferentes métodos de otimização em um conjunto de redes convolucionais de referência. Ao final, os modelos propostos demonstraram alta precisão na detecção e classificação dessas condições, alcançando acurácias superiores a 99%.

Preocupados com o gargalo no processo de anotação de imagens histológicas, em Lutnick *et al.* [144], os autores propõem um sistema de anotação de imagens que inclui um pré-anotador, que facilita o trabalho dos patologistas. O sistema segmenta automaticamente os glomérulos e entrega as anotações para refinamento pelos patologistas. O pré-anotador é baseado em um modelo semi-supervisionado denominado *DatasetGAN* [145]. Os resultados mostram que o tempo de anotação com o uso desse sistema torna-se significativamente menor em comparação com o processo de anotação tradicional.

Em Minamiguchi *et al.* [141], é apresentado um método não supervisionado para correlacionar imagens de tecido com nefropatia, coradas em HE. No trabalho de Lutnick *et al.* [20], os autores propõem uma ferramenta de anotação de imagens executada na nuvem, com alta escalabilidade e colaboração entre patologistas.

Por fim, no trabalho de Ginley *et al.* [146], é realizada a segmentação de diversas estruturas do tecido renal, incluindo os glomérulos. Os métodos utilizados são redes de segmentação panópticas, uma classe mais recente de algoritmos de segmentação de imagem, que distinguem objetos do tipo instância (contáveis, como túbulos renais) de objetos do tipo grupo (incontáveis, como o interstício renal). O conjunto de dados utilizado

consiste em 126 imagens de lâminas coradas em PAS, de nefropatia diabética nativa, nefrite lúpica e biópsias renais de vigilância de transplante, com anotações contemplando interstício, glomérulos, glomérulos escleróticos globalmente, túbulos e árvore arterial (artérias/arteríolas). Os resultados mostram que as redes panópticas apresentaram alto desempenho, comparáveis ao dos patologistas renais, para todas as classes anotadas em um conjunto de teste de biópsias renais de transplante.

3.2 Trabalhos com foco em podócitos

Entre os trabalhos de imagens histológicas com foco específico em podócitos, foi encontrado um conjunto pequeno de propostas, entre os quais é possível destacar: Zeng *et al.* [33], Maraszek *et al.* [30], Govind *et al.* [28], Zimmerman *et al.* [31], Santo *et al.* [34] e Farhat *et al.* [147].

O trabalho de Zeng *et al.* [33] tem como objetivo localizar glomérulos, identificar lesões glomerulares (esclerose glomerular global e segmentar, crescente e nenhuma das anteriores), e identificar e quantificar diferentes células glomerulares intrínsecas. As imagens utilizadas estão coradas em PAS. Para a tarefa de localização dos glomérulos, são usadas 360 lâminas para treinamento e 40 para teste. O método utilizado é a rede *Attention U-Net* [148]. Na tarefa de classificação dos glomérulos, são utilizadas 1438 imagens de glomérulos, com aproximadamente 300 imagens por classe. O método adotado para a classificação é a combinação das redes *DenseNet-121* [85] e *LSTM-SENet* [149]. Para a segmentação das células internas (incluindo os podócitos), são utilizadas 460 imagens de glomérulos, com aproximadamente 70 mil células anotadas. A rede utilizada é uma modificação da *U-net*, batizada de *2D V-Net*. Os resultados obtidos são: 93,1% de precisão média e 94,9% de *recall* médio para localização de glomérulos; 95% de precisão para a classificação das lesões glomerulares; e 88,2% de precisão média e 87,9% de *recall* médio para detecção das células internas dos glomérulos.

Em Govind *et al.* [28], a tarefa é a detecção e quantificação de podócitos para reconhecimento de tumor de *Wilms*'. O conjunto de dados utilizado é composto por 240 *patches* (imagens de glomérulos) coradas em PAS, originadas de camundongos. O método proposto segue os seguintes passos: inicialmente, as imagens do conjunto de dados são adquiridas em imunofluorescência, e depois os mesmos glomérulos são corados em PAS. As imagens desses dois conjuntos são utilizadas para o treinamento de uma GAN [150], que converte imagens PAS para imunofluorescência. Após o treinamento, o método recebe uma imagem em PAS, converte-a para sua versão artificial em imunofluorescência, e então a imagem é processada e os podócitos são segmentados através da máscara criada.

O resultado obtido é de 0,87 de sensibilidade e 0,93 de especificidade para a detecção dos podócitos.

Em Maraszek *et al.* [30], o objetivo também é detectar e quantificar podócitos renais, porém com o objetivo de associá-los à presença de diabetes *Mellitus*. Assim como no trabalho de [28], o método proposto utiliza imagens de imunofluorescência e as mesmas imagens em PAS. A imagem em imunofluorescência é processada com métodos clássicos de processamento digital de imagens e finalizada como uma máscara para segmentação na imagem correspondente em PAS. O conjunto de dados é formado por 883 imagens de glomérulos de ratos. No final do trabalho, os autores calculam os danos nos glomérulos através de análise morfológica dos podócitos e sua distribuição intra-glomerular. Os resultados obtidos são de sensibilidade de 0,727, especificidade de 0,999 e acurácia de 0,959 na localização dos podócitos.

No trabalho de Santo *et al.* [34], os autores propõem o *PodoCount*, uma ferramenta computacional para quantificação automatizada de podócitos em tecidos marcados de forma imuno-histoquímica. O conjunto de dados utilizado conta com 35 lâminas de tecido com diabetes nefrótica. Nos experimentos relatados, a partir de glomérulos segmentados de seções de tecido (WSI), os núcleos podocitários são segmentados, sendo a partir deste momento aplicada uma análise de imagens para calcular medidas de depleção e morfometria nuclear. Os resultados obtidos indicam uma segmentação dos núcleos podocitários com sensibilidade de 0,85 e especificidade de 0,99.

Na proposta de Zimmerman *et al.* [31], os autores utilizam um conjunto de dados bastante robusto, com 1095 imagens de imunofluorescência, contendo um total de 27.696 podócitos anotados. O objetivo do trabalho também é detectar podócitos, mas com o intuito de associá-los à doença glomerulonefrite associada a anticorpos anti-neutrófilos citoplasmáticos (ANCA-GN). A rede utilizada para segmentar os glomérulos e os podócitos é a *U-net*. O coeficiente Dice obtido nas tarefas de segmentação tanto dos glomérulos quanto dos podócitos é de 0,92.

Finalmente, em Farhat *et al.* [147], propõe-se o uso de redes neurais convolucionais para automatizar a segmentação de células e capilares glomerulares em biópsias de transplante renal, avaliando a correlação dessas estruturas com a função do enxerto. O estudo inclui 215 pacientes, divididos em três grupos: o grupo de treinamento, composto por 37 pacientes cujas células e capilares foram anotados manualmente para treinar as redes; o grupo de teste, com 24 pacientes, utilizado para comparar as anotações manuais às previsões automatizadas; e o grupo de aplicação, com 154 biópsias, utilizado para examinar fatores preditivos em relação à função renal e ao prognóstico. No resultado geral, considerando a segmentação de todas as estruturas, a rede neural alcança métricas de precisão, revocação, *F1-score* e *IoU* superiores a 0,92, 0,85, 0,89 e 0,74, respectivamente.

O resultado específico para a segmentação de podócitos apresenta 0,75 de IoU, obtida em um teste com 633 podócitos anotados por especialistas humanos.

A Tabela 3.1 resume os trabalhos com imagens histológicas que se concentram especificamente na análise automatizada de podócitos em glomérulos renais. Esses estudos são os principais trabalhos correlatos quando se considera o escopo do presente estudo.

Tabela 3.1: Resumo dos trabalhos correlatos mais relevantes.

| Autores | Tarefa | Conjunto de Dados | Métodos | Resultados |
|------------------------------|---|--|--|--|
| Zeng <i>et al.</i> [33] | Detecção de glomérulos em lâminas WSI, classificação dos glomérulos e segmentação de células intra-glomerulares | 60 imagens de glomérulos, contendo aproximadamente 70 mil células anotadas | 2D V-Net e U-Net | Precisão média de 88,2% e recall médio de 87,9% na detecção das células internas dos glomérulos |
| Maraszek <i>et al.</i> [30] | Detecção e quantificação de podócitos renais para associá-los à presença de diabetes mellitus | 883 imagens de glomérulos de rato em imunofluorescência | Métodos tradicionais de processamento de imagens | Sensibilidade de 72,7%, especificidade de 99,9% e acurácia de 95,9% na localização dos podócitos |
| Govind <i>et al.</i> [28] | Detecção e quantificação de podócitos para reconhecimento de tumor de Wilm's | 40 imagens de glomérulos coradas em PAS, originadas de camundongo | Conversão de estilo com GAN e operações tradicionais de processamento de imagens | Sensibilidade de 87% e especificidade de 93% para detecção dos podócitos |
| Zimmerman <i>et al.</i> [31] | Detecção de podócitos para associá-los à doença ANCA-GN | 1095 imagens de imunofluorescência (27.696 podócitos anotados) | U-Net | Coefficiente Dice de 92% para segmentação de podócitos |
| Santo <i>et al.</i> [34] | Segmentação, quantificação de podócitos e análise de depleção | 35 lâminas coradas em PAS, com imagens de tecido de rato | Métodos tradicionais de processamento de imagens | Sensibilidade de 85% e especificidade de 99% |
| Farhat <i>et al.</i> [147] | Segmentação de células intra-glomerulares e correlações para transplante | 1170 podócitos anotados | Mask R-CNN com Inception ResNet V2 | IoU de 75% para segmentação de podócitos |

3.3 Lacunas e desafios identificados

A análise dos trabalhos correlatos evidencia que nenhum estudo se concentra especificamente na podocitopatia, embora muitos utilizem a segmentação de podócitos para associar perda celular ou alterações morfológicas a condições como diabetes nefrótica. Além disso, os conjuntos de dados analisados não estão disponíveis para outros pesquisadores e carecem de diversidade, como imagens de podócitos com diferentes tipos de lesões (hiperplasia, hipertrofia e degeneração), variação de coloração e desbalanceamento entre as classes.

Os estudos focados na análise de células glomerulares internas, particularmente podócitos, têm como objetivo principal a localização e quantificação para o estudo de diversas patologias. Em comum, muitas dessas pesquisas utilizam imagens de imunofluorescência com marcadores específicos para facilitar a segmentação dos podócitos. Entretanto, a aplicação desses métodos é frequentemente limitada, pois nem todos os conjuntos de dados disponíveis possuem imagens de imunofluorescência ou outros tipos de coloração, como PAS, indispensáveis para replicar essas abordagens.

As metodologias empregadas nesses trabalhos incluem redes neurais profundas e técnicas clássicas de processamento de imagens. Os conjuntos de dados utilizados são tipicamente obtidos sob condições controladas, utilizando o mesmo corante, a mesma resolução e provenientes de um único laboratório. Embora essa padronização facilite a segmentação, a dependência de imagens obtidas em ambientes tão controlados limita a generalização dos métodos para conjuntos de dados mais variados.

Por fim, enquanto a utilização de imagens imunofluorescentes com pré-marcação dos podócitos facilita a segmentação, sua aplicação prática pode ser restrita pela falta de imagens de diferentes tipos de coloração e pela indisponibilidade de dados em alguns casos clínicos, o que prejudica a replicabilidade das abordagens.

Capítulo 4

Metodologia

A metodologia adotada neste trabalho está diretamente relacionada aos objetivos específicos da tese, apresentados na Seção 1.4. Desse modo, os experimentos foram organizados em quatro etapas principais:

1. Elaboração dos conjuntos de dados.
2. Identificação de podocitopatia em imagens de glomérulos.
3. Estudo sobre o impacto do uso de classificadores automatizados na prática dos patologistas.
4. Segmentação de lesões podocitárias em imagens de glomérulos.

Na etapa de elaboração dos conjuntos de dados, inicialmente um grupo de patologistas voluntários analisou imagens histológicas de glomérulos renais, atribuindo a cada imagem um rótulo binário indicando a presença ("com lesão") ou ausência ("sem lesão") de podocitopatia. Em seguida, uma nova tarefa foi proposta, na qual os mesmos patologistas indicaram, por meio de segmentação, as regiões específicas nas imagens que continham diferentes tipos de lesões podocitárias.

A etapa de identificação de podocitopatia em imagens de glomérulos teve como objetivo desenvolver modelos de classificação binária capazes de inferir automaticamente a presença ou ausência de podocitopatia nas imagens analisadas. Nesta fase, diferentes métodos de referência da literatura foram treinados e comparados com uma nova solução proposta para classificação automatizada das imagens.

O estudo sobre o impacto da utilização de classificadores automatizados visou avaliar, por meio de experimentos controlados, as variações de desempenho dos patologistas ao analisar imagens com e sem o suporte desses modelos. O objetivo foi investigar se a utilização dessas ferramentas melhora ou acelera o diagnóstico clínico dos especialistas.

Por fim, a etapa de segmentação de lesões podocitárias compreendeu uma comparação rigorosa entre modelos já existentes na literatura e uma nova abordagem proposta neste trabalho, especificamente desenvolvida para a tarefa de segmentação das lesões em um cenário de dados limitados, com múltiplas colorações e distribuição desbalanceada das classes. A Figura 4.1 ilustra esquematicamente cada uma dessas etapas.

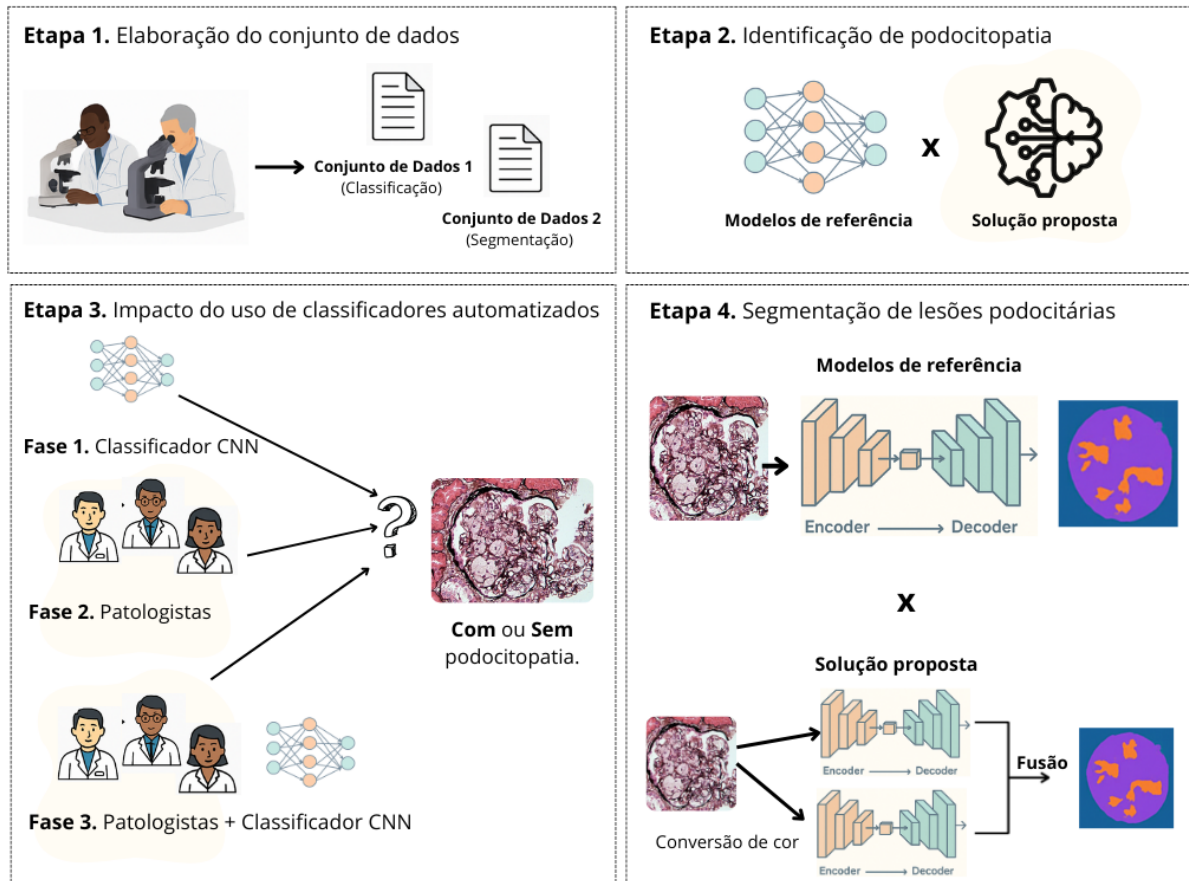


Figura 4.1: Resumo ilustrativo das fases e experimentos realizados no trabalho.

Cada uma dessas etapas metodológicas está intimamente relacionada. A elaboração cuidadosa dos conjuntos de dados, realizada na primeira etapa, é um requisito fundamental para o sucesso das etapas subsequentes, uma vez que esses dados servirão como base tanto para os experimentos de classificação quanto para a segmentação das lesões podocitárias. A identificação automática de podocitopatia, por sua vez, é um passo necessário para avaliar o impacto clínico da automação no desempenho dos patologistas, permitindo verificar como as ferramentas computacionais afetam a prática diagnóstica real. Por fim, a etapa de segmentação visa aprofundar ainda mais o detalhamento das lesões, ampliando o potencial de aplicação prática das soluções propostas ao fornecer informações mais es-

pecíficas sobre as regiões afetadas, contribuindo diretamente para a precisão diagnóstica e decisão clínica final.

Capítulo 5

Elaboração dos Conjuntos de Dados

A elaboração dos conjuntos de dados deste trabalho iniciou-se a partir de um acervo de 30 mil imagens não rotuladas de glomérulos renais, cedido pelo Centro de Pesquisas Gonçalo Muniz, Fiocruz. Essas imagens foram analisadas visualmente por dois patologistas de forma independente, sem validação cruzada entre os especialistas, resultando na seleção de 835 imagens classificadas binariamente como “com” ou “sem” lesões podocitárias. Este primeiro conjunto de dados foi denominado $D1v1$. Posteriormente, uma segunda rodada de análise permitiu a identificação de novas imagens, ampliando o conjunto para 1043 imagens, o que originou a versão $D1v2$.

A partir do conjunto expandido, foi desenvolvido o $D2$, com o objetivo de validar o desempenho diagnóstico dos patologistas na tarefa de classificação binária, incorporando, além da inspeção visual, a confirmação clínica baseada em níveis de proteinúria. Dessa forma, os conjuntos de dados $D1v1$, $D1v2$ e $D2$ foram construídos com o propósito comum de apoiar a tarefa de **classificação binária**, visando identificar a presença ou ausência de podocitopatia nas imagens de glomérulos renais.

Complementarmente, elaborou-se o conjunto de dados $D3$, destinado especificamente à tarefa de **segmentação semântica multiclasse**. Nesse conjunto, os patologistas anotaram manualmente as regiões das imagens correspondentes a diferentes tipos de lesões podocitárias (hiperplasia, hipertrofia e degeneração), bem como áreas de podócitos sem lesão.

Os conjuntos de dados $D1v1$, $D1v2$, $D2$ e $D3$ representam uma evolução sequencial de um mesmo acervo de imagens, sendo que cada nova versão aperfeiçoou o nível de anotação e contribuiu diretamente para a melhoria da análise automatizada a ser realizada nos modelos treinados nas fases posteriores deste trabalho.

Cabe destacar, como limitação metodológica, que, devido à restrição de tempo e ao número reduzido de patologistas voluntários, em nenhuma das rodadas houve reavaliação

cruzada das anotações. Essa limitação pode ter resultado em inconsistências pontuais nas imagens rotuladas, aspecto considerado nas análises subsequentes.

Adicionalmente, todos os conjuntos de dados gerados estão disponíveis para uso acadêmico e podem ser solicitados mediante o preenchimento de formulário específico, acessível na página oficial do grupo de pesquisa Pathospotter (<https://pathospotter.bahia.fiocruz.br/>). A Figura 5.1 resume todo o processo de anotação e a evolução dos conjuntos até a obtenção do conjunto *D3*.

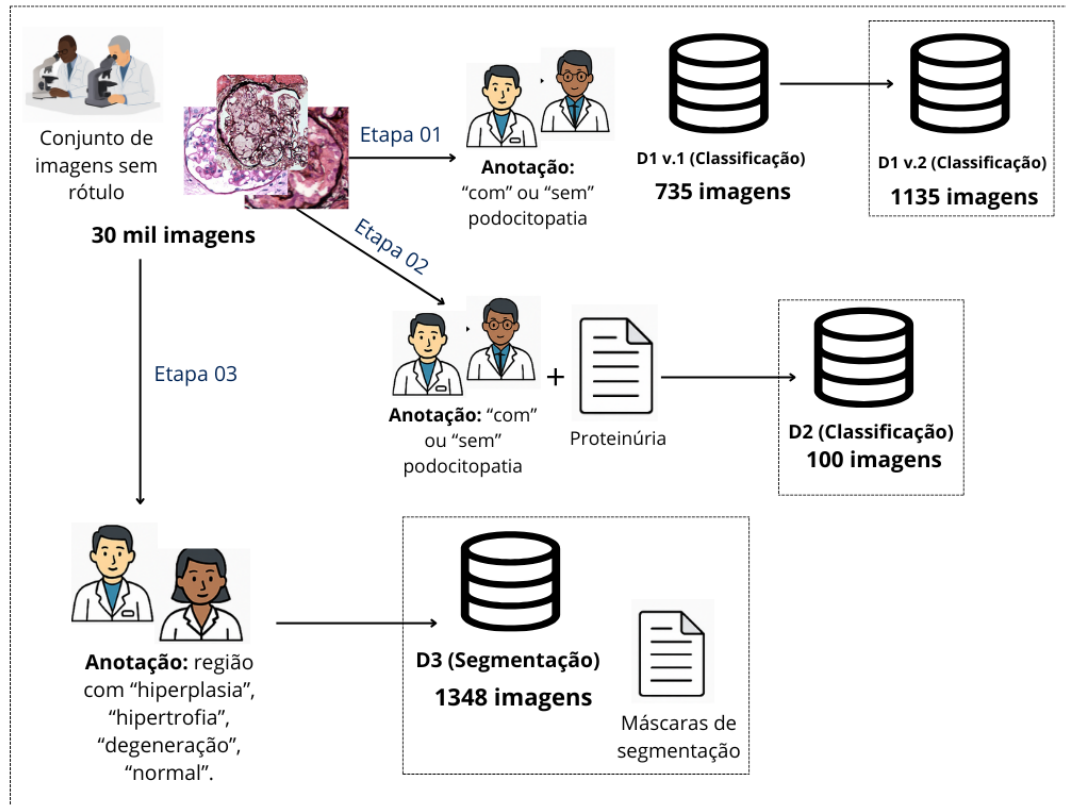


Figura 5.1: Etapas de anotação para obtenção dos conjuntos de dados.

5.0.1 Conjunto de dados para classificação: D1v.1

Este conjunto de dados é composto por 835 imagens de glomérulos renais, das quais 374 apresentam lesão podocitária e 461 não apresentam lesão. As anotações sobre a presença ou ausência de lesão podocitária foram realizadas por dois patologistas, ambos nefropatologistas ¹.

¹A aquisição das imagens foi conduzida de acordo com a Resolução nº 466/12 do Conselho Nacional de Saúde. Todos os procedimentos foram aprovados pelo Comitê de Ética em Pesquisa com Seres Humanos do Instituto Gonçalo Moniz da Fundação Oswaldo Cruz (CPqGM/FIOCRUZ), sob os Protocolos nº 188/09 e nº 1.817.574.

As imagens foram obtidas utilizando diferentes câmeras, microscópios e scanners, apresentando variados formatos (.jpg, .png e .tif) e resoluções, que variam de 238×201 a 1920×1440 pixels. Os corantes empregados nas amostras são: Tricrômico (173 imagens), *Periodic Acid-Schiff* (PAS) (409 imagens), *Periodic Acid Methenamine Silver* (PAMS) (169 imagens) e *Hematoxylin and Eosin* (HE) (84 imagens).

As imagens rotuladas como "com podocitopatia" incluem amostras com diferentes tipos de lesões podocitárias: hipertrofia, hiperplasia e degeneração. Exemplos de imagens representativas do conjunto de dados podem ser visualizados na Figura 5.2. Vale destacar que, tanto no grupo de imagens com lesão podocitária quanto no grupo sem lesão, também estão presentes outros tipos de alterações glomerulares, como hiper celularidade, esclerose e alterações membranasas.

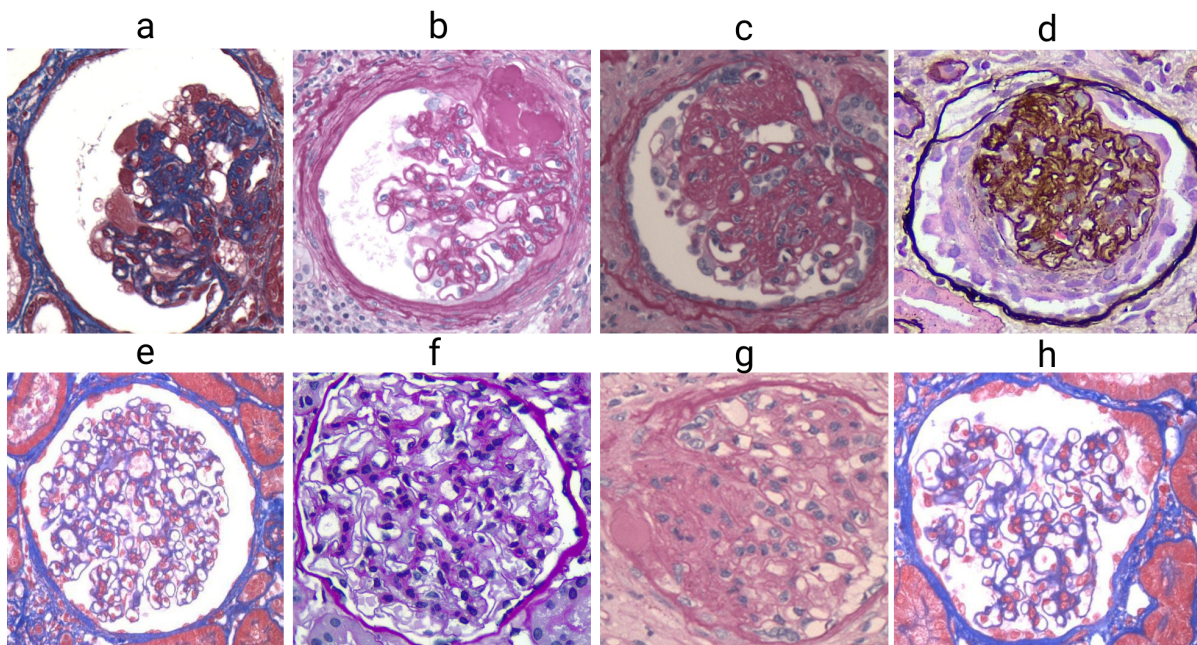


Figura 5.2: Exemplos de imagens que compõem o conjunto de dados. Com lesão (a-d) e sem lesão (e-h).

5.0.2 Conjunto de dados para classificação: D1v2 e D2

Os conjuntos de dados *D1v2* e *D2* foram utilizados especificamente na etapa de estudo sobre o impacto do uso de classificadores automatizados na prática médica. O conjunto de dados *D1v2* foi utilizado para treinar e validar o classificador automatizado, enquanto o conjunto de dados *D2* foi empregado para avaliação de desempenho tanto do melhor modelo de classificador quanto dos patologistas. Ambos os conjuntos incluem imagens de glomerúlos renais com diversos tipos de lesões podocitárias, como degeneração, hiperplasia

Tabela 5.1: Distribuição dos corantes entre as imagens do conjunto de dados *D1v2*.

| Rótulo | HE | PAS | Tricrômico | PAMS | Total |
|-----------|-------------|-------------|-------------|-------------|-------------|
| Com lesão | 44 | 275 | 104 | 90 | 513 (49%) |
| Sem lesão | 64 | 254 | 94 | 118 | 530 (51%) |
| Total | 108 (10,8%) | 529 (50,8%) | 198 (18,9%) | 208 (19,9%) | 1043 (100%) |

Tabela 5.2: Distribuição dos corantes entre as imagens do conjunto de dados *D2*.

| Rótulo | HE | PAS | Tricrômico | PAMS | Total |
|-----------|----------|----------|------------|----------|------------|
| Com lesão | 18 | 3 | 2 | 17 | 50 (50%) |
| Sem lesão | 12 | 24 | 6 | 8 | 50 (50%) |
| Total | 30 (30%) | 37 (37%) | 8 (8%) | 25 (25%) | 100 (100%) |

e hipertrofia. Essa diversidade proporciona uma representação realista dos casos encontrados na prática clínica, permitindo que os modelos computacionais treinados reflitam a complexidade dos cenários do mundo real.

O conjunto de dados *D1v2* é composto por 1.043 imagens, das quais 513 são rotuladas como “com lesão podocitária” e 530 como “sem lesão podocitária”. As imagens apresentam quatro tipos de coloração: *Periodic Acid-Schiff* (PAS), *Hematoxylin and Eosin* (HE), *Periodic Acid Methenamine Silver* (PAMS) e Tricrômico, conforme a distribuição mostrada na Tabela 5.1. Essas imagens foram obtidas em diferentes instituições, utilizando câmeras, microscópios e scanners variados, com resoluções espaciais que variam de 238×201 a 1920×1440 *pixels*. Em ambas as classes de imagens (com e sem lesão podocitária), podem ser observadas outras alterações glomerulares associadas, como hiper celularidade, esclerose e alterações membranosas.

O conjunto de dados *D2* foi criado seguindo um protocolo distinto. Ele é composto por 100 imagens de glomérulos renais (50 com e 50 sem lesão podocitária), coradas com PAS, HE, PAMS e Tricrômico. Diferentemente do conjunto *D1v2*, que foi rotulado exclusivamente com base na inspeção visual dos patologistas, cada imagem do conjunto *D2* foi validada tanto pela análise visual quanto por dados clínicos referentes à taxa de proteinúria. Além disso, cada imagem de *D2* foi obtida de um paciente distinto (totalizando 100 pacientes), de modo a favorecer a generalização dos resultados obtidos pelos modelos computacionais. Essa abordagem também aumentou a diversidade de casos utilizados para avaliar o desempenho dos classificadores. A distribuição dos corantes no conjunto de dados *D2* está apresentada na Tabela 5.2.

5.0.3 Conjunto de dados para segmentação: D3

O conjunto de dados *D3* é o mais abrangente deste estudo, reunindo imagens provenientes dos conjuntos *D1v2*, *D2* e de novas adições, totalizando 1.401 imagens. Este conjunto foi

elaborado para suprir a carência de bases públicas que contemplem anotações específicas para diferentes tipos de lesões podocitárias e múltiplas técnicas de coloração histológica. A iniciativa resultou em uma coleção de imagens de glomérulos renais estruturada para representar o problema de segmentação de lesões podocitárias em casos variados, desafiando, assim, os modelos de redes neurais convolucionais (CNNs) com cenários mais próximos da prática clínica real.

O processo de anotação das segmentações, que se estendeu por aproximadamente 18 meses, envolveu três patologistas, cada um com mais de cinco anos de experiência em nefropatologia. A colaboração entre os especialistas foi fundamental para assegurar anotações precisas e consistentes em todo o conjunto de dados.

O conjunto final inclui 204 imagens coradas com *Hematoxylin and Eosin* (HE), 828 imagens com *Periodic Acid-Schiff* (PAS), 124 imagens com tricrômico de Masson, 39 imagens com tricrômico de Gomori e 206 imagens com *Periodic Acid Methenamine Silver* (PAMS). A Figura 5.3 ilustra exemplos representativos das amostras do conjunto de dados, juntamente com exemplos das anotações realizadas pelos patologistas (verdade de campo) em uma das amostras. A tabela 5.3 resume a percentual de imagens por corante.

Tabela 5.3: Distribuição do número de imagens por técnica de coloração no conjunto de dados $D3$.

| Técnica de coloração | Número de imagens |
|--|--------------------------|
| <i>Hematoxylin and Eosin</i> (HE) | 204 |
| <i>Periodic Acid-Schiff</i> (PAS) | 828 |
| Tricrômico de Masson | 124 |
| Tricrômico de Gomori | 39 |
| <i>Periodic Acid Methenamine Silver</i> (PAMS) | 206 |
| Total | 1401 |

Durante o processo de anotação, os patologistas utilizaram o software *Labelme* [151] para marcar as regiões das imagens contendo podócitos sem lesões, bem como aquelas apresentando lesões específicas: degeneração, hipertrofia e hiperplasia. Esse procedimento resultou no registro de quatro classes distintas nas imagens. Ao final, foram anotados um total de 54.760 objetos, sendo 48.471 podócitos sem lesões (controle normal), 2.782 instâncias de degeneração, 1.872 de hiperplasia e 1.635 de hipertrofia.

O conjunto de dados $D3$ representa uma contribuição estratégica para o avanço das pesquisas em segmentação semântica multiclasse de lesões podocitárias. Sua diversidade de técnicas de coloração, variedade de tipos de lesões e quantidade significativa de anotações proporcionam uma base robusta para o treinamento, validação e teste de modelos computacionais em um cenário realista e desafiador. Dessa forma, $D3$ não apenas supre a

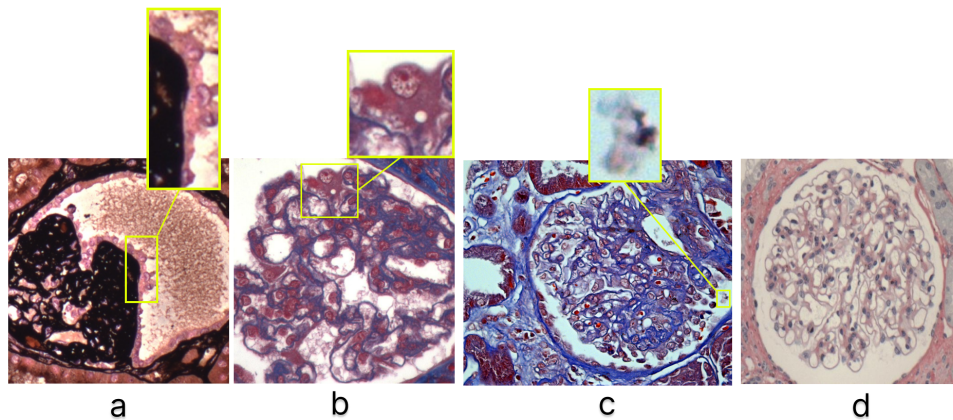


Figura 5.3: Exemplos de amostras do conjunto de dados. As condições observadas incluem hiperplasia (a), hipertrofia (b), degeneração (c) e ausência de lesões evidentes (d)..

escassez de conjuntos de dados públicos voltados à podocitopatia, mas também viabiliza o desenvolvimento de abordagens mais generalizáveis e aplicáveis na prática médica.

Capítulo 6

Identificação de Podocitopatia: PodNet

Nesta etapa do trabalho, o objetivo foi identificar a presença ou ausência de podocitopatia em imagens de glomérulos renais, ou seja, realizar uma classificação binária dessas imagens. O conjunto de dados utilizado foi o *D1v1*. Esta etapa foi conduzida em duas fases: avaliação de modelos de referência (*baselines*) e desenvolvimento de uma nova proposta de solução capaz de superar o desempenho dos modelos de referência.

6.0.1 Avaliação dos modelos de referência (*baselines*)

Na avaliação dos modelos de referência, foram analisadas seis arquiteturas de redes convolucionais (Seção 2.3): *ResNet101 v2* [84], *VGG19* [79], *DenseNet201* [85], *Inception ResNet v2* [81], *Inception v3* [80] e *Xception* [86]. Cada modelo foi treinado em dois cenários distintos: (i) com transferência de aprendizado, utilizando redes pré-treinadas no conjunto de dados ImageNet [139], e (ii) com treinamento a partir do zero, empregando inicialização aleatória dos pesos.

Além disso, os modelos foram avaliados em duas versões do conjunto de dados: imagens em RGB e imagens convertidas para escala de cinza. Essa abordagem visou investigar possíveis vantagens ou desvantagens da utilização da informação de cor no processo de aprendizado das redes.

As métricas utilizadas para avaliação dos modelos de classificação foram: precisão (*precision*), revocação (*recall*), *f1-score* e área sob a curva ROC (AUC) [119] (Seção 2.6). O ranqueamento dos melhores modelos foi baseado no *f1-score*, métrica que equilibra precisão e revocação, sendo especialmente adequada para cenários com conjuntos de dados desbalanceados.

Antes do treinamento dos modelos, realizou-se um ajuste de hiperparâmetros utilizando uma divisão aleatória em conjuntos de treino e teste. A estratégia adotada foi o *Hyperband* [88], implementado por meio da biblioteca Keras Tuner [152]. Os hiperparâmetros ajustados e seus respectivos espaços de busca foram: tamanho do *batch* (16, 32 e 64), número de neurônios nas últimas camadas densas (2048, 1024, 512 e 256), taxa de aprendizado (de 0,1 a 10^{-7}), funções de ativação das últimas camadas densas (*ReLU* e *tanh*), otimizadores (*Adam*, *SGD* e *Adamax*), *momentum* [24] (0,3, 0,6 e 0,9) e funções de perda (*binary cross-entropy* e *hinge*).

Após o ajuste de hiperparâmetros, uma nova divisão do conjunto de dados foi realizada em dois subconjuntos: (i) conjunto de generalização (70%) e (ii) conjunto de validação final (30%). A divisão preservou a proporção de classes em ambos os conjuntos. Os modelos de referência foram treinados e validados no conjunto de generalização utilizando validação cruzada com 5 *folds*. Adicionalmente, antes do treinamento, aplicou-se aumento de dados (*data augmentation*) nos conjuntos de treino da validação cruzada.

As operações de aumento de dados consistiram em transformações clássicas que preservaram as características relevantes das imagens de glomérulos. Entre as operações aplicadas, destacam-se: inversão vertical e horizontal (*flip*), rotações de 30°, 90° e 270°, ajuste de brilho (variação de 0,1 a 0,3) e zoom aleatório (entre 0,1 e 0,3 vezes). Tais operações visaram aumentar a capacidade de generalização das redes, ampliando a diversidade de exemplos disponíveis no treinamento.

O treinamento dos modelos foi realizado com um limite máximo de 200 épocas. Entretanto, nenhum dos modelos treinados ultrapassou 130 épocas, nem permaneceu abaixo de 50 épocas, em função do critério de parada antecipada (*early stopping*) estabelecido: interrupção do treinamento após 5 épocas consecutivas sem melhoria na função de perda (*loss*) de validação. Essa estratégia foi adotada para evitar o sobreajuste (*overfitting*). Os pesos finais de cada modelo foram definidos a partir da época em que a *loss* de validação atingiu seu menor valor.

A validação cruzada com 5 *folds* gerou cinco modelos distintos, os quais foram avaliados no conjunto de validação final. O modelo que apresentou o melhor desempenho nesse conjunto foi selecionado como o melhor modelo de referência.

6.0.2 Solução proposta: Modelo PodNet

Após a avaliação dos modelos *baselines* nas configurações apresentadas, foi proposto um método para a classificação de imagens de glomérulos renais quanto à presença ou ausência de podocitopatia.

A solução desenvolvida fundamentou-se em duas hipóteses: (i) a conversão de imagens de glomérulos para espaços de cor que isolam informações dos corantes em canais

distintos pode permitir que redes convolucionais extraíam características complementares, conforme observado nos estudos de Barros *et al.* [18] e Bukowy *et al.* [153]; e (ii) a extração de características por redes convolucionais pré-treinadas pode gerar resultados superiores aos obtidos com o treinamento de redes fim a fim, como demonstrado nos trabalhos de Chagas *et al.* [19], Minamiguchi *et al.* [141] e Mathur *et al.* [136].

O modelo proposto, denominado *PodNet* [52], é um sistema de reconhecimento de podocitopatia em imagens de glomérulos renais, organizado em três etapas principais: pré-processamento, extração de características e classificação. A Figura 6.1 apresenta a topologia geral da rede e suas respectivas etapas.

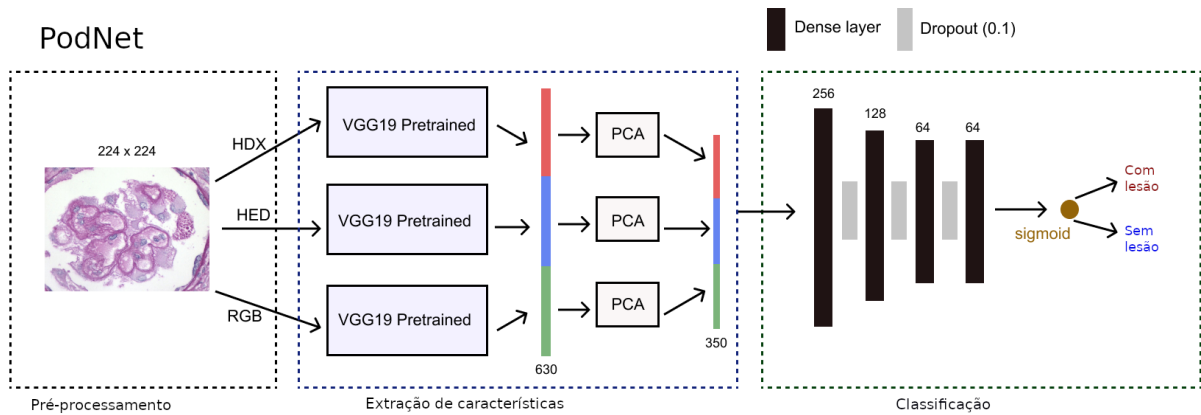


Figura 6.1: Etapas básicas da proposta de solução (PodNet).

Na primeira etapa, de pré-processamento, as imagens são convertidas de RGB para os espaços de cor HED e HDX e, em seguida, normalizadas para valores entre 0 e 1. O espaço de cor HED, utilizado por [18] e [153], gera imagens cujos canais representam as contribuições dos corantes Hematoxilina, Eosina e DAB. Por sua vez, o espaço de cor HDX produz imagens com canais correspondentes aos corantes Hematoxilina e PAS.

A segunda etapa corresponde à extração de características. Nessa fase, as imagens convertidas para os espaços de cor HED e HDX, juntamente com a imagem original em RGB, são processadas pela arquitetura VGG19, previamente treinada no conjunto de dados ImageNet [139]. A arquitetura foi adaptada para que sua saída correspondesse ao resultado da operação de *max pooling* [66] da última camada convolucional. Em seguida, aplica-se uma operação de *flattening*, convertendo a matriz tridimensional em um vetor unidimensional [24], gerando um vetor de características para cada imagem. Esses vetores são posteriormente redimensionados utilizando o algoritmo PCA (*Principal Component Analysis*) [154], com o objetivo de reduzir a dimensionalidade, acelerar o treinamento e eliminar características redundantes ou menos relevantes. Por fim, os três vetores de

características (RGB, HED e HDX) são concatenados, resultando em um único vetor integrado.

Na terceira etapa, de classificação, utiliza-se uma rede neural densa composta por quatro camadas ocultas, contendo 256, 128, 64 e 64 neurônios, respectivamente. A regularização entre as camadas ocultas é realizada com a técnica de *dropout*, adotando uma taxa de 0,1. Os hiperparâmetros dessa rede densa, incluindo o número de neurônios por camada, foram ajustados por meio do método *grid search* [155], o qual avalia sistematicamente diferentes combinações de parâmetros. A camada de saída é composta por um único neurônio, ativado pela função *sigmoid*.

Capítulo 7

Implicações do Uso de Classificadores Automatizados na Prática Médica

Nesta etapa do trabalho, o objetivo foi estudar o impacto do uso de classificadores automatizados na tomada de decisão dos patologistas quanto à presença ou ausência de podocitopatia em imagens de glomérulos. O protocolo experimental adotado foi estruturado em três fases e baseou-se no trabalho de Ligabue *et al.* [156], que comparou o desempenho de um modelo de rede neural convolucional (CNN) com o de um grupo de três patologistas.

Na primeira fase, objetivou-se obter um classificador automatizado capaz de reconhecer glomérulos com lesão podocitária visível por microscopia óptica. Diferentes modelos de CNN foram ajustados, treinados e validados utilizando o conjunto de dados denominado *D1v2*. O modelo de melhor desempenho foi selecionado com base nos resultados obtidos na classificação de imagens de um segundo conjunto de dados, *D2*.

Na segunda fase, avaliou-se o desempenho de um grupo de três patologistas na classificação das imagens do conjunto *D2*, sem qualquer suporte computacional, permitindo a comparação direta de seu desempenho com o do modelo automatizado.

Na terceira fase, após um intervalo de 30 dias — adotado para minimizar o viés de memória —, os mesmos patologistas reavaliaram as imagens de *D2*, desta vez com acesso às classificações fornecidas pelo classificador automatizado. O objetivo foi verificar se a disponibilização dessa informação influenciaria (isto é, melhoraria) a acurácia dos especialistas.

A Figura 7.1 apresenta uma visão geral do protocolo experimental. Na Etapa 1, foram avaliados diferentes modelos de CNN no conjunto *D1v2* e reavaliados no conjunto *D2*. Na Etapa 2, foi analisado o desempenho dos três patologistas na classificação das imagens

de *D2* sem suporte. Na Etapa 3, avaliou-se o desempenho dos mesmos patologistas na classificação das imagens de *D2* com o suporte do classificador automatizado.

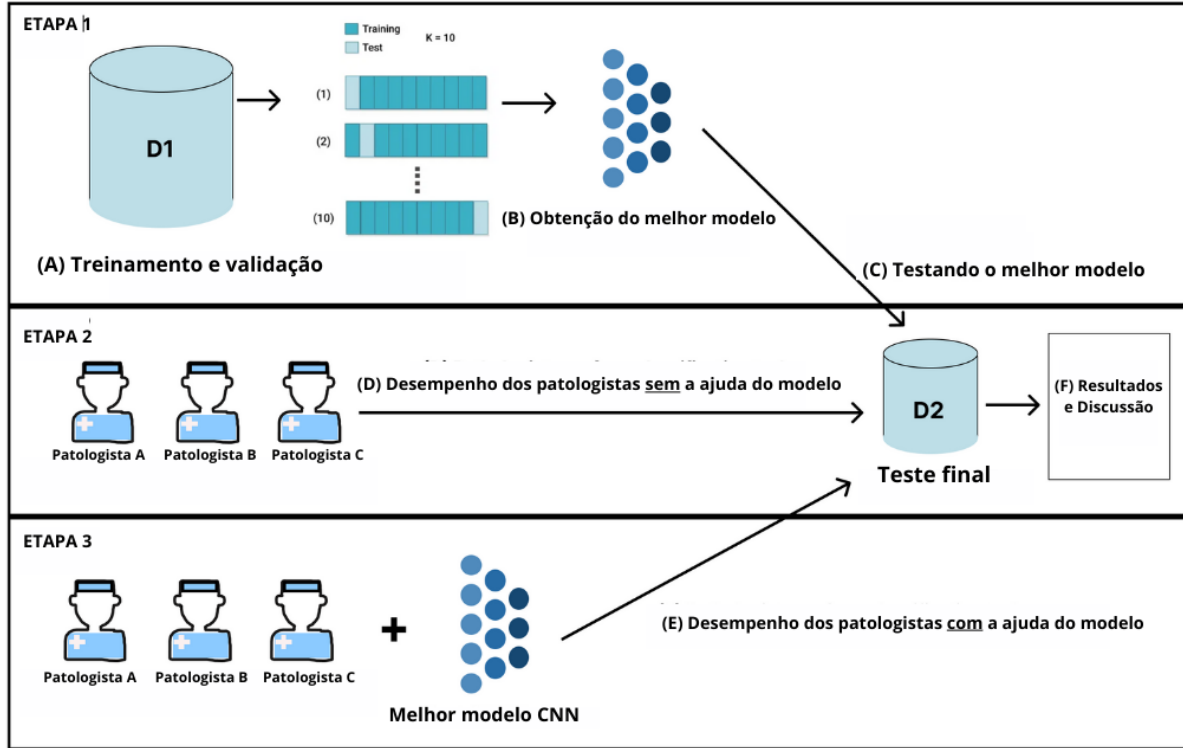


Figura 7.1: Visão geral do protocolo experimental.

7.0.1 Treinamento e validação dos modelos de classificação

Três arquiteturas clássicas de redes neurais convolucionais (CNNs) [157] foram selecionadas para a classificação automatizada de imagens: *Inception ResNet-v2* [158], *DenseNet201* [159] e *EfficientNet B3* [160].

A escolha dessas arquiteturas foi motivada por dois fatores principais: i) a confiabilidade demonstrada em tarefas semelhantes [161, 162, 163]; e ii) suas diferenças estruturais, tanto em profundidade (número fixo de camadas em *DenseNet201* e *Inception ResNet-v2*, e variável em *EfficientNet B3*) quanto em estratégias de aprendizado (*e.g.*, blocos residuais no *Inception ResNet-v2* e conexões densas no *DenseNet201*).

Cada arquitetura foi treinada em dois cenários distintos:

- **Treinamento do zero** (*from scratch* — FS): inicialização aleatória dos pesos;

- **Aprendizado por transferência** (*transfer learning* — TL): utilizando pesos pré-treinados no conjunto *ImageNet* [63].

Essa abordagem resultou em seis modelos distintos, considerando as combinações entre arquitetura e estratégia de treinamento.

Os hiperparâmetros de cada modelo foram ajustados utilizando o algoritmo *Hyperband* [164, 165], que emprega uma estratégia de otimização bayesiana associada à alocação adaptativa de recursos e interrupção antecipada para acelerar o processo de busca.

Para o ajuste dos hiperparâmetros, o conjunto de dados *D1v2* foi dividido aleatoriamente em dois grupos: 70% para treinamento e 30% para teste, mantendo a proporção entre as classes (com e sem degeneração podocitária). Essa divisão foi utilizada exclusivamente para o ajuste dos hiperparâmetros, visando mitigar problemas de subajuste (*underfitting*) e sobreajuste (*overfitting*).

O algoritmo *Hyperband* ajustou hiperparâmetros críticos, tais como: taxa de aprendizado (*learning rate*), otimizador, tamanho do *batch*, função de perda (*loss function*) e o número de neurônios nas camadas densas. A Tabela 7.1 apresenta os valores considerados para cada hiperparâmetro, enquanto a Tabela 7.2 mostra os melhores valores obtidos para cada modelo.

Tabela 7.1: Intervalo de valores considerados durante o ajuste de hiperparâmetros.

| Hiperparâmetro | Valores avaliados |
|--|--|
| Tamanho do <i>batch</i> | 32, 16, 8, 4, 2 |
| Função de perda | <i>Binary Cross-entropy</i> , <i>Hinge</i> |
| Otimizador | SGD, Adam, Adagrad, RMSprop |
| Taxa de aprendizado | 0,1; 0,001; 0,0001; 0,00001 |
| Número de neurônios nas camadas densas | 2048, 1024, 512, 256 |

Tabela 7.2: Melhores valores de hiperparâmetros para cada modelo avaliado. LR: *Learning rate*

| Modelo | Batch size | Perda | Otimizador | Camadas | LR |
|-------------------------------|------------|--------------|------------|------------|--------|
| <i>Inception ResNet-v2</i> TL | 8 | <i>BCE</i> | RMSprop | 1024, 512 | 0,0001 |
| <i>DenseNet201</i> TL | 8 | <i>BCE</i> | Adam | 512, 512 | 0,001 |
| <i>EfficientNet B3</i> TL | 4 | <i>Hinge</i> | RMSprop | 512, 512 | 0,0001 |
| <i>DenseNet201</i> FS | 4 | <i>BCE</i> | SGD | 2048, 1024 | 0,001 |
| <i>Inception ResNet-v2</i> FS | 8 | <i>BCE</i> | Adam | 1024, 512 | 0,0001 |
| <i>EfficientNet B3</i> FS | 8 | <i>BCE</i> | Adam | 512, 512 | 0,0001 |

Após a definição da melhor combinação de hiperparâmetros para cada modelo, conforme ilustrado na Figura 7.1, foi realizada uma nova divisão do conjunto de dados, e os modelos candidatos foram treinados por meio de validação cruzada estratificada em

10 subconjuntos [122]. Em cada iteração, um subconjunto foi utilizado para validação, enquanto os demais serviram para treinamento. Esse processo gerou 10 modelos distintos para cada uma das seis combinações avaliadas.

Durante o treinamento na validação cruzada, foi aplicada uma estratégia de aumento de dados (*data augmentation*) com o objetivo de aumentar a diversidade do conjunto de treinamento e mitigar problemas de sobreajuste [24]. As seguintes transformações foram empregadas:

- Rotações em ângulos variando de 30° a 310° (com intervalos de 30°);
- Espelhamento vertical e horizontal;
- Ajustes aleatórios de brilho;
- Variações de contraste.

Essas transformações foram cuidadosamente selecionadas para preservar a morfologia das estruturas histológicas dos glomérulos, evitando a introdução de artefatos que pudessem prejudicar o treinamento dos modelos.

Graças à aplicação da técnica de aumento de dados, o número de imagens nos conjuntos de treinamento foi ampliado de 1.043 para 12.516, com cada imagem original gerando 12 novas variações. Essa expansão contribuiu significativamente para a melhoria da capacidade de generalização dos classificadores.

Adicionalmente, foi empregada a estratégia de parada antecipada (*early stopping*) [24], interrompendo o treinamento sempre que, em uma sequência de 10 épocas, não fosse observada redução na função de perda (*loss*). Essa prática é amplamente utilizada para evitar o sobreajuste e o treinamento excessivo dos modelos.

O melhor modelo foi selecionado com base no maior valor médio de *F1-score* obtido nas validações cruzadas. Em seguida, os modelos ótimos de cada combinação de arquitetura e estratégia de treinamento foram comparados entre si.

Por fim, após a seleção final, o melhor modelo avaliado no conjunto *D1v2* foi também testado no conjunto *D2*, permitindo validar seu desempenho em um conjunto de dados independente.

7.0.2 Classificação dos patologistas

Para estabelecer as bases de comparação do desempenho entre os modelos de CNN, três patologistas classificaram as imagens do conjunto *D2* como "com" ou "sem" degeneração podocitária. Essa tarefa foi realizada exclusivamente por meio da inspeção visual das imagens, sem qualquer informação adicional.

É importante destacar que os patologistas envolvidos na classificação do conjunto $D2$ não participaram da preparação ou anotação dos conjuntos de dados utilizados neste estudo ($D1v2$ e $D2$). Cada especialista estava vinculado a uma instituição distinta de nefropatologia, o que garantiu maior diversidade e independência nas avaliações.

Para mitigar possíveis vieses de classificação decorrentes da reanálise imediata, foi implementado um intervalo de 30 dias entre as duas sessões de avaliação. Após esse período, os mesmos patologistas reclassificaram as imagens de $D2$, desta vez com acesso às predições fornecidas pelo classificador automatizado. Em outras palavras, antes de tomar uma decisão sobre cada imagem, os especialistas puderam consultar a recomendação gerada pelo modelo computacional.

Essa abordagem permitiu avaliar o desempenho dos patologistas tanto em condições de análise independente quanto sob o suporte do classificador, minimizando o risco de dependência excessiva em relação às sugestões automatizadas.

Para cada uma das duas sessões de classificação realizadas, foi calculado o coeficiente de concordância entre os nefropatologistas. Essa métrica foi fundamental para avaliar eventuais alterações na consistência diagnóstica decorrentes da utilização do suporte computacional.

Capítulo 8

Segmentação de Lesões Podocitárias

Neste capítulo, são apresentados os procedimentos adotados para a avaliação dos modelos de referência (*baselines*) e o desenvolvimento da solução proposta para a segmentação semântica multiclasse de lesões podocitárias em imagens de glomérulos renais. Inicialmente, são descritas as estratégias de tratamento do desequilíbrio entre classes e colorações no conjunto de dados, bem como a estruturação dos subconjuntos de treinamento, validação e teste. Em seguida, detalha-se a seleção e treinamento de múltiplas arquiteturas clássicas de segmentação combinadas a diferentes codificadores, formando o conjunto de modelos de referência. Posteriormente, é apresentada a abordagem inovadora desenvolvida neste trabalho, baseada na fusão de características extraídas de espaços de cores distintos (RGB e HED). Essa solução visa aprimorar a capacidade dos modelos em lidar com a heterogeneidade cromática e o desequilíbrio de classes, aspectos críticos para a segmentação precisa de estruturas histológicas lesionadas.

8.0.1 Avaliação dos modelos de referência (*baselines*)

Antes do treinamento e validação dos modelos de referência, foi crucial abordar o desequilíbrio inerente entre classes e colorações no conjunto de dados $D3$, a fim de garantir um treinamento efetivo dos modelos. Para isso, foram implementadas estratégias específicas. Uma delas consistiu na redução direcionada de imagens contendo apenas podócitos sem lesões, todas coradas com PAS, diminuindo seu número de 828 para 528. Essa redução foi essencial devido à incidência significativamente maior de podócitos sem lesões, servindo como medida para mitigar o desequilíbrio entre as classes.

Além disso, foi aplicada uma ampliação de dados às imagens com casos de lesões, conforme detalhado na próxima seção. O conjunto de dados foi então dividido em três subconjuntos — treinamento, validação e teste — garantindo que a proporção de imagens com lesões e diferentes colorações fosse mantida de forma consistente em todos os

subconjuntos. A Tabela 8.1 descreve a distribuição das imagens por corante em cada subconjunto.

Tabela 8.1: Proporção de colorações presentes em cada subconjunto.

| Subconjuntos | PAS | HE | Gomori | Mason | PAMS |
|--------------|-----------|-----------|---------|-----------|-----------|
| Treinamento | 368 | 125 | 26 | 85 | 143 |
| Validação | 53 | 18 | 4 | 13 | 21 |
| Teste | 107 | 36 | 9 | 26 | 42 |
| Total | 528 (49%) | 179 (17%) | 39 (4%) | 124 (11%) | 206 (19%) |

Para combater os efeitos do desequilíbrio entre as classes no processo de treinamento, foi empregada uma estratégia de aumento de dados (*data augmentation*). Foram aplicadas aleatoriamente combinações de rotações, espelhamentos, suavizações, ajustes de contraste e transformações de iluminação nas imagens do conjunto de treinamento. Essa abordagem aumentou artificialmente o tamanho do conjunto de dados, contribuindo para mitigar o desequilíbrio entre imagens com e sem lesões. Como resultado, o número de imagens com lesões foi ampliado em um fator de 30, enquanto o número de imagens sem lesões cresceu em um fator de 5. Dessa forma, o conjunto de treinamento expandiu-se de 747 para 14.111 imagens, criando uma base mais generalizada e diversificada para o treinamento dos modelos.

O treinamento dos modelos de referência serviu como ponto de partida para o desenvolvimento da abordagem proposta para a segmentação das classes de lesões podocitárias. Como os trabalhos relacionados não abordam diretamente a segmentação dos diferentes tipos dessas lesões, foi necessário explorar um conjunto mais amplo e diversificado de arquiteturas de rede e *encoders*, adotando diferentes estratégias de aprendizado para superar essa lacuna.

Os modelos de referência foram construídos a partir de nove arquiteturas clássicas de segmentação: *DeepLabv3*, *DeepLabv3+* [106], *Unet* [103], *Unet++* [104], *Linknet* [109], *PSPnet* [110], *FPN* [111], *MAnet* [112] e *PAN* [113]. Essas arquiteturas são bem estabelecidas na literatura [166], e algumas delas, como *Unet* e *Unet++*, são amplamente utilizadas na segmentação de imagens histológicas [167].

Os modelos avaliados foram gerados pela combinação dessas nove arquiteturas com trinta e cinco (35) *encoders* — todos disponíveis na biblioteca Segmentation Models Pytorch¹ — resultando em trezentos e quinze (315) combinações (35×9). Cada modelo foi treinado em duas condições de inicialização de pesos: inicialização aleatória e pesos pré-treinados no ImageNet [139]. Além disso, as imagens foram representadas em dois espaços de cor distintos (RGB e HED), totalizando 1260 modelos avaliados (315×4).

¹https://github.com/qubvel/segmentation_models.pytorch

Cada um desses 1260 modelos foi treinado tanto no espaço de cor RGB quanto no espaço HED, gerando, ao final, um total de 2520 (1260×2) modelos de referência analisados.

Durante o treinamento, os hiperparâmetros foram padronizados para todos os modelos: a taxa de aprendizado foi inicializada em 0,0001, com um tamanho de *batch* de 16. Utilizou-se a função de ativação *Softmax 2d* e a função de perda *Dice loss* [168]. Foram atribuídos pesos específicos para cada classe (0,0040, 0,0840, 0,4408, 0,3498 e 0,1214), como uma estratégia de aprendizado sensível ao custo, considerando a proporção de cada classe no conjunto de dados [169, 170]. As métricas de avaliação utilizadas foram o *F-score* e o coeficiente de *Jaccard* [171], tanto em nível global quanto por classe.

O número máximo de épocas foi fixado em 500, com adoção da estratégia de parada antecipada: o treinamento era interrompido se não houvesse melhora na função de perda após 15 épocas consecutivas, medida empregada para evitar sobreajuste. A versão do modelo com a menor perda de validação foi salva como o melhor resultado.

Após o treinamento e validação de todos os modelos de referência, foram selecionados trinta modelos com base em seu *F1-score* para análise comparativa da nova solução proposta (Seção 8.0.2). Esse conjunto foi composto pelos 10 modelos com melhor desempenho, pelos 10 modelos com pior desempenho e por 10 modelos selecionados aleatoriamente (sem sobreposição com os dois primeiros grupos). Todos os modelos foram retreinados utilizando o novo método proposto, mantendo inalterados os hiperparâmetros e critérios de parada, de modo que a comparação focasse exclusivamente no impacto da nova estratégia de aprendizado.

A partir dos resultados obtidos com os modelos de referência, identificou-se a necessidade de desenvolver uma abordagem mais robusta e sensível às particularidades do conjunto de dados estudado. Nesse contexto, a próxima seção apresenta a solução proposta, baseada na fusão de características extraídas de múltiplos espaços de cores, com o objetivo de aprimorar a segmentação semântica das lesões podocitárias.

8.0.2 Solução proposta

Optou-se pelo desenvolvimento de uma nova abordagem de fusão de características, que combina informações dos espaços de cores RGB e HED para treinar modelos de segmentação semântica. O foco principal é enfrentar os desafios impostos por conjuntos de dados desequilibrados e multicoloridos na segmentação de lesões podocitárias. Seja θ um modelo de segmentação semântica com arquitetura baseada em codificadores e decodificadores, conforme definido nas Equações 8.1 e 8.2.

$$\theta : x_1 \times x_2 \rightarrow G, \quad (8.1)$$

$$(M_1, M_2) = \theta(x_1, x_2), \quad (8.2)$$

Onde:

- G é o conjunto dos possíveis mapas de atributos, sendo $G \subseteq \mathbb{R}^{h \times w \times c'}$, onde c' representa o número de classes e $h \times w$ a resolução das imagens;
- x_1 e x_2 são as entradas correspondentes às imagens nos espaços de cor RGB e HED, respectivamente;
- M_1 e M_2 são os mapas de características resultantes dos blocos decodificadores nos espaços RGB e HED, respectivamente.

Após a extração dos mapas de características M_1 e M_2 por meio de seus respectivos codificadores e decodificadores, é realizada uma operação de fusão (\oplus), resultando no mapa combinado N (Equação 8.3).

$$N = M_1 \oplus M_2. \quad (8.3)$$

Finalmente, o mapa resultante N é submetido ao bloco de segmentação original da rede (S), gerando a máscara de saída Y (Equação 8.4):

$$Y = S(N). \quad (8.4)$$

Desenvolvimento e validação do método proposto

O método desenvolvido modifica o fluxo convencional de treinamento de redes de segmentação semântica multiclasse, criando uma arquitetura multicanal. A Figura 8.1 ilustra a proposta, que segue as etapas descritas a seguir:

1. Conversão das imagens do espaço de cores RGB para HED;
2. Duplicação do fluxo de dados, com codificadores e decodificadores específicos para cada espaço de cor;
3. Fusão dos mapas de características gerados antes do bloco de segmentação;
4. Aplicação do bloco de segmentação original da arquitetura.

A inspiração para o desenvolvimento dessa abordagem, bem como a ideia inicial de utilizar imagens em diferentes espaços de cores, surgiu de estudos que investigam o impacto do uso de múltiplos espaços de cor no treinamento de CNNs para problemas de

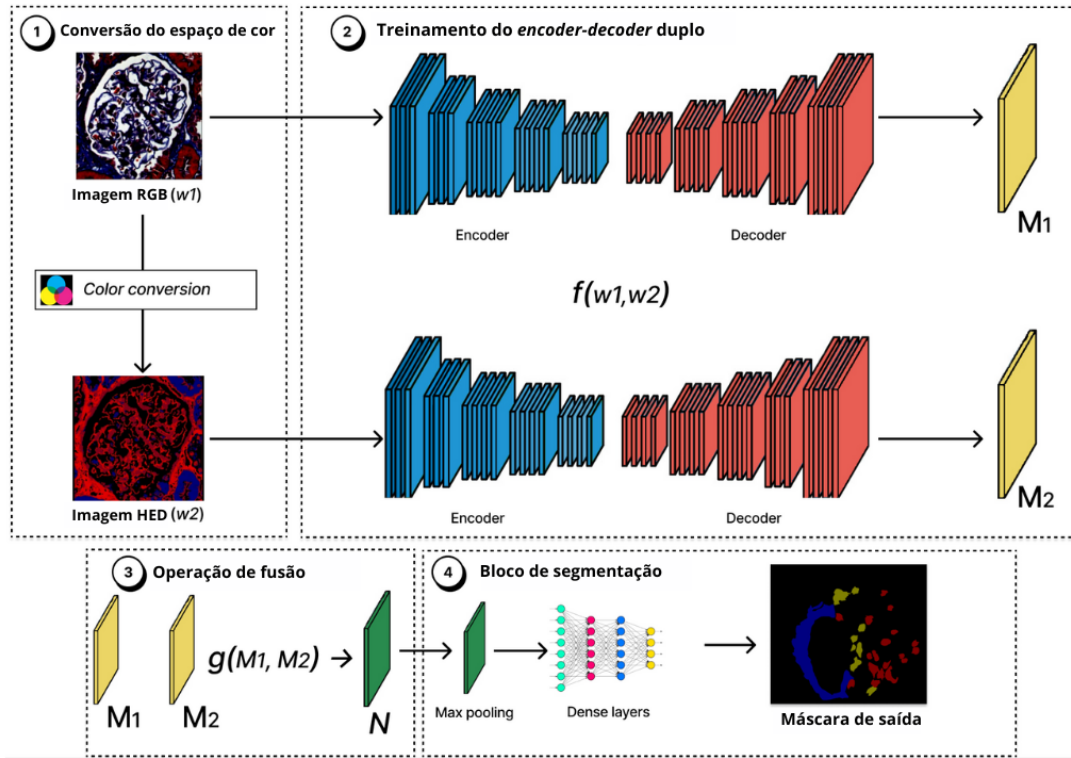


Figura 8.1: Visão geral da abordagem proposta: (1) conversão do espaço de cores, (2) extração de características usando blocos encoder-decoder, (3) fusão das saídas dos decoders e (4) bloco de segmentação.

classificação de imagens histológicas [54, 55], e na melhoria do desempenho por meio de diferentes representações cromáticas [172, 173].

Considerando que o espaço HED oferece uma representação especializada para imagens histológicas [56], esta abordagem propõe combinar o HED com o espaço RGB, de modo a explorar informações complementares.

Para extrair características de cada representação de forma independente, foi empregada a duplicação dos blocos de codificadores e decodificadores, formando um novo modelo multicanal. Essa estratégia foi inspirada em abordagens que propõem arquiteturas multicanal ou *multi-stream* [174, 175, 176, 177, 178, 179].

A fusão dos mapas de características extraídos dos decodificadores foi baseada em estudo que avalia diferentes estratégias de fusão para redes multicanal [118]. As operações de fusão consideradas estão descritas na Tabela 8.2, visando investigar os efeitos específicos de cada estratégia no desempenho da segmentação.

As hipóteses consideradas para cada operação de fusão foram:

Tabela 8.2: Operações de fusão avaliadas no método proposto.

| Operação de fusão | Definição |
|------------------------------|---|
| Soma das matrizes de entrada | $N = M_1 + M_2$ (8.5) |
| Diferença absoluta | $N = M_1 - M_2 $ (8.6) |
| Máximo elemento a elemento | $N = \max(M_{1,i,j}, M_{2,i,j})$ (8.7) |
| Mínimo elemento a elemento | $N = \min(M_{1,i,j}, M_{2,i,j})$ (8.8) |
| Média elemento a elemento | $N = \frac{M_{1,i,j} + M_{2,i,j}}{2}$ (8.9) |

- **Soma** (Equação 8.5): permite a integração das informações complementares dos mapas, mas pode amplificar ruídos caso ambos contenham informações imprecisas.
- **Diferença absoluta** (Equação 8.6): destaca regiões de alta discrepância entre os mapas, mas pode negligenciar informações contextuais em áreas de baixo contraste.
- **Máximo elemento a elemento** (Equação 8.7): enfatiza as características mais proeminentes, mas pode levar à perda de detalhes sutis quando ignora valores inferiores relevantes.
- **Mínimo elemento a elemento** (Equação 8.8): reforça regiões de consenso entre os mapas, reduzindo o impacto de ruído, embora possa eliminar informações relevantes associadas a valores mais altos.
- **Média elemento a elemento** (Equação 8.9): promove uma combinação equilibrada dos mapas, mas pode suavizar excessivamente características importantes.

Na etapa final, o mapa de características fundido N é encaminhado para o bloco de segmentação original da arquitetura, sem modificações adicionais.

A abordagem desenvolvida neste trabalho propõe uma estratégia inovadora de fusão de características, combinando informações extraídas dos espaços de cor RGB e HED para o aprimoramento da segmentação semântica de lesões podocitárias. Ao duplicar o fluxo de codificadores e decodificadores e integrar os mapas de atributos por meio de diferentes operações de fusão, a proposta visa explorar representações complementares das imagens histológicas, mitigar os efeitos do desequilíbrio de classes e melhorar a robustez dos modelos frente à variabilidade de coloração. Dessa forma, a solução apresentada contribui de maneira significativa para a literatura de patologia computacional, oferecendo uma alternativa promissora para a segmentação multiclasse em cenários desafiadores, caracterizados por heterogeneidade cromática e escassez de dados anotados.

Capítulo 9

Resultados

Este capítulo apresenta os resultados obtidos pelos modelos de referência (*baselines*) e pelas soluções propostas nas etapas de identificação de podocitopatia, estudo sobre o impacto do uso de classificadores na prática médica e segmentação de lesões podocitárias, conforme descrito no Capítulo 4.

9.1 Identificação de podocitopatia: PodNet

A Tabela 9.1 apresenta os resultados obtidos para todos os modelos de classificação avaliados. Os resultados no conjunto de generalização correspondem às médias obtidas na validação cruzada, acompanhadas de seus respectivos desvios padrão. Já os resultados da validação final foram calculados com base no conjunto de teste separado no início dos experimentos.

A PodNet (modelo proposto) alcançou os melhores resultados de *F1-score* no conjunto de validação final. Nos conjuntos de generalização, a média de *F1-score* obtida foi equivalente à do modelo ResNet101 v2 (treinado em RGB com transferência de aprendizado), que atingiu $90,2 \pm 3,54$, enquanto a PodNet obteve $90,1 \pm 1,70$. Considerando a diferença ínfima entre os modelos e o menor desvio padrão apresentado pela PodNet, também é possível considerá-la superior nos conjuntos de generalização.

As Figuras 9.1, 9.2 e 9.3 apresentam as curvas ROC dos modelos de referência em cada contexto de treinamento. Entre os modelos treinados com inicialização aleatória de pesos e utilizando imagens em escala de cinza, o melhor desempenho foi obtido pela *Inception v3* (Figura 9.2). No mesmo cenário de inicialização aleatória, porém utilizando imagens em RGB, o modelo de melhor desempenho foi a *Inception ResNet v2* (Figura 9.1). Já no contexto de treinamento com transferência de aprendizado e imagens em RGB, o destaque foi para a *DenseNet201* (Figura 9.3).

A Figura 9.4 apresenta a curva ROC dos quatro melhores modelos classificados pelo *F1-score*: *PodNet*, *DenseNet201 TL (RGB)*, *Inception v3 TL (RGB)* e *ResNet101 v2 TL (RGB)*.

Tabela 9.1: Resultados obtidos nos modelos de classificação. TL: *transfer learning*. FS: *from scratch*.

| Classificação | Conjunto de generalização (média) | | | Conjunto de validação | | | |
|----------------------------------|-----------------------------------|------------------|------------------|-----------------------|----------------|---------------|--------------|
| Modelos | <i>Prec</i> (%) | <i>Rec</i> (%) | <i>F1</i> (%) | <i>Prec</i> (%) | <i>Rec</i> (%) | <i>F1</i> (%) | <i>AUC</i> |
| PodNet (solução proposta) | 90,6±3,07 | 89,6±1,36 | 90,1±1,70 | 88,9 | 93,2 | 90,9 | 0.959 |
| Densenet201 TL (RGB) | 90,0±3,66 | 90,0±5,31 | 88,0±3,89 | 85,0 | 91,0 | 87,8 | 0.935 |
| Inception v3 TL (RGB) | 87,0 ± 1,03 | 88,0 ± 8,93 | 87,4 ± 4,76 | 81,0 | 90,0 | 85,2 | 0.928 |
| Resnet101 v2 TL (RGB) | 94,0±2,65 | 86,0±7,44 | 90,2±3,54 | 83,00 | 86,0 | 84,4 | 0.927 |
| VGG19 TL (RGB) | 93,0±2,56 | 86,0±6,03 | 89,3±4,11 | 87,0 | 81,0 | 83,8 | 0.919 |
| Xception TL (RGB) | 89,0±2,66 | 90,0±8,26 | 88,4±4,69 | 82,0 | 84,0 | 82,9 | 0.893 |
| Inception Resnet v2 FS (RGB) | 86,0 ± 5,9 | 75,0 ± 12,02 | 80,4 ± 8,02 | 79,0 | 87,0 | 82,8 | 0.921 |
| Inception Resnet v2 TL (RGB) | 92,0±3,76 | 90,0±5,75 | 87,4±3,22 | 79,0 | 86,0 | 82,3 | 0.896 |
| Densenet201 FS (RGB) | 82 ± 7,9 | 83 ± 9,02 | 82,0 ± 8,52 | 75,0 | 88,0 | 80,9 | 0.895 |
| Inception v3 FS (GL) | 80,0 ± 6,94 | 67,0 ± 6,02 | 78,2 ± 8,02 | 77,0 | 84,0 | 80,3 | 0.915 |
| Resnet101 v2 FS (GL) | 80,0 ± 6,7 | 82,0 ± 1,79 | 76,7 ± 9,23 | 72,00 | 92,0 | 80,3 | 0.801 |
| Inception Resnet v2 FS (GL) | 79,0 ± 7,67 | 88,0 ± 9,60 | 83,2 ± 7,44 | 83,0 | 77,0 | 79,8 | 0.888 |
| Densenet201 FS (GL) | 72,0 ± 7,8 | 84,0 ± 1,36 | 88,0 ± 7,1 | 69,0 | 91,0 | 78,4 | 0.865 |
| VGG19 FS (GL) | 86,0±1,41 | 61,0±14,4 | 71,3±8,44 | 78,0 | 79,0 | 78,4 | 0.883 |
| Xception FS (GL) | 82,0 ± 5,02 | 69,0 ± 5,82 | 75,3 ± 5,28 | 70,0 | 80,0 | 74,6 | 0.835 |
| Inception v3 FS (RGB) | 83,0 ± 4,02 | 78,0 ± 10,0 | 80,4 ± 7,42 | 71,0 | 78,0 | 74,3 | 0.816 |
| Resnet101 v2 FS (RGB) | 72,0 ± 10,1 | 65,0 ± 14,0 | 69,6 ± 11,0 | 71,00 | 74,0 | 72,4 | 0.784 |
| VGG19 FS (RGB) | 89,0±4,80 | 80,0±9,80 | 84,3±5,80 | 87,0 | 61,0 | 71,8 | 0.833 |
| Xception FS (RGB) | 76,0 ± 5,92 | 69,0 ± 12,0 | 69,4 ± 8,80 | 71,0 | 69,0 | 69,9 | 0.781 |

9.2 Impacto do uso de classificadores automatizados na prática dos patologistas

As métricas utilizadas para avaliar o desempenho do classificador automatizado e dos patologistas foram: acurácia, precisão, revocação, *F1-score* e área sob a curva (*Area Under the Curve* - AUC) [123, 122]. Além dessas métricas, também foi calculada a confiabilidade do acordo entre os três patologistas utilizando o índice *Fleiss' Kappa* [127], com o objetivo de determinar se a inclusão dos resultados do classificador automatizado alterou o nível de concordância entre os observadores.

A Tabela 9.2 apresenta os valores médios alcançados pelos modelos de CNN. O modelo que obteve o maior *F1-score* na validação cruzada com 10 subconjuntos foi o *Inception ResNet101 v2* com transferência de aprendizado, apresentando média de 94,7% de acurácia, 95,5% de precisão, 92,3% de revocação e 93,7% de *F1-score*.

As curvas ROC [123] foram utilizadas para ilustrar o desempenho dos classificadores em função da variação do limiar de decisão. Um desempenho ideal seria caracterizado por uma taxa de verdadeiros positivos igual a 1 e uma taxa de falsos positivos igual a 0.

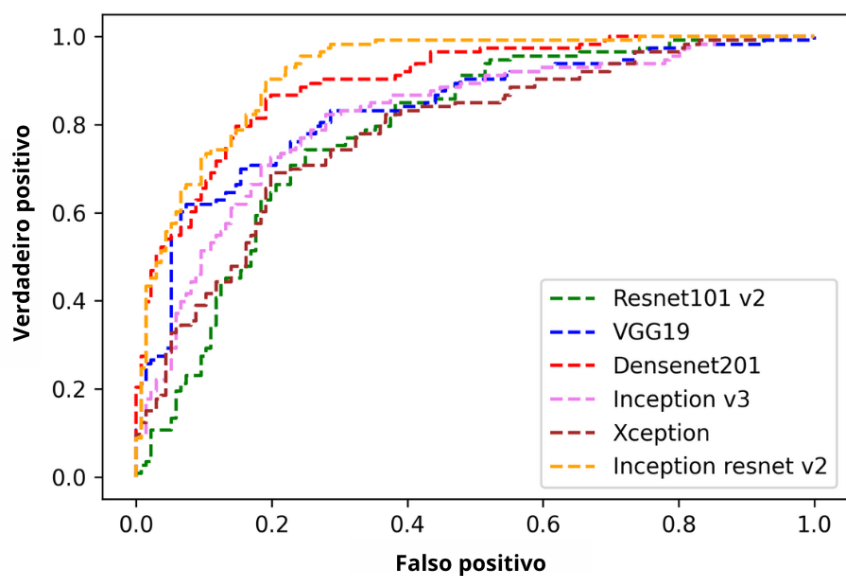


Figura 9.1: Curvas ROC: Modelos treinados do zero em RGB.

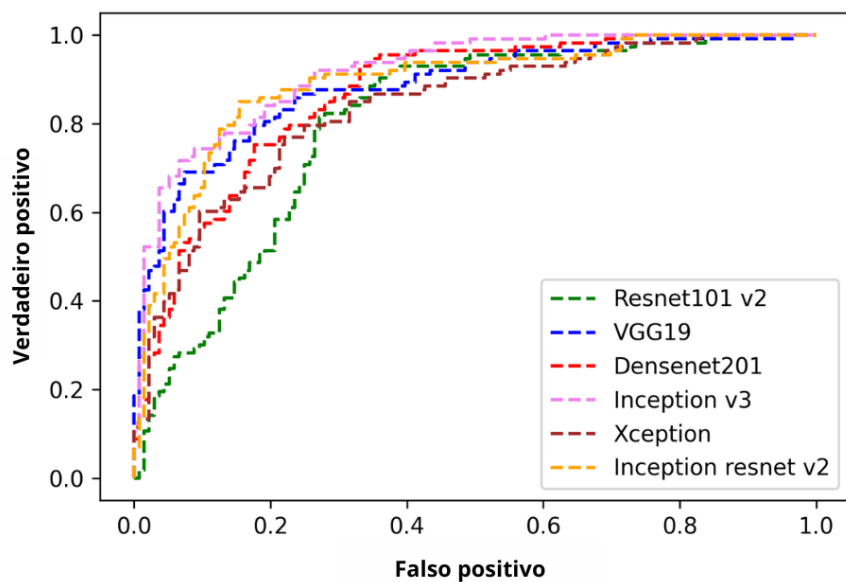


Figura 9.2: Curvas ROC: Modelos treinados do zero em nível de cinza.

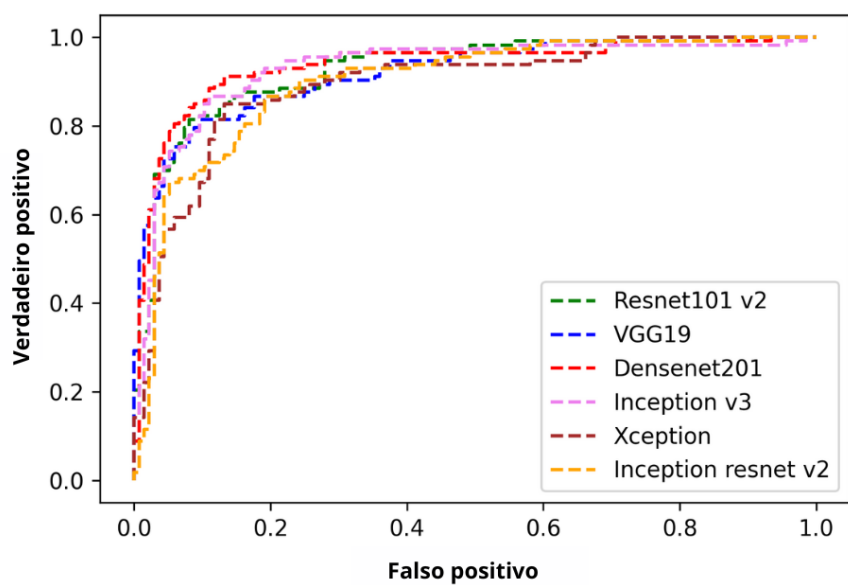


Figura 9.3: Curvas ROC: Modelos treinados com *transfer learning* em RGB.

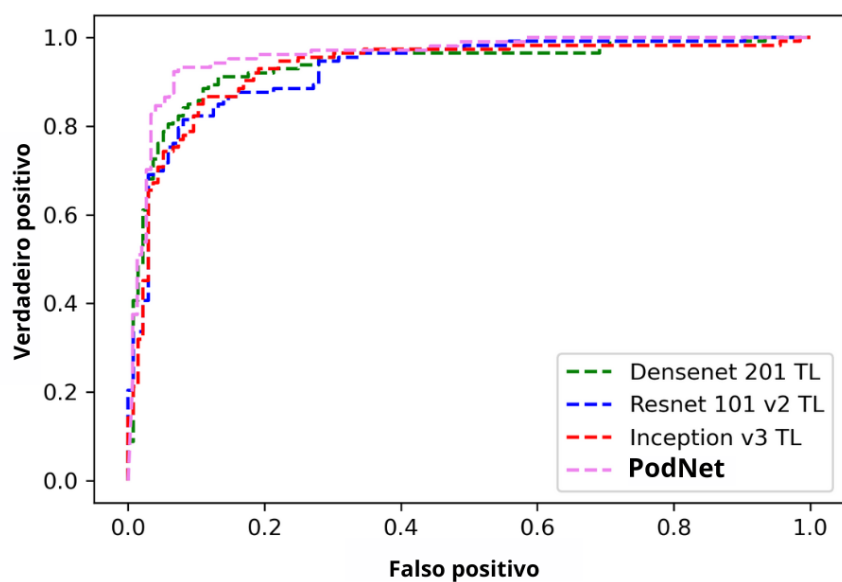


Figura 9.4: Curvas ROC: Top-4 melhores modelos.

Tabela 9.2: Resultados da validação cruzada k-fold para os modelos de CNN. Modelos com a sigla TL (*Transfer Learning*) foram pré-treinados; modelos com a sigla FS (*From Scratch*) foram treinados com inicialização aleatória de pesos.

| Modelo | Acc.(%) | Prec.(%) | Rev.(%) | F1(%) |
|----------------------------------|-----------------|-----------------|-----------------|-----------------|
| <i>Inception ResNet101 v2 TL</i> | 94,7±2,6 | 95,5±3,6 | 92,3±4,3 | 93,7±2,6 |
| <i>DenseNet201 TL</i> | 91,3±3,5 | 92,5±3,8 | 90,2±7,1 | 91,1±3,9 |
| <i>EfficientNet B3 TL</i> | 91,1±2,4 | 92,8±3,4 | 89,0±7,1 | 90,6±2,8 |
| <i>DenseNet201 FS</i> | 88,4±2,9 | 90,6±5,9 | 86,0±5,8 | 88,0±3,1 |
| <i>Inception ResNet101 v2 FS</i> | 86,6±3,9 | 89,3±6,0 | 82,2±9,7 | 85,1±4,9 |
| <i>EfficientNet B3 FS</i> | 83,3±3,1 | 84,1±5,2 | 82,2±7,1 | 82,8±3,4 |

A Figura 9.5 apresenta as curvas ROC geradas por cada modelo ao classificar as imagens do conjunto *D2*. A comparação entre modelos treinados *from scratch* e com aprendizado por transferência evidencia que os modelos pré-treinados apresentaram maior AUC em todos os casos. Entre eles, o *Inception ResNet v2* destacou-se, alcançando uma AUC de 0,95 (linha pontilhada verde).

O modelo *Inception ResNet101 v2* pré-treinado apresentou a melhor capacidade discriminativa no conjunto de dados *D2*, com uma área sob a curva (AUC) de 0,95, em comparação com 0,90 obtido pela *DenseNet201* pré-treinada. Todos os modelos pré-treinados superaram suas respectivas versões treinadas *from scratch*, evidenciando a vantagem do aprendizado por transferência para esta tarefa.

O desempenho dos patologistas também melhorou significativamente ao utilizar os resultados do classificador automatizado como suporte. A Tabela 9.3 mostra que a acurácia média dos patologistas aumentou de 88,3±10,7% para 95,0±2,6% com o auxílio da CNN.

Tabela 9.3: Desempenho de três patologistas com e sem o auxílio de um classificador automatizado na classificação de imagens do conjunto *D2*.

| Classificador | Acc.(%) | Prec.(%) | Rev.(%) | F1(%) | Fleiss' Kappa |
|------------------------|-----------------|-----------------|-----------------|-----------------|---------------|
| Patologista A | 94,0 | 94,0 | 94,1 | 94,0 | 0,59 |
| Patologista B | 91,0 | 87,0 | 96,0 | 91,0 | |
| Patologista C | 74,0 | 80,0 | 64,0 | 71,1 | |
| Média dos patologistas | 88,3±10,7 | 87,0±7,0 | 84,6±17,9 | 91,4±12,5 | - |
| CNN | 91,0 | 86,0 | 95,0 | 90,9 | - |
| Patologista A+CNN | 96,0 | 100,0 | 92,5 | 96,1 | 0,83 |
| Patologista B+CNN | 97,0 | 94,0 | 100,0 | 96,9 | |
| Patologista C+CNN | 92,0 | 86,0 | 97,7 | 91,4 | |
| Média Patologista+CNN | 95,0±2,6 | 93,3±7,0 | 96,7±3,7 | 96,1±2,9 | - |

A análise dos resultados gerais revelou que o uso do classificador automatizado impactou positivamente tanto a precisão quanto a revocação dos patologistas. Sem o suporte do modelo, a precisão média dos especialistas foi de 87,0% e a revocação de 84,6%, indicando uma tendência de maior variabilidade na identificação correta das imagens positivas. Com

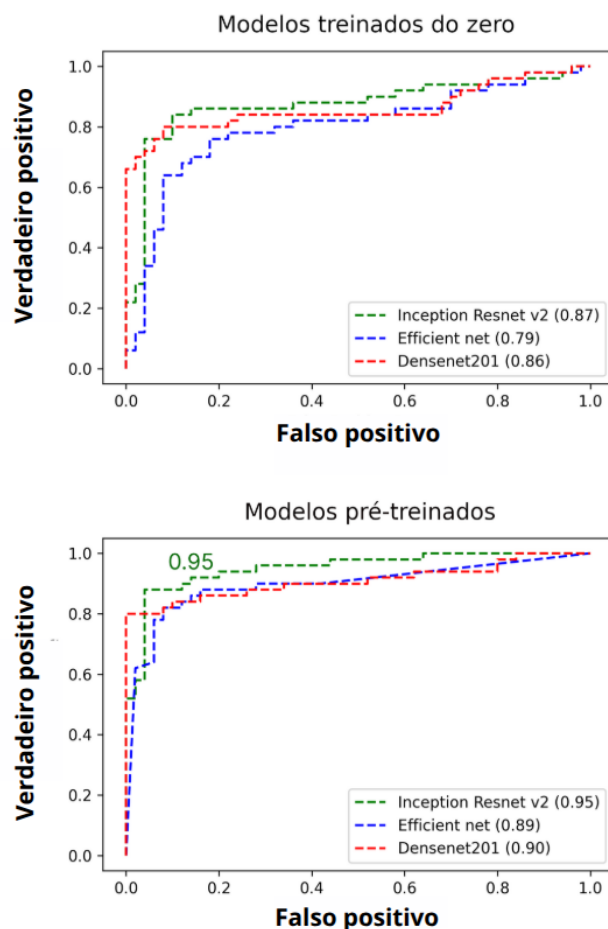


Figura 9.5: Área sob a curva (AUC) dos modelos ao classificar imagens no conjunto de dados *D2*.

a utilização do classificador, a precisão média aumentou para 93,3% e a revocação para 96,7%, demonstrando um aprimoramento significativo na capacidade de detecção das imagens com lesões podocitárias. Esses ganhos sugerem que o classificador automatizado não apenas auxiliou na redução de erros falsos positivos e falsos negativos, mas também conferiu maior estabilidade e consistência ao processo de decisão diagnóstica, o que é fundamental para elevar a qualidade da prática médica e apoiar a formação de novos especialistas.

Em termos de técnica de coloração, os resultados de acurácia observaram-se da seguinte forma:

- Para imagens coradas com PAS, a acurácia aumentou de $83,7 \pm 9\%$ para $94,5 \pm 2\%$;
- Para imagens coradas com HE, a acurácia aumentou de $80,0 \pm 17\%$ para $95,5 \pm 5\%$;
- Para imagens coradas com tricrômico, a acurácia manteve-se estável em $91,6 \pm 7\%$;

- Para imagens coradas com PAMS, houve uma leve redução, de $96,0 \pm 6\%$ para $94,6 \pm 4\%$, ainda dentro do intervalo de variação esperado.

Além disso, o grau de concordância (*Fleiss' Kappa*) entre os patologistas aumentou de 0,59 (concordância moderada) para 0,83 (concordância substancial) após a introdução do classificador automatizado, indicando uma melhoria de aproximadamente 40%.

Este resultado é particularmente relevante, considerando que os patologistas participantes pertenciam a diferentes centros, o que historicamente tende a reduzir a concordância [180, 181, 182]. A utilização de patologistas de centros distintos confere maior rigor à avaliação da eficácia do classificador automatizado como ferramenta de apoio diagnóstico.

A Figura 9.6 apresenta diagramas de *Venn* que ilustram a distribuição dos erros cometidos pelos patologistas nas análises com e sem o auxílio do classificador. Observa-se que, após a introdução do suporte automatizado, houve uma clara redistribuição dos erros entre os especialistas, indicando uma mudança na forma como as imagens foram analisadas.

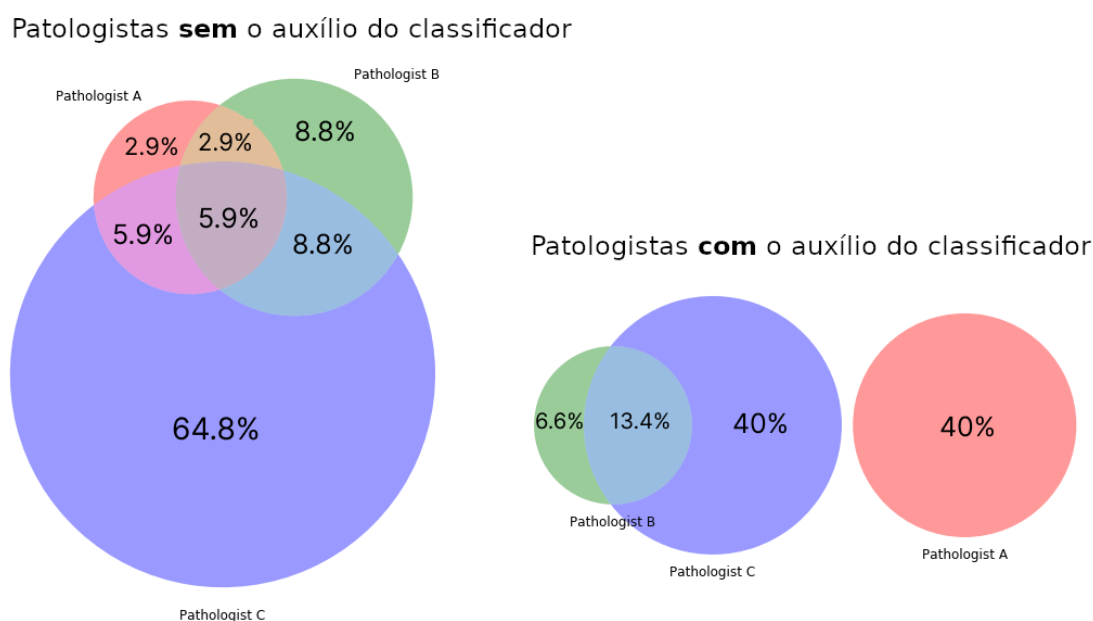


Figura 9.6: Diagramas de *Venn* ilustrando a distribuição dos erros de classificação cometidos pelos patologistas com e sem auxílio do classificador automatizado..

Os resultados obtidos evidenciam que a introdução do classificador automatizado não apenas elevou a acurácia individual dos patologistas, mas também aumentou significativamente a concordância entre os especialistas de diferentes instituições. Esses achados reforçam o potencial dos modelos de inteligência artificial como ferramentas de apoio

à prática diagnóstica, proporcionando decisões mais consistentes, seguras e precisas na análise de imagens de glomérulos renais com suspeita de podocitopatia.

9.3 Segmentação de lesões podocitárias

Considerando todas as circunstâncias desafiadoras de um conjunto de dados do mundo real — caracterizado por alta diversidade morfológica, diferentes colorações e desbalanceamento entre classes —, mesmo com a adoção de estratégias para mitigar essas dificuldades, a expectativa para esta etapa do estudo era que os modelos apresentassem desempenho inferior ao relatado em trabalhos relacionados baseados em conjuntos de dados bem controlados [28, 29, 30, 31, 32].

Todos os modelos de referência treinados no espaço de cores RGB superaram seus correspondentes treinados em HED. Portanto, considerando apenas os resultados dos 1260 modelos de referência (Seção 8.0.1) treinados em RGB, o melhor desempenho geral foi alcançado pela arquitetura *Unet++* com o codificador *Inception v4*, inicializada com pesos pré-treinados no *ImageNet*. Este modelo obteve, no conjunto de teste, um *F1-score* geral de 30,31%, com 39,53% na classe "lesão", 15,88% na classe "hiperplasia" e falhou em segmentar as classes "hipertrofia" e "degeneração". Esse comportamento de dificuldade em identificar todas as classes foi semelhante ao observado na maioria dos modelos de referência avaliados.

Mesmo entre modelos capazes de gerar previsões para todas as classes, o desempenho permaneceu baixo. Por exemplo, o *DeepLab v3+* com codificador *EfficientNet B0* pré-treinado no *ImageNet* e treinado em RGB alcançou um *F1-score* geral de 29,97%, com 31,94% para "sem lesão", 5,70% para "hipertrofia", 3,62% para "degeneração" e 14,39% para "hiperplasia".

De maneira geral, ficou evidente que as estratégias de aumento de dados (*data augmentation*) e ponderação de classes no cálculo da função de perda (*loss weighting*) não foram suficientes para superar os desafios impostos por um conjunto de dados tão diverso.

Para uma avaliação mais detalhada, foi realizada uma análise comparativa entre os 10% melhores e 10% piores modelos (ver Figura 9.7). Entre os melhores modelos, observou-se maior incidência das arquiteturas *U-Net* e *U-Net++*, presentes em 36% dos casos, indicando a robustez dessas arquiteturas para imagens histológicas. Entre os codificadores mais frequentes nesse grupo destacaram-se o *Timm RegNetX*, *Timm RegNetY*, *EfficientNet B0* e *EfficientNet B2*, aparecendo juntos em 66,66% dos modelos.

No grupo dos 10% piores modelos, apenas quatro arquiteturas se fizeram presentes: *LinkNet*, *FPN*, *PAN* e *PSPNet*. Os codificadores com maior incidência nesse grupo foram *EfficientNet B0* e *EfficientNet B1*, responsáveis por 58,8% dos modelos. Esses codifica-

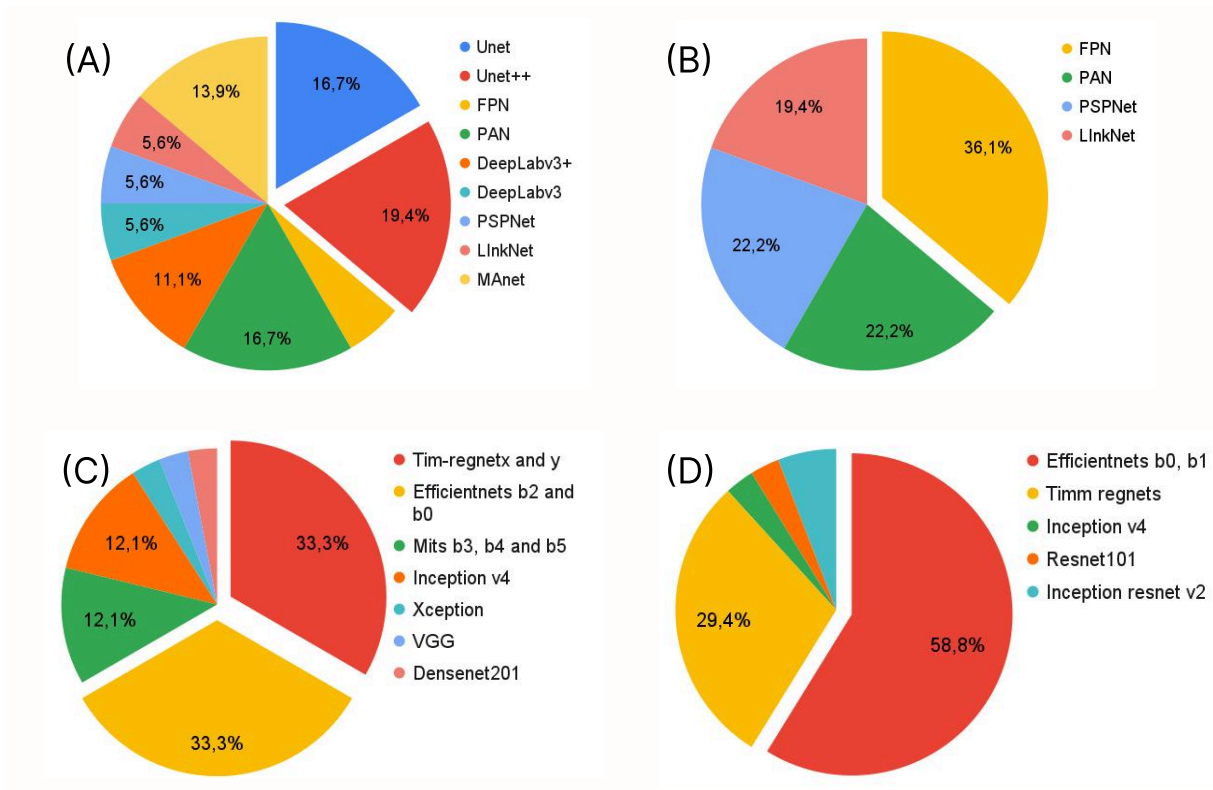


Figura 9.7: Distribuição das arquiteturas de rede (A, B) e codificadores (C, D) entre os 10% melhores e piores modelos de referência..

dores possuem menor custo computacional, o que pode ter impactado negativamente o desempenho.

Outro ponto relevante é que todos os melhores modelos de referência foram obtidos via *fine-tuning* de pesos pré-treinados do *ImageNet*, enquanto, entre os piores *baselines*, 40% dos modelos foram treinados *from scratch* (com inicialização aleatória de pesos).

9.3.1 Impacto da Solução Proposta

Para a avaliação da solução proposta (Seção 8.0.2), foram retreinados 30 modelos de referência: os 10 melhores, os 10 piores e 10 selecionados aleatoriamente. Após o retreinamento, observou-se uma melhoria substancial: 29 dos 30 modelos apresentaram avanços em termos de *Intersection over Union* (IoU) [120] e *F1-score* [119] geral em todas as classes. Mesmo o modelo inicialmente de pior desempenho foi capaz de prever classes anteriormente não identificáveis.

A Figura 9.8 apresenta um *boxplot* comparativo das melhorias de *F1-score* obtidas com a aplicação das diferentes estratégias de fusão propostas.

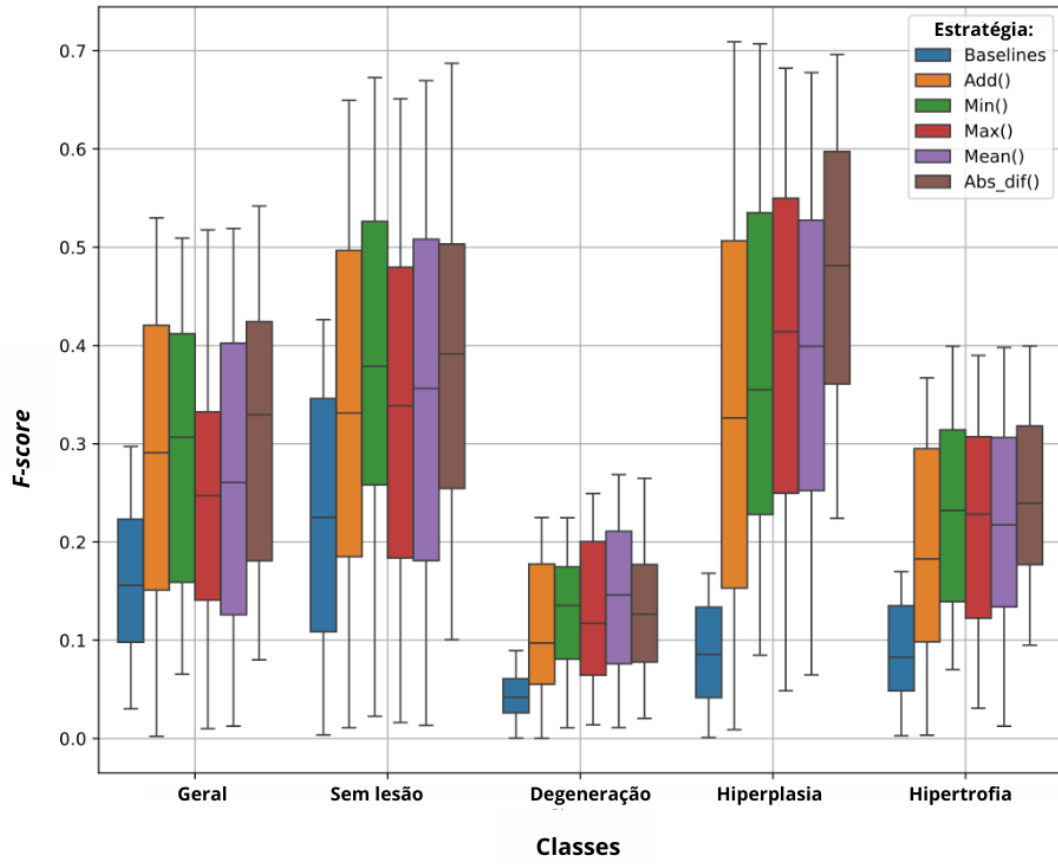


Figura 9.8: *Boxplot* das diferenças no $F1$ -score entre os modelos de referência (*Baselines*) e a solução proposta utilizando diferentes estratégias de fusão.

O modelo com melhor desempenho após o retreinamento foi o *FPN* com codificador *Timm-RegNetY 120*, que, utilizando a operação de fusão média, elevou seu $F1$ -score de 27,17% para 51,88%. Na classe "sem lesão", a acurácia passou de 40,21% para 67,03%; na classe "hiperplasia", de 15,13% para 64,94%; enquanto nas classes "hipertrofia" e "degeneração" — onde antes não havia predições — atingiu 36,84% e 26,98%, respectivamente.

Entre os 30 modelos reavaliados, observou-se variação nas operações de fusão que proporcionaram os melhores resultados. Como ilustrado na Figura 9.8, a operação de diferença absoluta obteve os melhores resultados gerais e nas classes "hipertrofia" e "degeneração". Para a classe "hiperplasia", as operações de mínimo e média se destacaram, enquanto para "lesão" a operação de máximo apresentou melhor desempenho.

As Tabelas 9.4, 9.5 e 9.6 resumem os valores de IoU e $F1$ -score obtidos para os grupos *Top-10*, *Random-10* e *Bottom-10*.

De forma geral, ao comparar o desempenho dos modelos de referência com o desempenho após a aplicação da solução proposta, observou-se um aumento médio no IoU de

18,25% para 34,20% e no $F1$ -score de 21,60% para 39,00%. Esses resultados confirmam a efetividade da abordagem multicanal baseada em fusão de características no contexto de segmentação multiclasse de lesões podocitárias.

A Figura 9.9 exemplifica os resultados visuais das predições realizadas pelo melhor modelo de referência — *Unet++* com *Timm-RegNetX* — retreinado utilizando a solução FFN proposta. Na amostra *a*, observa-se uma segmentação parcial da classe hiperplasia, com IoU de 0,56, além da ausência de predições para as classes degeneração e hipertrofia, que estavam presentes na máscara de rótulos fornecida pelos patologistas. Na amostra *b*, destaca-se um excelente desempenho na predição da classe hipertrofia (IoU de 0,81) e uma predição moderada de podócitos sem lesão (IoU de 0,41), sem predição para a classe de podócitos com hiperplasia. Por fim, a amostra *c* apresenta um IoU de 0,60 na predição da classe podócitos sem lesão.

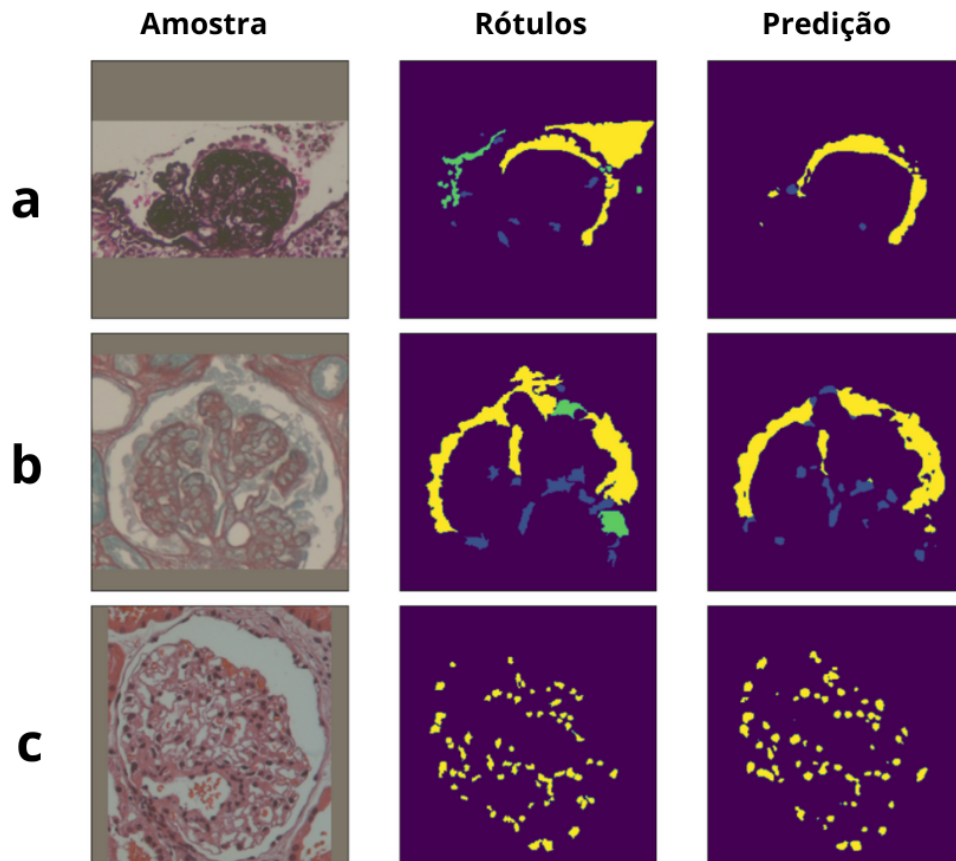


Figura 9.9: Predições do melhor modelo obtido com a adoção da solução proposta (*Unet++* com *Timm-RegNetX*) em três amostras do conjunto de teste..

Tabela 9.4: *IoU* e *F1-score* do grupo *Top-10* (%)

| Modelo | Fusão | Geral | | Sem lesão | | Hipertrofia | | Degeneração | | Hiperplasia | |
|------------------------|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. |
| Unet++ timm-regnetx320 | Add | 5,27 | 5,39 | 1,90 | 3,72 | 5,18 | 9,64 | 6,92 | 12,00 | 11,72 | 20,57 |
| | Mean | 21,11 | 22,36 | 28,73 | 44,59 | 19,16 | 31,18 | 1,47 | 2,80 | 24,65 | 38,88 |
| | Min | 7,41 | 7,77 | 10,29 | 18,62 | 4,54 | 8,46 | 0,77 | 1,49 | 4,59 | 8,48 |
| | Max | 14,38 | 15,03 | 19,62 | 32,72 | 14,67 | 24,36 | 1,30 | 2,52 | 7,30 | 13,24 |
| | AbsDif | 22,85 | 24,36 | 24,34 | 38,96 | 21,87 | 34,74 | 4,42 | 7,79 | 25,97 | 39,42 |
| | Baseline | 23,62 | 27,06 | 26,00 | 39,45 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| FPN timm-regnety120 | Add | 24,21 | 27,37 | 25,70 | 40,73 | 7,64 | 13,90 | 11,07 | 19,45 | 30,95 | 46,32 |
| | Mean | 44,54 | 51,88 | 50,69 | 67,03 | 23,42 | 36,84 | 16,18 | 26,98 | 48,71 | 64,94 |
| | Min | 34,42 | 37,79 | 35,79 | 52,44 | 18,18 | 28,38 | 8,44 | 14,19 | 44,01 | 59,03 |
| | Max | 43,75 | 51,34 | 49,19 | 65,65 | 23,12 | 36,52 | 14,33 | 23,88 | 49,84 | 66,06 |
| | AbsDif | 34,70 | 41,65 | 37,50 | 54,42 | 22,48 | 35,84 | 16,27 | 27,35 | 41,00 | 57,53 |
| | Baseline | 23,60 | 27,17 | 26,29 | 40,21 | 0,00 | 0,00 | 0,00 | 0,00 | 13,01 | 15,13 |
| Unet++ densenet201 | Add | 6,76 | 6,89 | 3,68 | 7,10 | 5,24 | 9,68 | 2,65 | 5,07 | 15,73 | 26,42 |
| | Mean | 9,30 | 9,70 | 12,13 | 21,59 | 5,02 | 9,30 | 3,84 | 7,25 | 7,28 | 13,14 |
| | Min | 17,30 | 17,70 | 22,99 | 37,21 | 11,46 | 18,98 | 4,28 | 8,02 | 9,49 | 16,42 |
| | Max | 16,54 | 17,51 | 19,57 | 32,62 | 18,50 | 8,50 | 4,68 | 8,61 | 15,98 | 26,65 |
| | AbsDif | 26,29 | 28,73 | 24,63 | 39,45 | 20,31 | 31,89 | 3,66 | 6,76 | 38,30 | 53,86 |
| | Baseline | 23,31 | 27,93 | 16,12 | 26,34 | 3,66 | 5,23 | 3,00 | 3,62 | 9,00 | 12,00 |
| PAN timm-regnety120 | Add | 13,91 | 14,58 | 9,28 | 16,91 | 16,25 | 26,55 | 5,11 | 9,04 | 25,62 | 38,96 |
| | Mean | 25,21 | 27,27 | 21,93 | 35,85 | 17,39 | 27,50 | 9,06 | 15,80 | 39,78 | 55,24 |
| | Min | 18,51 | 19,35 | 15,45 | 26,44 | 17,51 | 27,46 | 6,40 | 10,70 | 28,88 | 42,94 |
| | Max | 32,31 | 36,13 | 33,91 | 50,43 | 18,40 | 29,66 | 9,79 | 16,97 | 41,26 | 56,65 |
| | AbsDif | 8,33 | 8,71 | 6,72 | 12,51 | 9,12 | 15,98 | 3,46 | 6,31 | 12,93 | 22,53 |
| | Baseline | 25,93 | 30,31 | 24,86 | 38,52 | 0,00 | 0,00 | 0,00 | 0,00 | 12,78 | 17,24 |
| Unet++ inceptionv4 | Add | 5,01 | 5,08 | 6,36 | 11,91 | 4,42 | 8,31 | 1,10 | 2,15 | 3,72 | 7,10 |
| | Mean | 28,49 | 30,92 | 36,31 | 52,91 | 10,71 | 18,74 | 4,04 | 7,47 | 24,65 | 38,88 |
| | Min | 16,69 | 17,84 | 18,34 | 30,66 | 13,61 | 23,10 | 3,42 | 6,43 | 18,10 | 29,61 |
| | Max | 3,11 | 3,17 | 1,77 | 3,47 | 13,39 | 22,41 | 2,18 | 4,41 | 2,96 | 5,70 |
| | AbsDif | 9,05 | 9,40 | 5,53 | 10,47 | 14,18 | 23,47 | 4,60 | 7,50 | 16,31 | 27,04 |
| | Baseline | 26,70 | 30,31 | 25,94 | 39,53 | 0,00 | 0,00 | 0,00 | 0,00 | 11,78 | 15,78 |
| FPN efficientnet-b2 | Add | 41,21 | 48,11 | 45,96 | 62,77 | 21,13 | 32,88 | 11,01 | 18,92 | 48,84 | 64,77 |
| | Mean | 42,86 | 49,71 | 47,48 | 64,21 | 20,98 | 32,73 | 12,49 | 21,14 | 51,45 | 67,16 |
| | Min | 38,36 | 43,19 | 41,60 | 58,28 | 19,52 | 30,31 | 10,80 | 18,42 | 46,13 | 61,60 |
| | Max | 44,30 | 52,40 | 50,68 | 67,02 | 20,35 | 32,09 | 13,09 | 22,11 | 49,96 | 64,50 |
| | AbsDif | 43,84 | 52,15 | 51,94 | 68,16 | 19,64 | 31,22 | 13,99 | 23,93 | 46,37 | 62,40 |
| | Baseline | 25,68 | 30,33 | 21,37 | 33,91 | 5,22 | 7,51 | 0,00 | 0,00 | 10,92 | 15,06 |
| Unet inception v4 | Add | 10,56 | 11,00 | 10,88 | 19,55 | 16,97 | 27,68 | 3,73 | 7,04 | 10,27 | 18,39 |
| | Mean | 8,50 | 8,83 | 6,54 | 12,25 | 5,13 | 9,61 | 2,52 | 4,86 | 16,08 | 27,27 |
| | Min | 21,80 | 23,17 | 17,32 | 29,27 | 17,22 | 27,69 | 5,03 | 8,68 | 36,70 | 52,51 |
| | Max | 11,25 | 11,56 | 11,42 | 20,30 | 15,39 | 25,54 | 3,40 | 5,70 | 11,40 | 20,33 |
| | AbsDif | 23,80 | 25,84 | 18,78 | 31,52 | 21,39 | 34,05 | 9,73 | 16,77 | 38,98 | 55,08 |
| | Baseline | 24,36 | 27,97 | 29,56 | 43,95 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Unet++ densenet169 | Add | 0,24 | 0,24 | 0,00 | 0,00 | 0,25 | 0,49 | 0,41 | 0,81 | 0,70 | 0,14 |
| | Mean | 22,98 | 23,13 | 1,24 | 2,44 | 7,50 | 12,55 | 0,80 | 1,50 | 3,90 | 7,30 |
| | Min | 6,98 | 7,80 | 5,00 | 9,00 | 4,42 | 7,83 | 2,13 | 4,00 | 13,93 | 23,70 |
| | Max | 24,01 | 25,79 | 23,11 | 37,20 | 16,11 | 26,70 | 5,73 | 10,20 | 34,19 | 49,67 |
| | AbsDif | 33,10 | 33,73 | 37,32 | 53,99 | 19,83 | 31,90 | 7,20 | 12,63 | 36,86 | 51,69 |
| | Baseline | 24,72 | 29,20 | 19,69 | 31,57 | 4,80 | 6,20 | 0,00 | 0,00 | 9,99 | 13,81 |
| Unet++ vgg19 | Add | 6,15 | 6,24 | 4,73 | 8,95 | 3,10 | 5,89 | 0,32 | 0,64 | 12,34 | 21,21 |
| | Mean | 20,67 | 21,90 | 22,19 | 36,17 | 15,33 | 25,96 | 0,50 | 0,98 | 25,96 | 40,39 |
| | Min | 21,80 | 23,17 | 17,32 | 29,27 | 17,22 | 27,69 | 5,03 | 8,68 | 36,70 | 52,51 |
| | Max | 11,25 | 11,56 | 11,42 | 20,30 | 15,39 | 25,54 | 3,40 | 5,70 | 11,40 | 20,33 |
| | AbsDif | 23,80 | 25,84 | 18,78 | 31,52 | 21,39 | 34,05 | 9,73 | 16,77 | 38,98 | 55,08 |
| | Baseline | 24,36 | 27,97 | 29,56 | 43,95 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Manet mitb3 | Add | 42,17 | 48,22 | 46,87 | 63,37 | 20,62 | 33,19 | 10,42 | 18,01 | 49,93 | 65,98 |
| | Mean | 41,73 | 47,78 | 45,30 | 62,25 | 23,35 | 40,60 | 8,92 | 15,67 | 49,32 | 65,49 |
| | Min | 32,25 | 37,78 | 33,57 | 49,95 | 24,83 | 38,77 | 12,96 | 22,35 | 38,34 | 54,91 |
| | Max | 32,66 | 37,57 | 32,33 | 48,70 | 17,25 | 28,60 | 13,60 | 22,94 | 43,76 | 60,04 |
| | AbsDif | 38,40 | 43,81 | 36,33 | 53,01 | 26,44 | 40,27 | 12,13 | 20,78 | 54,66 | 70,03 |
| | Baseline | 24,83 | 28,95 | 22,39 | 35,41 | 0,00 | 0,00 | 0,00 | 0,00 | 10,00 | 13,67 |

Tabela 9.5: *IoU* e *F1-score* do grupo *Random-10* (%)

| Modelo | Fusão | Geral | | Sem lesão | | Hipertrofia | | Degeneração | | Hiperplasia | |
|--------------------------|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. |
| Unet mitb3 | Add | 26,80 | 29,76 | 32,11 | 48,45 | 9,03 | 14,85 | 5,54 | 9,62 | 27,38 | 40,87 |
| | Mean | 30,36 | 34,47 | 35,34 | 51,93 | 12,50 | 20,25 | 8,55 | 14,47 | 31,86 | 46,03 |
| | Min | 19,43 | 20,59 | 17,62 | 29,85 | 9,44 | 15,52 | 4,60 | 8,12 | 30,75 | 45,54 |
| | Max | 21,43 | 23,09 | 22,32 | 36,37 | 10,72 | 17,05 | 6,75 | 11,78 | 27,38 | 40,77 |
| | AbsDif | 19,71 | 21,82 | 21,30 | 35,01 | 11,98 | 19,81 | 6,29 | 11,30 | 23,95 | 36,54 |
| | Baseline | 20,33 | 23,49 | 14,85 | 24,62 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Unet++ efficientnet b4 | Add | 46,60 | 53,95 | 49,46 | 65,98 | 23,90 | 36,43 | 12,19 | 20,59 | 57,79 | 72,39 |
| | Mean | 44,13 | 50,49 | 47,61 | 64,16 | 25,99 | 40,25 | 11,33 | 19,33 | 53,09 | 62,39 |
| | Min | 35,55 | 39,05 | 34,90 | 51,39 | 22,70 | 35,60 | 10,26 | 17,56 | 49,21 | 64,92 |
| | Max | 41,81 | 47,80 | 42,57 | 59,47 | 25,23 | 38,65 | 14,45 | 24,44 | 54,30 | 69,57 |
| | AbsDif | 47,34 | 55,12 | 52,43 | 68,63 | 26,66 | 40,80 | 14,11 | 23,90 | 55,03 | 70,00 |
| | Baseline | 24,20 | 27,84 | 29,21 | 43,53 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| PSPnet timm-regnety 080 | Add | 0,21 | 0,21 | 0,70 | 0,15 | 0,10 | 0,30 | 0,17 | 0,34 | 0,56 | 1,10 |
| | Mean | 8,37 | 8,60 | 6,30 | 11,83 | 6,11 | 11,09 | 1,50 | 2,86 | 15,77 | 26,45 |
| | Min | 12,53 | 13,19 | 8,77 | 16,00 | 11,00 | 18,28 | 2,93 | 4,84 | 23,54 | 36,19 |
| | Max | 1,81 | 1,91 | 1,80 | 3,53 | 3,72 | 7,00 | 0,43 | 0,84 | 2,15 | 4,17 |
| | AbsDif | 11,73 | 12,66 | 13,42 | 21,89 | 8,93 | 15,36 | 1,00 | 2,00 | 13,44 | 23,13 |
| | Baseline | 22,60 | 25,51 | 13,56 | 23,48 | 2,41 | 3,28 | 0,00 | 0,00 | 4,99 | 6,58 |
| MAnet vgg19 | Add | 33,51 | 40,36 | 42,45 | 59,32 | 12,24 | 20,95 | 13,41 | 23,15 | 29,43 | 43,60 |
| | Mean | 31,31 | 37,28 | 39,39 | 56,13 | 13,39 | 22,29 | 12,26 | 21,13 | 25,94 | 39,39 |
| | Min | 38,40 | 45,16 | 48,18 | 64,63 | 12,69 | 20,75 | 6,43 | 11,83 | 33,79 | 48,64 |
| | Max | 38,85 | 46,83 | 49,54 | 66,12 | 12,31 | 21,12 | 12,35 | 21,29 | 35,15 | 50,32 |
| | AbsDif | 40,88 | 50,41 | 51,88 | 67,31 | 10,36 | 17,98 | 11,23 | 19,63 | 38,05 | 52,72 |
| | Baseline | 20,15 | 22,43 | 10,22 | 17,15 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Manet mitb5 | Add | 0,49 | 0,49 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| | Mean | 41,27 | 48,26 | 45,12 | 61,89 | 24,65 | 36,29 | 11,90 | 20,27 | 47,20 | 62,38 |
| | Min | 9,74 | 10,00 | 7,58 | 14,00 | 11,27 | 19,62 | 4,27 | 8,08 | 15,20 | 25,80 |
| | Max | 30,35 | 33,90 | 27,76 | 43,24 | 24,41 | 37,85 | 14,87 | 25,16 | 41,99 | 58,60 |
| | AbsDif | 30,37 | 34,39 | 35,84 | 52,50 | 19,91 | 32,31 | 9,30 | 16,61 | 29,67 | 45,11 |
| | Baseline | 24,27 | 27,86 | 29,13 | 43,38 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Manet mitb4 | Add | 42,76 | 48,87 | 46,23 | 62,94 | 23,51 | 36,11 | 12,73 | 21,79 | 52,36 | 67,76 |
| | Mean | 41,64 | 48,28 | 42,95 | 59,91 | 24,22 | 37,47 | 12,03 | 20,44 | 54,53 | 69,61 |
| | Min | 40,48 | 45,20 | 43,17 | 60,00 | 23,25 | 35,72 | 13,07 | 22,30 | 50,25 | 65,87 |
| | Max | 41,91 | 47,57 | 45,52 | 62,29 | 20,81 | 33,60 | 9,68 | 16,99 | 50,50 | 66,00 |
| | AbsDif | 37,05 | 42,02 | 36,67 | 53,19 | 20,65 | 33,27 | 12,97 | 21,90 | 50,38 | 66,06 |
| | Baseline | 24,59 | 28,69 | 21,74 | 34,36 | 0,00 | 0,00 | 0,00 | 0,00 | 9,71 | 13,57 |
| Linknet timm-regnety 120 | Add | 34,39 | 38,34 | 32,22 | 48,21 | 24,12 | 37,34 | 10,00 | 17,13 | 49,41 | 65,24 |
| | Mean | 28,79 | 31,70 | 23,59 | 37,84 | 21,64 | 34,54 | 10,46 | 18,09 | 45,23 | 61,17 |
| | Min | 43,58 | 50,04 | 44,52 | 61,53 | 26,60 | 40,28 | 14,32 | 23,17 | 55,46 | 70,75 |
| | Max | 31,59 | 35,93 | 38,25 | 55,09 | 15,38 | 25,88 | 8,07 | 14,35 | 27,77 | 42,41 |
| | AbsDif | 37,14 | 42,47 | 38,84 | 55,80 | 21,53 | 34,40 | 12,91 | 21,96 | 48,01 | 63,92 |
| | Baseline | 24,83 | 28,95 | 22,39 | 35,41 | 0,00 | 0,00 | 0,00 | 0,00 | 9,99 | 13,67 |
| Linknet efficientnet b1 | Add | 41,97 | 48,30 | 49,46 | 65,91 | 18,00 | 28,19 | 9,17 | 15,56 | 41,26 | 56,66 |
| | Mean | 31,69 | 34,59 | 36,59 | 53,42 | 18,11 | 28,49 | 7,13 | 12,03 | 31,49 | 45,54 |
| | Min | 39,39 | 45,07 | 46,73 | 63,53 | 20,55 | 32,82 | 9,32 | 15,90 | 38,62 | 53,82 |
| | Max | 34,67 | 38,12 | 38,65 | 55,61 | 21,83 | 34,25 | 6,94 | 11,94 | 38,71 | 53,44 |
| | AbsDif | 10,01 | 10,40 | 9,00 | 16,63 | 8,97 | 16,21 | 2,52 | 4,85 | 15,77 | 26,60 |
| | Baseline | 23,78 | 27,79 | 22,68 | 35,17 | 0,00 | 0,00 | 5,22 | 8,53 | 0,00 | 0,00 |
| Unet timmregnety 080 | Add | 19,76 | 22,06 | 18,22 | 30,65 | 15,61 | 24,89 | 7,11 | 12,84 | 29,33 | 43,52 |
| | Mean | 29,74 | 32,83 | 32,37 | 48,67 | 15,38 | 23,92 | 6,18 | 10,45 | 36,46 | 51,46 |
| | Min | 31,28 | 36,22 | 39,11 | 55,63 | 8,13 | 12,84 | 5,85 | 9,90 | 28,57 | 41,86 |
| | Max | 35,27 | 39,87 | 46,38 | 63,22 | 4,35 | 7,88 | 7,72 | 13,52 | 29,52 | 43,30 |
| | AbsDif | 25,92 | 28,56 | 27,33 | 42,77 | 14,21 | 22,93 | 7,92 | 13,83 | 34,54 | 49,22 |
| | Baseline | 19,23 | 22,13 | 8,83 | 15,75 | 0,00 | 0,00 | 0,00 | 0,00 | 1,87 | 2,85 |
| DeepLab v3 resnet50 | Add | 3,10 | 3,13 | 0,70 | 1,38 | 0,77 | 1,44 | 0,34 | 0,65 | 9,81 | 16,63 |
| | Mean | 0,90 | 0,90 | 0,20 | 0,40 | 0,10 | 0,20 | 0,28 | 0,56 | 2,97 | 5,60 |
| | Min | 6,50 | 6,61 | 0,57 | 1,13 | 4,18 | 7,39 | 0,96 | 1,85 | 21,41 | 33,04 |
| | Max | 1,57 | 1,57 | 0,24 | 0,48 | 1,34 | 2,57 | 1,38 | 2,59 | 4,75 | 8,77 |
| | AbsDif | 22,98 | 24,56 | 26,60 | 41,84 | 10,94 | 17,41 | 5,19 | 9,24 | 25,00 | 39,21 |
| | Baseline | 19,88 | 22,16 | 9,65 | 16,25 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |

Tabela 9.6: *IoU* e *F1-score* do grupo *Bottom-10* (%)

| Modelo | Fusão | Geral | | Sem lesão | | Hipertrofia | | Degeneração | | Hiperplasia | |
|-------------------------|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. | IoU | F-sc. |
| PAN timm-regnetx 120 | Add | 17,18 | 18,01 | 16,09 | 27,51 | 13,86 | 22,46 | 6,73 | 11,99 | 23,62 | 36,80 |
| | Mean | 22,12 | 23,52 | 22,77 | 36,93 | 18,94 | 30,21 | 5,36 | 9,37 | 27,18 | 40,93 |
| | Min | 10,48 | 10,73 | 8,60 | 15,70 | 11,26 | 18,56 | 3,72 | 6,59 | 16,97 | 27,87 |
| | Max | 7,55 | 7,70 | 7,86 | 14,53 | 5,28 | 9,31 | 0,80 | 1,56 | 10,25 | 17,88 |
| | AbsDif | 17,35 | 18,81 | 20,28 | 33,56 | 5,27 | 9,61 | 7,30 | 13,32 | 18,71 | 30,89 |
| | Baseline | 3,92 | 6,29 | 15,14 | 24,91 | 5,00 | 10,00 | 0,00 | 0,00 | 3,95 | 5,51 |
| FPN resnet50 | Add | 28,68 | 32,06 | 37,71 | 54,56 | 13,09 | 21,18 | 8,00 | 13,96 | 20,28 | 31,25 |
| | Mean | 28,10 | 31,76 | 35,97 | 52,75 | 9,35 | 15,60 | 6,74 | 11,85 | 24,37 | 37,92 |
| | Min | 26,12 | 29,35 | 26,74 | 41,97 | 14,68 | 23,03 | 13,11 | 22,54 | 32,56 | 47,08 |
| | Max | 30,04 | 33,69 | 39,24 | 56,12 | 10,50 | 17,67 | 7,00 | 12,26 | 22,75 | 35,32 |
| | AbsDif | 29,44 | 34,38 | 36,82 | 53,52 | 11,75 | 20,17 | 8,58 | 15,34 | 27,04 | 41,60 |
| | Baseline | 2,38 | 4,20 | 0,00 | 0,00 | 10,40 | 17,50 | 0,00 | 0,00 | 10,91 | 3,56 |
| PSPNet efficientnet b4 | Add | 11,80 | 13,47 | 11,38 | 20,09 | 4,50 | 7,94 | 5,68 | 10,34 | 15,98 | 26,09 |
| | Mean | 15,35 | 17,37 | 19,74 | 32,45 | 4,93 | 8,43 | 1,76 | 3,29 | 13,76 | 20,58 |
| | Min | 35,43 | 42,63 | 44,83 | 61,28 | 16,18 | 25,34 | 7,66 | 13,39 | 32,17 | 46,16 |
| | Max | 31,90 | 37,93 | 41,91 | 58,53 | 6,08 | 10,67 | 6,36 | 11,46 | 25,76 | 38,80 |
| | AbsDif | 27,58 | 33,52 | 37,15 | 53,60 | 9,87 | 16,20 | 7,45 | 13,30 | 18,74 | 29,24 |
| | Baseline | 2,36 | 4,10 | 0,00 | 0,00 | 9,88 | 16,95 | 0,00 | 0,00 | 1,91 | 3,56 |
| PAN efficientnet b4 | Add | 32,50 | 36,80 | 33,77 | 50,35 | 17,09 | 26,87 | 8,44 | 14,66 | 42,82 | 58,36 |
| | Mean | 34,46 | 38,89 | 36,55 | 53,30 | 17,67 | 27,92 | 8,34 | 14,17 | 44,74 | 59,86 |
| | Min | 26,10 | 28,15 | 29,07 | 44,61 | 13,69 | 21,46 | 4,73 | 7,99 | 29,62 | 43,00 |
| | Max | 29,30 | 33,46 | 32,35 | 48,73 | 14,23 | 23,06 | 5,86 | 10,64 | 35,91 | 50,93 |
| | AbsDif | 27,48 | 31,69 | 31,47 | 47,66 | 14,26 | 22,28 | 7,63 | 13,36 | 30,73 | 45,05 |
| | Baseline | 1,76 | 3,20 | 6,48 | 11,67 | 0,00 | 0,00 | 2,40 | 4,50 | 2,10 | 3,80 |
| FPN timm-regnetx 080 | Add | 20,57 | 22,73 | 37,71 | 54,56 | 13,09 | 21,18 | 8,00 | 13,96 | 20,28 | 31,25 |
| | Mean | 21,56 | 23,75 | 35,97 | 52,75 | 9,35 | 15,60 | 6,76 | 11,85 | 24,37 | 37,92 |
| | Min | 27,94 | 31,93 | 26,74 | 41,97 | 14,68 | 23,03 | 13,11 | 22,54 | 32,56 | 47,08 |
| | Max | 22,08 | 24,63 | 39,24 | 56,12 | 10,50 | 17,67 | 7,00 | 12,26 | 22,75 | 35,32 |
| | AbsDif | 28,92 | 34,70 | 36,82 | 53,52 | 11,75 | 20,17 | 8,58 | 15,34 | 27,04 | 41,60 |
| | Baseline | 20,70 | 24,40 | 16,46 | 27,17 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| PAN efficientnet b2 | Add | 31,13 | 35,86 | 38,53 | 55,25 | 14,08 | 22,34 | 6,07 | 10,93 | 29,34 | 43,84 |
| | Mean | 19,50 | 21,38 | 22,70 | 36,83 | 8,32 | 14,09 | 3,72 | 6,68 | 22,06 | 33,97 |
| | Min | 34,16 | 40,41 | 43,04 | 59,93 | 13,21 | 21,40 | 10,00 | 17,38 | 29,76 | 44,17 |
| | Max | 33,07 | 38,08 | 40,30 | 57,05 | 15,48 | 24,55 | 10,97 | 18,69 | 29,43 | 43,53 |
| | AbsDif | 39,87 | 47,12 | 53,16 | 69,14 | 10,16 | 16,73 | 3,99 | 6,88 | 28,20 | 42,91 |
| | Baseline | 2,28 | 4,10 | 8,88 | 15,90 | 0,00 | 0,00 | 0,26 | 0,51 | 2,33 | 4,70 |
| FPN timm-regnetx 080 | Add | 21,33 | 23,40 | 27,74 | 43,31 | 7,98 | 13,73 | 3,63 | 6,46 | 16,60 | 26,85 |
| | Mean | 18,83 | 20,45 | 21,74 | 35,57 | 6,14 | 10,70 | 3,66 | 6,49 | 21,53 | 34,12 |
| | Min | 13,22 | 13,89 | 14,56 | 25,35 | 3,87 | 6,99 | 1,40 | 2,84 | 17,27 | 27,82 |
| | Max | 26,91 | 30,58 | 35,14 | 51,85 | 8,40 | 14,04 | 5,67 | 10,07 | 20,98 | 32,95 |
| | AbsDif | 21,76 | 26,46 | 25,40 | 40,43 | 6,30 | 11,43 | 3,23 | 6,00 | 25,38 | 39,32 |
| | Baseline | 1,86 | 3,35 | 4,30 | 8,00 | 0,00 | 0,00 | 1,32 | 2,37 | 3,69 | 6,27 |
| Linknet efficientnet b2 | Add | 26,49 | 30,17 | 33,12 | 49,59 | 12,61 | 19,26 | 4,00 | 7,42 | 23,96 | 36,90 |
| | Mean | 24,66 | 26,53 | 23,19 | 37,35 | 11,82 | 18,61 | 4,24 | 7,32 | 37,61 | 52,59 |
| | Min | 43,71 | 51,36 | 52,93 | 68,78 | 18,29 | 28,94 | 13,60 | 23,09 | 46,24 | 61,29 |
| | Max | 11,28 | 12,32 | 13,33 | 23,25 | 11,04 | 16,75 | 2,90 | 5,50 | 11,11 | 19,20 |
| | AbsDif | 25,35 | 28,64 | 27,30 | 42,79 | 15,32 | 24,20 | 8,88 | 15,44 | 32,31 | 46,23 |
| | Baseline | 17,67 | 22,15 | 8,20 | 14,67 | 2,40 | 3,77 | 1,24 | 1,92 | 5,57 | 7,82 |
| FPN efficientnet b5 | Add | 33,46 | 38,58 | 39,64 | 56,46 | 15,69 | 25,28 | 7,90 | 13,67 | 33,56 | 47,61 |
| | Mean | 30,22 | 34,59 | 35,63 | 52,30 | 13,94 | 22,59 | 9,60 | 15,97 | 29,19 | 43,14 |
| | Min | 34,46 | 39,57 | 42,99 | 59,99 | 15,60 | 24,64 | 9,06 | 15,53 | 31,44 | 45,27 |
| | Max | 35,15 | 40,94 | 42,30 | 58,93 | 16,45 | 26,25 | 9,33 | 16,02 | 34,11 | 48,68 |
| | AbsDif | 26,15 | 32,50 | 30,60 | 46,74 | 9,04 | 16,31 | 8,96 | 16,16 | 30,00 | 44,88 |
| | Baseline | 12,48 | 14,41 | 8,67 | 15,66 | 0,29 | 0,57 | 0,00 | 0,00 | 3,42 | 5,84 |
| Linknet efficientnet b5 | Add | 27,68 | 30,49 | 33,76 | 50,31 | 14,17 | 23,74 | 2,40 | 4,44 | 27,77 | 41,98 |
| | Mean | 34,06 | 38,42 | 40,50 | 57,45 | 14,05 | 22,86 | 7,14 | 12,53 | 35,91 | 51,78 |
| | Min | 37,35 | 42,66 | 43,15 | 60,00 | 16,55 | 26,63 | 6,86 | 11,98 | 42,13 | 57,53 |
| | Max | 36,46 | 42,43 | 41,24 | 58,12 | 16,40 | 26,47 | 8,89 | 15,45 | 43,12 | 58,83 |
| | AbsDif | 28,34 | 31,90 | 34,26 | 50,88 | 11,42 | 19,22 | 5,21 | 9,15 | 28,48 | 42,94 |
| | Baseline | 10,29 | 15,75 | 6,38 | 11,79 | 0,44 | 0,87 | 1,80 | 3,25 | 5,20 | 8,30 |

Capítulo 10

Considerações Finais

Esta tese teve como objetivo ampliar a fronteira da Patologia Computacional por meio do desenvolvimento de técnicas computacionais para auxiliar na análise de imagens de glomérulos renais. Para isso, foram propostos métodos inovadores para a identificação de podocitopatia e segmentação de lesões podocitárias, bem como a análise do impacto do suporte computacional no desempenho de patologistas.

A seguir, detalhamos o atendimento a cada objetivo específico proposto na Seção 1.4:

- **Fornecer novos conjuntos de dados sobre lesões podocitárias, simulando cenários de multicoloração e desbalanceamento.** Este objetivo foi atingido com a criação e disponibilização, para fins acadêmicos, do conjunto de dados *D3*, composto por 1.401 imagens de glomérulos renais, anotadas em quatro classes distintas: controle normal, hiperplasia, hipertrofia e degeneração. A elaboração desse conjunto de dados, caracterizado por multicoloração e desbalanceamento entre classes, preenche uma lacuna crítica na área de patologia computacional, oferecendo à comunidade científica um recurso inédito e mais representativo da variabilidade observada em cenários clínicos reais.
- **Aperfeiçoar o desempenho de modelos de classificação binária para identificação de podocitopatia.** Para a tarefa de classificação da presença ou ausência de podocitopatia, foi desenvolvido o classificador *PodNet*. Este modelo alcançou os melhores resultados de *F1-score* no conjunto de validação final, superando arquiteturas tradicionais como *ResNet101v2*. A solução proposta demonstrou robustez mesmo em cenários com dados limitados e multicoloração.
- **Estudar o impacto do uso de classificadores automatizados na prática dos patologistas.** A avaliação da interação entre patologistas e o classificador automatizado mostrou que o uso do suporte computacional aumentou a acurácia dos especialistas de 88,3% para 95,0%, e elevou o índice de concordância (*Fleiss'*

Kappa) de 0,59 para 0,83. Esses resultados indicam que o uso de inteligência artificial pode reduzir variabilidade entre observadores e aprimorar a prática diagnóstica.

- **Propor modificações de topologia e arquitetura para aprimorar o desempenho de modelos de segmentação semântica.** Foi proposta uma solução baseada em *Feature Fusion Network* (FFN), uma estratégia de fusão de características a partir dos espaços de cor RGB e HED. A nova abordagem gerou ganhos médios de 9,13% em *F1-score* e de 8,56% em *IoU* em relação aos modelos tradicionais de segmentação. A estratégia de fusão baseada na diferença absoluta mostrou-se, em geral, a mais eficaz, embora variações como média e mínimo tenham se destacado em classes específicas.

Outro ponto importante é que este trabalho abordou um problema ainda inexplorado na literatura: a segmentação multiclasse de lesões podocitárias em glomérulos renais, em um cenário realista de multicoloração e desbalanceamento de classes. Assim, não foi possível realizar comparações diretas com trabalhos anteriores, que lidam majoritariamente com conjuntos controlados ou restritos a outras patologias glomerulares. Esse fator ressalta o caráter pioneiro e desafiador deste estudo.

Além das contribuições específicas, esta pesquisa também promove avanços para a compreensão da morfologia dos podócitos e sua relação com doenças renais como a nefropatia diabética e glomerulopatias diversas. A identificação precisa dos podócitos e suas alterações é um passo essencial para o diagnóstico precoce e o manejo clínico mais eficaz.

10.1 Perspectivas Futuras

Os resultados promissores desta tese abrem diversas frentes de continuidade:

- Explorar o uso de outros espaços de cor além do RGB e HED para enriquecer a extração de características relevantes nas imagens histológicas.
- Investigar métodos de fusão adaptativos que escolham dinamicamente a melhor operação de fusão de acordo com o conteúdo da imagem.
- Ampliar o conjunto de dados com novas amostras, incluindo outras colorações e fontes diversas, visando aumentar a robustez dos modelos.
- Aplicar arquiteturas mais recentes, como modelos V-LLM (*Vision-Language Large Models*) [183] e *Transformers* [184], para explorar o potencial de estratégias multimodais na análise de lâminas histológicas.

Portanto, esta tese contribui de maneira sólida para o fortalecimento da Patologia Computacional, demonstrando que a integração entre métodos de inteligência artificial e a prática clínica pode oferecer suporte robusto e confiável para o diagnóstico médico.

10.2 Limitações

Apesar dos avanços alcançados neste trabalho, algumas limitações devem ser reconhecidas e podem orientar pesquisas futuras.

Em relação aos conjuntos de dados elaborados, as anotações foram realizadas por três patologistas experientes, mas sem um processo formal de reanálise cruzada entre os especialistas. A ausência de uma validação mais ampla pode ter introduzido variações interobservadores não detectadas, o que sugere a importância de, em trabalhos futuros, envolver um número maior de patologistas e adotar métodos sistemáticos de consenso, como votações majoritárias ou etapas de reconciliação de divergências.

Outro ponto relevante refere-se ao método de segmentação proposto, que se baseia na fusão de características extraídas dos espaços de cores RGB e HED. Embora o espaço HED ofereça uma representação especializada para imagens histológicas coradas com Hematoxilina-Eosina e derivados, ele pode não capturar de forma ideal as nuances cromáticas de outras técnicas de coloração presentes no conjunto de dados, como o tricrômico e o PAMS. Essa limitação pode impactar a extração de características em imagens provenientes dessas técnicas, indicando a necessidade de estudos futuros que explorem a conversão para espaços de cor adaptativos ou específicos para diferentes corantes.

Essas limitações não invalidam as contribuições desta tese, mas reforçam a necessidade de continuidade nos estudos para aprimorar a generalização e a robustez das soluções propostas, visando sua aplicação prática em cenários clínicos cada vez mais diversos e realistas.

10.3 Publicações

Cada uma das três etapas definidas na metodologia resultou na produção de um artigo. A Tabela 10.1 apresenta os títulos e os veículos em que cada artigo foi publicado ou submetido.

Os resultados relacionados à proposta de classificação de podocitopatia em imagens histológicas de glomérulos renais foram publicados na conferência VISAPP 2022¹, classificada pela CAPES com conceito Qualis A3. O estudo sobre as implicações do uso de

¹<https://www.scitepress.org/Link.aspx?doi=10.5220/0010828600003124>

um classificador automatizado em colaboração com a prática de patologistas foi publicado no periódico *IEEE Journal of Translational Engineering in Health and Medicine*², classificado com conceito Qualis A4. Por fim, a proposta de melhoria do desempenho de redes tradicionais para segmentar lesões podocitárias foi submetida ao periódico *Modern Pathology (Elsevier)*³, que possui conceito Qualis A1.

Tabela 10.1: Artigos produzidos no decorrer do doutoramento.

| Título | Veículo submetido | Status |
|---|---|------------------|
| Podnet: Ensemble-based classification of podocytopathy on kidney glomerular images | Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (2022) | Publicado |
| Enhancing podocyte degenerative changes identification with pathologist collaboration: Implications for improved diagnosis in kidney diseases | IEEE Journal of Translational Engineering in Health and Medicine (2024) | Publicado |
| Segmentation of Lesion-Bearing Podocytes: A Proposal for Multistain and Imbalanced Data Contexts | Modern Pathology (Elsevier) (2025) | Submetido |

²<https://ieeexplore.ieee.org/document/10675352>

³<https://www.sciencedirect.com/journal/modern-pathology>

Referências

- [1] Sharifani, Koosha e Mahyar Amini: *Machine learning and deep learning: A review of methods and applications*. World Information Technology and Engineering Journal, 10(07):3897–3904, 2023. 1
- [2] Yanase, Juri e Evangelos Triantaphyllou: *A systematic survey of computer-aided diagnosis in medicine: Past and present developments*. Expert Systems with Applications, 138:112821, 2019, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417419305238>. 1
- [3] Torres, Richard e Eben Olson: *AI: What Have You Done for Us Lately?* Journal of the American Society of Nephrology, 29(8):2031–2032, 2018, ISSN 1046-6673. <https://jasn.asnjournals.org/content/29/8/2031>. 1
- [4] Abels, Esther, Liron Pantanowitz, Famke Aeffner, Mark D Zarella, Jeroen van der Laak, Marilyn M Bui, Venkata Np Vemuri, Anil V Parwani, Jeff Gibbs, Emmanuel Agosto-Arroyo, Andrew H Beck e Cleopatra Kozlowski: *Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association*. The Journal of pathology, 249(3):286–294, nov 2019, ISSN 1096-9896. <https://pubmed.ncbi.nlm.nih.gov/31355445https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6852275/>. 1
- [5] Kumar, V., A.K. Abbas, N. Fausto e R.N. Mitchell: *Robbins Patologia Humana*. Elsevier Science Health Science Division, 2008, ISBN 9788480863322. <https://books.google.com.br/books?id=VmwbvQqs-3AC>. 1
- [6] Mohan, H. e S. Mohan: *Essential Pathology for Dental Students*. Jaypee Brothers Medical Publishers Pvt. Limited, 2016, ISBN 9789386107749. https://books.google.com.br/books?id=Omm_DQAAQBAJ. 1
- [7] Song, Andrew H, Guillaume Jaume, Drew F K Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R Miller e Faisal Mahmood: *Artificial intelligence for digital and computational pathology*. Nature Reviews Bioengineering, 1(12):930–949, 2023, ISSN 2731-6092. <https://doi.org/10.1038/s44222-023-00096-8>. 1
- [8] Rashidi, Parisa e Azra Bihorac: *Artificial intelligence approaches to improve kidney care*. Nature Reviews Nephrology, 16:71–72, 2020. <https://doi.org/10.1038/s41581-019-0243-3>. 1
- [9] Barisoni, Laura, Kyle J Lafata, Stephen M Hewitt, Anant Madabhushi e Ulysses G J Balis: *Digital pathology and computational image analysis in nephropathology*.

- Nature Reviews Nephrology, 16(11):669–685, 2020, ISSN 1759-507X. <https://doi.org/10.1038/s41581-020-0321-6>. 1, 24
- [10] Gonzalez, Rafael C. e Richard E. Woods: *Digital image processing*. Prentice Hall, Upper Saddle River, N.J., 2008, ISBN 9780131687288 013168728X 9780135052679 013505267X. <http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>. 1, 16, 27
 - [11] Kang, Lei, Xiufeng Li, Yan Zhang e Terence T W Wong: *Deep learning enables ultraviolet photoacoustic microscopy based histological imaging with near real-time virtual staining*. Photoacoustics, 25:100308, 2022, ISSN 2213-5979. <https://www.sciencedirect.com/science/article/pii/S2213597921000689>. 2
 - [12] Yuan, Zheng, Andre Esteva e Ran Xu: *MetaHistoSeg: A Python Framework for Meta Learning in Histopathology Image Segmentation*. Em Engelhardt, Sandy, Ilkay Oksuz, Dajiang Zhu, Yixuan Yuan, Anirban Mukhopadhyay, Nicholas Heller, Sharon Xiaolei Huang, Hien Nguyen, Raphael Sznitman e Yuan Xue (editores): *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections*, páginas 268–275, Cham, 2021. Springer International Publishing, ISBN 978-3-030-88210-5. 2
 - [13] Anisuzzaman, D M, Hosein Barzekar, Ling Tong, Jake Luo e Zeyun Yu: *A deep learning study on osteosarcoma detection from histological images*. Biomedical Signal Processing and Control, 69:102931, 2021, ISSN 1746-8094. <https://www.sciencedirect.com/science/article/pii/S1746809421005280>. 2
 - [14] Bouteldja, Nassim, Barbara M Klinkhammer, Roman D Bülow, Patrick Droste, Simon W Otten, Saskia von Stillfried, Julia Moellmann, Susan M Sheehan, Ron Korstanje, Sylvia Menzel, Peter Bankhead, Matthias Mietsch, Charis Drummer, Michael Lehrke, Rafael Kramann, Jürgen Floege, Peter Boor e Dorit Merhof: *Deep Learning-Based Segmentation and Quantification in Experimental Kidney Histopathology*. Journal of the American Society of Nephrology, 32(1):52–68, 2021, ISSN 1046-6673. <https://jasn.asnjournals.org/content/32/1/52>. 2
 - [15] Jia, Xibin, Zheng Sun, Qing Mi, Dawei Yang e Zhenghan Yang: *A multimodality-contribution-aware tripnet for histologic grading of hepatocellular carcinoma*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, páginas 1–1, 2021. 2
 - [16] Lemley, Kevin V: *Machine Learning Comes to Nephrology*. Journal of the American Society of Nephrology, 30(10):1780–1781, 2019, ISSN 1046-6673. <https://jasn.asnjournals.org/content/30/10/1780>. 2
 - [17] Farris, Alton B, Juan Vizcarra, Mohamed Amgad, Lee A D Cooper, David Gutman e Julien Hogan: *Artificial intelligence and algorithmic computational pathology: an introduction with renal allograft examples*. Histopathology, 78(6):791–804, 2021. <https://onlinelibrary.wiley.com/doi/abs/10.1111/his.14304>. 2, 3

- [18] Barros, George O, Brenda Navarro, Angelo Duarte e Washington L C Dos-Santos: *PathoSpotter-K: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys*. Scientific Reports, 7(1):46769, 2017, ISSN 2045-2322. <https://doi.org/10.1038/srep46769>. 2, 25, 27, 45
- [19] Chagas, Paulo, Luiz Souza, Ikaro Araújo, Nayze Aldeman, Angelo Duarte, Michele Angelo, Washington L C Dos-Santos e Luciano Oliveira: *Classification of glomerular hypercellularity using convolutional features and support vector machine*. Artificial Intelligence in Medicine, 103:101808, 2020, ISSN 0933-3657. <https://www.sciencedirect.com/science/article/pii/S0933365719303148>. 2, 25, 27, 45
- [20] Lutnick, Brendon, David Manthey, Jan U Becker, Brandon Ginley, Katharina Moos, Jonathan E Zuckerman, Luis Rodrigues, Alexander J Gallan, Laura Barisoni, Charles E Alpers, Xiaoxin X Wang, Komuraiah Myakala, Bryce A Jones, Moshe Levi, Jeffrey B Kopp, Teruhiko Yoshida, Seung Seok Han, Sanjay Jain, Avi Z Rosenberg, Kuang Yu. Jen, Pinaki Sarder e for the Kidney Precision Medicine Project: *A user-friendly tool for cloud-based whole slide image segmentation, with examples from renal histopathology*. bioRxiv, 2021. <https://www.biorxiv.org/content/early/2021/08/17/2021.08.16.456524>. 2, 28, 29
- [21] Nagata, Michio: *Podocyte injury and its consequences*. Kidney international, 89(6):1221–1230, jun 2016, ISSN 1523-1755 (ELECTRONIC). 2, 8
- [22] Boor, Peter: *Artificial intelligence in nephropathology*. Nat Rev Nephrol, 16:4–6, 2020. <https://doi.org/10.1038/s41581-019-0220-x>. 2
- [23] Singh, Prabhat *et al.*: *Artificial intelligence in nephrology: Clinical applications and challenges*. Kidney Medicine, 7:100927, 2024. [https://www.kidneymedicinejournal.org/article/S2590-0595\(24\)00138-9/fulltext](https://www.kidneymedicinejournal.org/article/S2590-0595(24)00138-9/fulltext). 2
- [24] Goodfellow, Ian, Yoshua Bengio e Aaron Courville: *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. 2, 11, 12, 13, 14, 20, 44, 45, 50
- [25] Zhang, Tao e Mingyang Shi: *Multi-modal neuroimaging feature fusion for diagnosis of alzheimer’s disease*. Journal of Neuroscience Methods, 341:108795, 2020, ISSN 0165-0270. <https://www.sciencedirect.com/science/article/pii/S0165027020302181>. 2, 18
- [26] Kotsifa, E e VK Mavroeidis: *Present and future applications of artificial intelligence in kidney transplantation*. Journal of Clinical Medicine, 13(19):5939, 2024. 2
- [27] Chen, Hui Mei, Zhi Hong Liu, Cai Hong Zeng, Shi Jun Li, Qing Wen Wang e Lei Shi Li: *Podocyte lesions in patients with obesity-related glomerulopathy*. American journal of kidney diseases : the official journal of the National Kidney Foundation, 48(5):772–779, nov 2006, ISSN 1523-6838 (ELECTRONIC). 2, 8
- [28] Govind, Darshana, Briana A Santo, Brandon Ginley, Rabi Yacoub, Avi Z Rosenberg, Kuang Yu Jen, Vighnesh Walavalkar, Gregory E Wilding, Amber M Worrall, Imtiaz Mohammad e Pinaki Sarder: *Automated detection and quantification of Wilms’ Tumor 1-positive cells in murine diabetic kidney disease*. Em Tomaszewski, John

- E e Aaron D Ward (editores): *Medical Imaging 2021: Digital Pathology*, volume 11603, páginas 76–82. International Society for Optics and Photonics, SPIE, 2021. <https://doi.org/10.1117/12.2581387>. 2, 3, 4, 8, 30, 31, 32, 65
- [29] Saga, Nobuyuki, Kazuo Sakamoto, Taiji Matsusaka e Michio Nagata: *Glomerular filtrate affects the dynamics of podocyte detachment in a model of diffuse toxic podocytopathy*. *Kidney International*, 2021, ISSN 0085-2538. <https://www.sciencedirect.com/science/article/pii/S008525382100137X>. 2, 4, 8, 65
- [30] Maraszek, Kathryn E, Briana A Santo, Rabi Yacoub, John E Tomaszewski, Intiaz Mohammad, Amber M Worrall e Pinaki Sarder: *The presence and location of podocytes in glomeruli as affected by diabetes mellitus*. Em Tomaszewski, John E e Aaron D Ward (editores): *Medical Imaging 2020: Digital Pathology*, volume 11320, páginas 298–307. International Society for Optics and Photonics, SPIE, 2020. <https://doi.org/10.1117/12.2548904>. 2, 4, 8, 30, 31, 32, 65
- [31] Zimmermann, Marina, Martin Klaus, Milagros N Wong, Ann Katrin Thebille, Lukas Gernhold, Christoph Kuppe, Maurice Halder, Jennifer Kranz, Nicola Wanner, Fabian Braun, Sonia Wulf, Thorsten Wiech, Ulf Panzer, Christian F Krebs, Elion Hoxha, Rafael Kramann, Tobias B Huber, Stefan Bonn e Victor G Puelles: *Deep learning-based molecular morphometrics for kidney biopsies*. *JCI Insight*, 6(7), 2021. <https://doi.org/10.1172/jci.insight.144779>. 2, 3, 4, 8, 30, 31, 32, 65
- [32] Vaden, Shelly L: *Glomerular Disease*. *Topics in Companion Animal Medicine*, 26(3):128–134, 2011, ISSN 1938-9736. <https://www.sciencedirect.com/science/article/pii/S1938973611000493>. 2, 8, 65
- [33] Zeng, Caihong, Yang Nan, Feng Xu, Qunjuan Lei, Fengyi Li, Tingyu Chen, Shaoshan Liang, Xiaoshuai Hou, Bin Lv, Dandan Liang, WeiLi Luo, Chuanfeng Lv, Xiang Li, Guotong Xie e Zhihong Liu: *Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning*. *The Journal of Pathology*, 252(1):e5491, 2020. <https://onlinelibrary.wiley.com/doi/abs/10.1002/path.5491>. 3, 30, 32
- [34] Santo, Briana A, Darshana Govind, Parnaz Daneshpajouhnejad, Xiaoping Yang, Xiaoxin X Wang, Komuraiah Myakala, Bryce A Jones, Moshe Levi, Jeffrey B Kopp, Teruhiko Yoshida, Laura J Niedernhofer, David Manthey, Kyung Chul Moon, Seung Seok Han, Jarcy Zee, Avi Z Rosenberg e Pinaki Sarder: *PodoCount: A Robust, Fully Automated, Whole-Slide Podocyte Quantification Tool*. *Kidney International Reports*, 7(6):1377–1392, 2022, ISSN 2468-0249. <https://www.sciencedirect.com/science/article/pii/S2468024922012116>. 3, 30, 31, 32
- [35] Deng, Shujian, Xin Zhang, Wen Yan, Eric I Chao Chang, Yubo Fan, Maode Lai e Yan Xu: *Deep learning in digital pathology image analysis: a survey*. *Frontiers of Medicine*, 14(4):470–487, 2020, ISSN 2095-0225. <https://doi.org/10.1007/s11684-020-0782-9>. 3, 8

- [36] Campanella, Gabriele, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra e Thomas J Fuchs: *Clinical-grade computational pathology using weakly supervised deep learning on whole slide images*. Nature Medicine, 25(8):1301–1309, 2019, ISSN 1546-170X. <https://doi.org/10.1038/s41591-019-0508-1>. 3
- [37] Luzia Fátima Gonçalves Caputo, Lycia de Brito e Manso Gitirana e Pedro Paulo de Abreu: *Conceitos e métodos para a formação de profissionais em laboratórios de saúde. Capítulo 3: Técnicas Histológicas*. Escola Politécnica de Saúde Joaquim Venâncio, 2010, ISBN 9788598768410. https://www.epsjv.fiocruz.br/sites/default/files/capitulo_3_vol2.pdf. 3, 8, 9, 10
- [38] Jayapandian, Catherine P, Yijiang Chen, Andrew R Janowczyk, Matthew B Palmer, Clarissa A Cassol, Miroslav Sekulic, Jeffrey B Hodgins, Jarcy Zee, Stephen M Hewitt e John O Toole: *Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains*. Kidney international, 99(1)(June 2020):86–101, 2021. 4
- [39] Netter, F.H.: *Atlas of Human Anatomy*. Netter Basic Science. Elsevier Health Sciences, 2010, ISBN 9781455706938. https://books.google.com.br/books?id=nzOWtLe_kXYC. 6, 7
- [40] Nefrologia SBN, Sociedade Brasileira de: *Compreendendo os rins*, 2022. <https://www.sbn.org.br/o-que-e-nefrologia/compreendendo-os-rins/>, Last accessed 09 January 2022. 6
- [41] Silva Gouveia, Denise Sbrissia e, Alexandre Tortoza Bignelli, Silvia Regina Hoka-zono, Itamara Danucalov, Tobias August Siemens, Fernando Meyer, Luiz Sergio Santos, Ziliane Caetano Lopes Martins, Tiago César Mierzwa e Raquel Furquim: *Análise do impacto econômico entre as modalidades de terapia renal substitutiva*. Brazilian Journal of Nephrology, 39(2):162–171, 2017. 6
- [42] Chen, Liang-Chieh, George Papandreou, Florian Schroff e Hartwig Adam: *Rethinking atrous convolution for semantic image segmentation*. CoRR, abs/1706.05587, 2017. <http://arxiv.org/abs/1706.05587>. 6, 17, 18
- [43] Nerbass, Fabiana Baggio, Helbert do Nascimento Lima, José Andrade Moura-Neto, Jocemir Ronaldo Lugon e Ricardo Sesso: *Censo brasileiro de diálise 2022*. Jornal Brasileiro de Nefrologia, 46(2):160–167, 2024. 6
- [44] Field, M.J., C. Pollock e D. Harris: *The Renal System: Systems of the Body Series*. Systems of the Body. Elsevier Health Sciences, 2011, ISBN 9780702048852. <https://books.google.com.br/books?id=cYfGAAAAQBAJ>. 7
- [45] Arthur C. Guyton, John E. Hall: *Tratado de Fisiologia Médica*. Elsevier Inc., 12ª edição, 2011. 7
- [46] MedicinaNet: *Doenças glomerulares – roman gb bonegio david j salant*, 2021. https://www.medicinanet.com.br/conteudos/acp-medicine/5274/doencas_

glomerulares_%E2%80%93roman_gb_bonegio_david_j_salant.htm, Last accessed 11 october 2021. 7

- [47] Fuad, Ziyadeh N. e Wolf Gunter: *Pathogenesis of the Podocytopathy and Proteinuria in Diabetic Glomerulopathy*. Current Diabetes Reviews, 1(4), 2008. 8
- [48] Trimarchi, Hernán e Rosanna Coppo: *Podocytopathy in the mesangial proliferative immunoglobulin A nephropathy: new insights into the mechanisms of damage and progression*. Nephrology Dialysis Transplantation, 34(8):1280–1285, janeiro 2019, ISSN 0931-0509. <https://doi.org/10.1093/ndt/gfy413>. 8
- [49] Silvano, José, Augusta Praça, Inês Castro Ferreira, Ana Teresa Nunes, Pedro Rodrigues-pereira, Roberto Pestana Silva, Susana Sampaio e Manuel Pestana: *A case of lupus podocytopathy and a review of the literature*. Portuguese Journal of Nephrology & Hypertension, 35(April):39–42, 2021. 8
- [50] Diez-Sampedro, Ana, Oliver Lenz e Alessia Fornoni: *Podocytopathy in Diabetes: A Metabolic and Endocrine Disorder*. American Journal of Kidney Diseases, 58(4):637–646, 2011, ISSN 0272-6386. <https://www.sciencedirect.com/science/article/pii/S027263861100919X>. 8
- [51] Kim, K.: *1003-p: Cost of care for patients with diabetic kidney disease in veterans health administration*. Diabetes, 72(Supplement₁) : 1003 – –P, 2023.8
- [52] Barros., George, David Wanderley., Luciano Rebouças., Washington Santos., Angelo Duarte. e Flavio Vidal.: *Podnet: Ensemble-based classification of podocytopathy on kidney glomerular images*. Em *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP*., páginas 405–412. INSTICC, SciTePress, 2022, ISBN 978-989-758-555-5. 9, 10, 45
- [53] BiologiaNota10: *A preparação de lâminas para a visualização no microscópio de luz*, 2022. <https://biologianota10.com.br/post/6/a-preparacao-de-laminas-para-a-visualizacao-no-microscopio-de-luz>, Last accessed 09 January 2022. 9
- [54] Mo, Yujian, Yan Wu, Xinneng Yang, Feilin Liu e Yujun Liao: *Review the state-of-the-art technologies of semantic segmentation based on deep learning*. Neurocomputing, 493:626–646, 2022, ISSN 0925-2312. <https://www.sciencedirect.com/science/article/pii/S0925231222000054>. 10, 56
- [55] Velastegui, Ronny e Marius Pedersen: *The impact of using different color spaces in histological image classification using convolutional neural networks*. Em *2021 9th European Workshop on Visual Information Processing (EUVIP)*, páginas 1–6, 2021. 10, 56
- [56] Ruifrok, Arnout e Dennis Johnston: *Quantification of histochemical staining by color deconvolution*. Anal Quant Cytol Histol, 23, 2001. 10, 56

- [57] Costa, Eliana, Edna Costa, Alberto Lopes, Eduardo Sacramento, Adriana Lopes, Ewerton Matos, Laura Vasconcelos e Mitermayer Reis: *Perfil clínico e laboratorial de pacientes com leptospirose atendidos em hospital terciário*. Revista da Sociedade Brasileira de Medicina Tropical, 38(5):375–379, 2005. 10
- [58] Walt, Stefan Van der, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart e Tony Yu: *scikit-image: image processing in python*. PeerJ, 2:e453, 2014. 10
- [59] Russell, Stuart e Peter Norvig: *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edição, 2003. 11
- [60] Fukushima, Kuniyoshi: *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*. Biological Cybernetics, 36:193–202, 1980. 11
- [61] LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard e L. D. Jackel: *Backpropagation applied to handwritten zip code recognition*. Neural Computation, 1(4):541–551, 1989. 11
- [62] Krizhevsky, Alex, Ilya Sutskever e Geoffrey E Hinton: *Imagenet classification with deep convolutional neural networks*. Em Pereira, F., C.J. Burges, L. Bottou e K.Q. Weinberger (editores): *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. 11
- [63] Deng, Jia, Wei Dong, Richard Socher, Li Jia Li, Kai Li e Li Fei-Fei: *Imagenet: A large-scale hierarchical image database*. Em *2009 IEEE Conference on Computer Vision and Pattern Recognition*, páginas 248–255, 2009. 11, 49
- [64] Dumoulin, Vincent e Francesco Visin: *A guide to convolution arithmetic for deep learning*, 2018. 12, 13, 14
- [65] Agarap, Abien Fred: *Deep learning using rectified linear units (relu)*, 2019. <https://arxiv.org/abs/1803.08375>. 12
- [66] Matoba, Kyle, Nikolaos Dimitriadis e François Fleuret: *The theoretical expressiveness of maxpooling*, 2022. <https://arxiv.org/abs/2203.01016>. 12, 45
- [67] Smith, Leslie N.: *A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay*. CoRR, abs/1803.09820, 2018. <http://arxiv.org/abs/1803.09820>. 13, 14
- [68] Ruder, Sebastian: *An overview of gradient descent optimization algorithms.*, 2016. <http://arxiv.org/abs/1609.04747>, cite arxiv:1609.04747Comment: Added derivations of AdaMax and Nadam. 13
- [69] Duchi, John, Elad Hazan e Yoram Singer: *Adaptive subgradient methods for online learning and stochastic optimization*. Journal of Machine Learning Research, 12(Jul):2121–2159, 2011. 13

- [70] Dauphin, Yann N., Harm de Vries, Junyoung Chung e Yoshua Bengio: *Rmsprop and equilibrated adaptive learning rates for non-convex optimization*. CoRR, abs/1502.04390, 2015. <http://dblp.uni-trier.de/db/journals/corr/corr1502.html#DauphinVCB15>. 13
- [71] Zeiler, Matthew D.: *Adadelta: An adaptive learning rate method*, 2012. <https://arxiv.org/abs/1212.5701>. 13
- [72] Mannor, Shie, Dori Peleg e Reuven Rubinstein: *The cross entropy method for classification*. Em *Proceedings of the 22nd International Conference on Machine Learning, ICML '05*, página 561–568, New York, NY, USA, 2005. Association for Computing Machinery, ISBN 1595931805. <https://doi.org/10.1145/1102351.1102422>. 14
- [73] Alfarozi, Syukron Abu Ishaq, Kuntpong Woraratpanya e Kitsuchart Pasupa: *Hinge loss projection for classification*. Em Hirose, Akira, Seiichi Ozawa, Kenji Doya, Kazushi Ikeda, Minhoo Lee e Derong Liu (editores): *Proceedings of the 23rd International Conference on Neural Information Processing (ICONIP 2016), 16-21 Oct 2016, Kyoto, Japan*, volume 9948 de *Lecture Notes in Computer Science*, páginas 250–258, 2016. 14
- [74] Elfving, Stefan, Eiji Uchibe e Kenji Doya: *Sigmoid-weighted linear units for neural network function approximation in reinforcement learning*. CoRR, abs/1702.03118, 2017. <http://arxiv.org/abs/1702.03118>. 14
- [75] Liu, Xinyu e Xiaoguang Di: *Tanhexp: A smooth activation function with high convergence speed for lightweight neural networks*. CoRR, abs/2003.09855, 2020. <https://arxiv.org/abs/2003.09855>. 14
- [76] Agarap, Abien Fred: *Deep learning using rectified linear units (relu)*, 2018. <http://arxiv.org/abs/1803.08375>, cite arxiv:1803.08375Comment: 7 pages, 11 figures, 9 tables. 14
- [77] Wang, Yingying, Yibin Li, Yong Song e Xuewen Rong: *The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition*. Applied Sciences, 10(5), 2020, ISSN 2076-3417. <https://www.mdpi.com/2076-3417/10/5/1897>. 14
- [78] Smith, Samuel L., Pieter-Jan Kindermans e Quoc V. Le: *Don't decay the learning rate, increase the batch size*. CoRR, abs/1711.00489, 2017. <http://arxiv.org/abs/1711.00489>. 14
- [79] Simonyan, Karen e Andrew Zisserman: *Very deep convolutional networks for large-scale image recognition*, 2015. 14, 15, 43
- [80] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens e Zbigniew Wojna: *Rethinking the inception architecture for computer vision*, 2015. 14, 15, 43
- [81] Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke e Alex Alemi: *the Impact of Residual Connections on Learning*. Arxiv preprint, 2016. 14, 15, 43
- [82] Yu, Tong e Hong Zhu: *Hyper-parameter optimization: A review of algorithms and applications*. CoRR, abs/2003.05689, 2020. <https://arxiv.org/abs/2003.05689>. 15

- [83] Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever e Ruslan Salakhutdinov: *Dropout: A simple way to prevent neural networks from overfitting*. Journal of Machine Learning Research, 15(56):1929–1958, 2014. <http://jmlr.org/papers/v15/srivastava14a.html>. 15
- [84] He, Kaiming, Xiangyu Zhang, Shaoqing Ren e Jian Sun: *Deep residual learning for image recognition*, 2015. 15, 28, 43
- [85] Huang, Gao, Zhuang Liu, Laurens van der Maaten e Kilian Q. Weinberger: *Densely connected convolutional networks*, 2018. 15, 30, 43
- [86] Chollet, François: *Xception: Deep learning with depthwise separable convolutions*, 2017. 15, 43
- [87] Tan, Mingxing e Quoc V. Le: *Efficientnet: Rethinking model scaling for convolutional neural networks*. CoRR, abs/1905.11946, 2019. <http://arxiv.org/abs/1905.11946>. 15
- [88] Li, Lisha, Kevin G. Jamieson, Giulia DeSalvo, Afshin Rostamizadeh e Ameet Talwalkar: *Efficient hyperparameter optimization and infinitely many armed bandits*. CoRR, abs/1603.06560, 2016. <http://arxiv.org/abs/1603.06560>. 15, 16, 44
- [89] Xiao, Xueli, Ming Yan, Sunitha Basodi, Chunyan Ji e Yi Pan: *Efficient hyperparameter optimization in deep learning using a variable length genetic algorithm*. CoRR, abs/2006.12703, 2020. <https://arxiv.org/abs/2006.12703>. 15
- [90] Shekar, B. H e Guesh Dagnew: *Grid search-based hyperparameter tuning and classification of microarray cancer data*. Em *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, páginas 1–8, 2019. 15
- [91] Andonie, Razvan e Adrian-Catalin Florea: *Weighted random search for CNN hyperparameter optimization*. CoRR, abs/2003.13300, 2020. <https://arxiv.org/abs/2003.13300>. 15, 16
- [92] Slivkins, Aleksandrs: *Introduction to multi-armed bandits*. CoRR, abs/1904.07272, 2019. <http://arxiv.org/abs/1904.07272>. 16
- [93] Nixon, Mark e Alberto S. Aguado: *Feature Extraction & Image Processing, Second Edition*. Academic Press, Inc., USA, 2nd edição, 2008, ISBN 0123725380. 16
- [94] Chen, Leiyu, Shaobo Li, Qiang Bai, Jing Yang, Sanlong Jiang e Yanming Miao: *Review of Image Classification Algorithms Based on Convolutional Neural Networks*. Remote Sensing, 13(22), 2021, ISSN 2072-4292. <https://www.mdpi.com/2072-4292/13/22/4712>. 16
- [95] Li, Chuanzi, Jining Feng, Li Hu, Junhong Li e Haibin Ma: *Review of image classification method based on deep transfer learning*. Em *2020 16th International Conference on Computational Intelligence and Security (CIS)*, páginas 104–108, 2020. 16
- [96] Serag, Ahmed, Adrian Ion-Margineanu, Hammad Qureshi, Ryan McMillan, Marie Judith Saint Martin, Jim Diamond, Paul O'Reilly e Peter Hamilton: *Translational AI and Deep Learning in Diagnostic Pathology*. Frontiers in Medicine, 6, 2019, ISSN 2296-858X. <https://www.frontiersin.org/article/10.3389/fmed.2019.00185>. 16

- [97] Srinidhi, Chetan L, Ozan Ciga e Anne L Martel: *Deep neural network models for computational histopathology: A survey*. Medical Image Analysis, 67:101813, 2021, ISSN 1361-8415. <https://www.sciencedirect.com/science/article/pii/S1361841520301778>. 16
- [98] Solem, Jan Erick: *Programming computer vision with Python*. O'Reilly, Beijing; Cambridge; Sebastopol [etc.], 2012, ISBN 9781449316549 1449316549. http://www.amazon.com/dp/1449316549/ref=pe_259560_30100590_pd_re_dt_dt7. 16
- [99] Minaee, Shervin, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz e Demetri Terzopoulos: *Image segmentation using deep learning: A survey*. CoRR, abs/2001.05566, 2020. <https://arxiv.org/abs/2001.05566>. 16
- [100] Guo, Yanming, Yu Liu, Theodoros Georgiou e Michael S. Lew: *A review of semantic segmentation using deep neural networks*. International Journal of Multimedia Information Retrieval, 7(2):87–93, 2018. 16
- [101] Ronneberger, Olaf, Philipp Fischer e Thomas Brox: *U-net: Convolutional networks for biomedical image segmentation*. CoRR, abs/1505.04597, 2015. <http://arxiv.org/abs/1505.04597>. 17, 18
- [102] Oktay, Ozan, Jo Schlemper, Loïc Le Folgoc, Matthew C. H. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker e Daniel Rueckert: *Attention u-net: Learning where to look for the pancreas*. CoRR, abs/1804.03999, 2018. <http://arxiv.org/abs/1804.03999>. 17
- [103] Ronneberger, Olaf, Philipp Fischer e Thomas Brox: *U-net: Convolutional networks for biomedical image segmentation*, 2015. 17, 18, 53
- [104] Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh e Jianming Liang: *Unet++: Redesigning skip connections to exploit multiscale features in image segmentation*, 2020. 18, 53
- [105] Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy e Alan L. Yuille: *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*. CoRR, abs/1606.00915, 2016. <http://arxiv.org/abs/1606.00915>. 18, 28
- [106] Chen, Liang Chieh, George Papandreou, Florian Schroff e Hartwig Adam: *Rethinking atrous convolution for semantic image segmentation*, 2017. 18, 53
- [107] Chen, Liang Chieh, Yukun Zhu, George Papandreou, Florian Schroff e Hartwig Adam: *Encoder-decoder with atrous separable convolution for semantic image segmentation*, 2018. 18, 19
- [108] He, Kaiming, Xiangyu Zhang, Shaoqing Ren e Jian Sun: *Spatial pyramid pooling in deep convolutional networks for visual recognition*. CoRR, abs/1406.4729, 2014. <http://arxiv.org/abs/1406.4729>. 18

- [109] Chaurasia, Abhishek e Eugenio Culurciello: *Linknet: Exploiting encoder representations for efficient semantic segmentation*. Em *2017 IEEE Visual Communications and Image Processing (VCIP)*, páginas 1–4, 2017. 18, 53
- [110] Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang e Jiaya Jia: *Pyramid scene parsing network*, 2017. 18, 53
- [111] Lin, Tsung Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan e Serge Belongie: *Feature pyramid networks for object detection*, 2017. 18, 53
- [112] Li, Rui, Shunyi Zheng, Ce Zhang, Chenxi Duan, Jianlin Su, Libo Wang e Peter M. Atkinson: *Multiaattention network for semantic segmentation of fine-resolution remote sensing images*. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022, ISSN 1558-0644. <http://dx.doi.org/10.1109/TGRS.2021.3093977>. 18, 53
- [113] Li, Hanchao, Pengfei Xiong, Jie An e Lingxue Wang: *Pyramid attention network for semantic segmentation*, 2018. 18, 53
- [114] Zhang, Yifan, Xin Li, Yulin Wang e Liangpei Zhang: *A comprehensive survey on deep learning-based remote sensing image super-resolution methods*. *Expert Systems with Applications*, 222:119511, 2024. 18
- [115] Hesamian, Mohammad Hassan, Wugai Jia, Xiaohua He e Paul Kennedy: *Deep learning techniques for medical image segmentation: Achievements and challenges*. *Computers in Biology and Medicine*, 134:104141, 2022. 18
- [116] Song, Weiwei, Shutao Li, Leyuan Fang e Ting Lu: *Hyperspectral image classification with deep feature fusion network*. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6):3173–3184, 2018. 18
- [117] Feng, Yong, Enbo Luo, Hai Lu e SuWei Zhai: *Cross-modality feature fusion for night pedestrian detection*. *Frontiers in Physics*, 12, 2024, ISSN 2296-424X. <https://www.frontiersin.org/articles/10.3389/fphy.2024.1356248>. 18
- [118] Almeida, A.P.G.S. de e F de Barros Vidal: *L-CNN: a lattice cross-fusion strategy for multistream convolutional neural networks*. *Electronics Letters*, 55(22):1180–1182, 2019. <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/el.2019.2631>. 19, 56
- [119] James, Gareth, Daniela Witten, Trevor Hastie e Robert Tibshirani: *An Introduction to Statistical Learning*. Springer texts in statistics, 2013, ISBN 9781461471370. 19, 20, 21, 43, 66
- [120] Rezatofighi, Seyed Hamid, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian D. Reid e Silvio Savarese: *Generalized intersection over union: A metric and A loss for bounding box regression*. *CoRR*, abs/1902.09630, 2019. <http://arxiv.org/abs/1902.09630>. 19, 66
- [121] OPER: *Uma visão geral sobre machine learning – classificaÇÃo*, 2022. <https://operdata.com.br/blog/uma-visao-geral-sobre-machine-learning/>, Last accessed 11 January 2022. 20

- [122] James, Gareth, Daniela Witten, Trevor Hastie e Robert Tibshirani: *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013. <https://faculty.marshall.usc.edu/gareth-james/ISL/>. 20, 27, 50, 59
- [123] Silva, Daniel F. da e Renan M. de Almeida: *Aplicações da curva roc em estudos clínicos e experimentais*. Jornal Vascular Brasileiro, 19:e20200186, 2020. <https://www.scielo.br/j/jvb/a/8S8Pfqnz8csmQJVqwgZT8gH/>. 21, 59
- [124] Fatourech, Mehrdad, Rabab K. Ward, Steven G. Mason, Jane Huggins, Alois Schlögl e Gary E. Birch: *Comparison of evaluation metrics in classification applications with imbalanced datasets*. Em *2008 Seventh International Conference on Machine Learning and Applications*, páginas 777–782, 2008. 21
- [125] Beitzel, Steven M, Eric C Jensen e Ophir Frieder: *MAP*, páginas 1691–1692. Springer US, Boston, MA, 2009, ISBN 978-0-387-39940-9. https://doi.org/10.1007/978-0-387-39940-9_492. 21
- [126] Hui, Jonathan: *map (mean average precision) for object detection*, 2022. <https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173>, Last accessed 11 January 2022. 22
- [127] Fleiss, J.L. *et al.*: *Measuring nominal scale agreement among many raters*. Psychological Bulletin, 76(5):378–382, 1971. 22, 59
- [128] Kannan, Shruti, Laura A Morgan, Benjamin Liang, Mckenzie G Cheung, Christopher Q Lin, Dan Mun, Ralph G Nader, Mostafa E Belghasem, Joel M Henderson, Jean M Francis, Vipul C Chitalia e Vijaya B Kolachalama: *Segmentation of Glomeruli Learning*. Kidney International Reports, 4(7):955–962, 2019, ISSN 2468-0249. <https://doi.org/10.1016/j.ekir.2019.04.008>. 25, 26
- [129] Gallego, Jaime, Zaneta Swiderska-Chadaj, Tomasz Markiewicz, Michifumi Yamashita, M Alejandra Gabaldon e Arkadiusz Gertych: *A U-Net based framework to quantify glomerulosclerosis in digitized PAS and H&E stained human tissues*. Computerized Medical Imaging and Graphics, 89:101865, 2021, ISSN 0895-6111. <https://www.sciencedirect.com/science/article/pii/S0895611121000136>. 25, 26
- [130] Ginley, Brandon, Brendon Lutnick, Kuang Yu Jen, Agnes B Fogo, Sanjay Jain, Avi Rosenberg, Vighnesh Walavalkar, Gregory Wilding, John E Tomaszewski, Rabi Yacoub, Giovanni Maria Rossi e Pinaki Sarder: *Computational Segmentation and Classification of Diabetic Glomerulosclerosis*. Journal of the American Society of Nephrology, 30(10):1953–1967, 2019, ISSN 1046-6673. <https://jasn.asnjournals.org/content/30/10/1953>. 25, 27
- [131] Shubham, Shubham, Nikita Jain, Vedika Gupta, Senthilkumar Mohan, Mazeyanti Mohd Ariffin e Ali Ahmadian: *Identify glomeruli in human kidney tissue images using a deep learning approach*. Soft Computing, 2021, ISSN 1433-7479. <https://doi.org/10.1007/s00500-021-06143-z>. 25, 27

- [132] Haan, Kevin de, Yijie Zhang, Jonathan E Zuckerman, Tairan Liu, Anthony E Sisk, Miguel F P Diaz, Kuang Yu Jen, Alexander Nobori, Sofia Liou, Sarah Zhang, Rana Riahi, Yair Rivenson, W Dean Wallace e Aydogan Ozcan: *Deep learning-based transformation of H&E stained tissues into special stains*. Nature Communications, 12(1):4884, 2021, ISSN 2041-1723. <https://doi.org/10.1038/s41467-021-25221-2>. 25, 28, 29
- [133] Jiang, Lei, Wenkai Chen, Bao Dong, Ke Mei, Chuang Zhu, Jun Liu, Meishun Cai, Yu Yan, Gongwei Wang e Li Zuo: *MACHINE LEARNING , COMPUTATIONAL PATHOLOGY , AND BIOPHYSICAL IMAGING A Deep Learning-Based Approach for Glomeruli Instance Segmentation from Multistained Renal Biopsy Pathologic Images*. The American Journal of Pathology, 191(8):1431–1441, 2021, ISSN 0002-9440. <https://doi.org/10.1016/j.ajpath.2021.05.004>. 25, 28
- [134] Yang, Cheng Kun, Ching Yi Lee, Hsiang Sheng Wang, Shun Chen Huang, Peir In Liang, Jung Sheng Chen, Chang Fu Kuo, Kun Hua Tu, Chao Yuan Yeh e Tai Di Chen: *Glomerular disease classification and lesion identification by machine learning*. Biomedical Journal, 2021, ISSN 2319-4170. <https://www.sciencedirect.com/science/article/pii/S2319417021001116>. 25, 28
- [135] Uchino, Eiichiro, Kanata Suzuki, Noriaki Sato, Ryosuke Kojima e Yoshinori Tamada: *Classification of glomerular pathological findings using deep learning and nephrologist – AI collective intelligence approach*. International Journal of Medical Informatics, 141:1–14, 2020. 25, 26
- [136] Mathur, Puneet, Meghna P. Ayyar, Rajiv Ratn Shah e Shree G. Sharma: *Exploring classification of histological disease biomarkers from renal biopsy images*. Em 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), páginas 81–90, 2019. 25, 26, 45
- [137] Ginley, Brandon, Kuang Yu Jen, Seung Seok Han, Luis Rodrigues, Sanjay Jain, Agnes B Fogo, Jonathan Zuckerman, Vighnesh Walavalkar, Jeffrey C Miecznikowski, Yumeng Wen, Felicia Yen, Donghwan Yun, Kyung Chul Moon, Avi Rosenberg, Chirag Parikh e Pinaki Sarder: *Automated Computational Detection of Interstitial Fibrosis, Tubular Atrophy, and Glomerulosclerosis*. Journal of the American Society of Nephrology, 32(4):837–850, 2021, ISSN 1046-6673. <https://jasn.asnjournals.org/content/32/4/837>. 25, 28
- [138] He, Hongjie, Hongzhang Xu, Ying Zhang, Kyle Gao, Huxiong Li, Lingfei Ma e Jonathan Li: *Mask r-cnn based automated identification and extraction of oil well sites*. International Journal of Applied Earth Observation and Geoinformation, 112:102875, 2022, ISSN 1569-8432. <https://www.sciencedirect.com/science/article/pii/S1569843222000772>. 28
- [139] Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg e Li Fei-Fei: *ImageNet Large Scale Visual Recognition Challenge*. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015. 28, 43, 45, 53

- [140] Aron, Cleo, Weis Jan, Niklas Bindzus, Jonas Voigt, Marlen Runz, Svetlana Hertjens, Matthias M Gaida, Zoran V Popovic e Stefan Porubsky: *Assessment of glomerular morphological patterns by deep learning algorithms*. Journal of Nephrology, (0123456789), 2022, ISSN 1724-6059. <https://doi.org/10.1007/s40620-021-01221-9>. 28
- [141] Minamiguchi, Sachiko, Hironori Haga, Hideki Yokoi e Motoko Yanagita: *Evaluation of Kidney Histological Images*. Kidney International Reports, 6(9):2445–2454, 2021, ISSN 2468-0249. <https://doi.org/10.1016/j.ekir.2021.06.008>. 28, 29, 45
- [142] Chu, Casey, Andrey Zhmoginov e Mark Sandler: *CycleGAN, a master of steganography*. 2017. <https://arxiv.org/abs/1712.02950>. 29
- [143] Jagadeesh Kumar, S. J. K., G. P. Kanna, D. P. Raja e Y. Kumar: *A comprehensive study on deep learning models for the detection of ovarian cancer and glomerular kidney disease using histopathological images*. Archives of Computational Methods in Engineering, 32(1):35–61, 2025. 29
- [144] Lutnick, Brendon e Pinaki Sarder: *Generative modeling of histology tissue reduces human annotation effort for segmentation model development*. bioRxiv, 2021. <https://www.biorxiv.org/content/early/2021/10/16/2021.10.15.464564>. 29
- [145] Fan, Zong, Varun Kelkar, Mark A. Anastasio e Hua Li: *Application of datasetGAN in medical imaging: preliminary studies*, 2022. <https://arxiv.org/abs/2202.13463>. 29
- [146] Ginley, Brandon, Kuang Yu Jen e Pinaki Sarder: *Structural Knowledge Transfer of Panoptic Kidney Segmentation to Other Stains, Organs, and Species*. bioRxiv, 2021. <https://www.biorxiv.org/content/early/2021/10/23/2021.10.21.465370>. 29
- [147] Farhat, I., E. Maréchal, D. Calmo, M. Ansart, M. Paindavoine, P. Bard, G. Tarris, D. Ducloux, S. A. Felix, L. Martin, C. Tinel, J. B. Gibier, M. Funes de la Vega, J. M. Rebibou, J. Bamoulid e M. Legendre: *Recognition of intraglomerular histological features with deep learning in protocol transplant biopsies and their association with kidney function and prognosis*. Clinical Kidney Journal, 17(2):sfae019, Feb 2024. 30, 31, 32
- [148] Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker e Daniel Rueckert: *Attention u-net: Learning where to look for the pancreas*, 2018. <https://arxiv.org/abs/1804.03999>. 30
- [149] Tang, J., L. Zhao, M. Wu, Z. Jiang, J. Cao e X. Bao: *A se-densenet-lstm model for locomotion mode recognition in lower limb exoskeleton*. PeerJ Comput Sci, 10:e1881, 2024. 30
- [150] Aggarwal, Alankrita, Mamta Mittal e Gopi Battineni: *Generative adversarial network: An overview of theory and applications*. International Journal of Information Management Data Insights, 1(1):100004, 2021. 30
- [151] Wada, Kentaro: *labelme: Image polygonal annotation with python*. <https://github.com/wkentaro/labelme>, 2018. 41

- [152] O'Malley, Tom, Elie Bursztein, James Long, François Chollet, Haifeng Jin, Luca Invernizzi *et al.*: *Keras Tuner*. <https://github.com/keras-team/keras-tuner>, 2019. 44
- [153] Bukowy, John D, Alex Dayton, Dustin Cloutier, Anna D Manis, Alexander Staruschenko, Julian H Lombard, Leah C Solberg Woods, Daniel A Beard e Allen W Cowley: *Region-Based Convolutional Neural Nets for Localization of Glomeruli in Trichrome-Stained Whole Kidney Sections*. Journal of the American Society of Nephrology, 29(8):2081–2088, 2018, ISSN 1046-6673. <https://jasn.asnjournals.org/content/29/8/2081>. 45
- [154] Tipping, Michael E e Christopher M Bishop: *Mixtures of Probabilistic Principal Component Analysers*. Neural Computation 11(2), 2006. 45
- [155] Chollet, Francois *et al.*: *Keras*, 2015. <https://github.com/fchollet/keras>, Accessed: 2022-01-28. 46
- [156] Ligabue, Giulia, Federico Pollastri, Francesco Fontana, Marco Leonelli, Luciana Furci, Silvia Giovanella, Gaetano Alfano, Gianni Cappelli, Francesca Testa, Federico Bolelli, Costantino Grana e Riccardo Magistroni: *Evaluation of the Classification Accuracy of the Kidney Biopsy Direct Immunofluorescence through Convolutional Neural Networks*. Clinical journal of the American Society of Nephrology : CJASN, 15(10):1445–1454, oct 2020, ISSN 1555-905X (ELECTRONIC). 47
- [157] Salas, Joaquin, Flavio de Barros Vidal e Francisco Martinez-Trinidad: *Deep learning: Current state*. IEEE Latin America Transactions, 17(12):1925–1945, 2019. 48
- [158] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens e Zbigniew Wojna: *Rethinking the inception architecture for computer vision*. CoRR, abs/1512.00567, 2015. <http://arxiv.org/abs/1512.00567>. 48
- [159] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten e Kilian Q. Weinberger: *Densely connected convolutional networks*. Em *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 2261–2269, 2017. 48
- [160] Tan, Mingxing e Quoc V. Le: *Efficientnet: Rethinking model scaling for convolutional neural networks*. CoRR, abs/1905.11946, 2019. <http://arxiv.org/abs/1905.11946>. 48
- [161] Thomas, Simon M, James G Lefevre, Glenn Baxter e Nicholas A Hamilton: *Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer*. Medical Image Analysis, 68:101915, 2021, ISSN 1361-8415. <https://www.sciencedirect.com/science/article/pii/S1361841520302796>. 48
- [162] Li, Yuqian, Jun min Wu e Qisong Wu: *Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning*. IEEE Access, 7:21400–21408, 2019. 48
- [163] Hong, Runyu, Wenke Liu, Deborah DeLair, Narges Razavian e David Fenyö: *Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models*. Cell Reports Medicine, 2(9):100400, 2021, ISSN 2666-3791. <https://www.sciencedirect.com/science/article/pii/S2666379121002585>. 48

- [164] Li, Lisha, Kevin G. Jamieson, Giulia DeSalvo, Afshin Rostamizadeh e Ameet Talwalkar: *Efficient hyperparameter optimization and infinitely many armed bandits*. CoRR, abs/1603.06560, 2016. <http://arxiv.org/abs/1603.06560>. 49
- [165] Li, Lisha, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh e Ameet Talwalkar: *Hyperband: A novel bandit-based approach to hyperparameter optimization*, 2018. 49
- [166] Mo, Yujian, Yan Wu, Xinneng Yang, Feilin Liu e Yujun Liao: *Review the state-of-the-art technologies of semantic segmentation based on deep learning*. Neurocomputing, 493:626–646, 2022, ISSN 0925-2312. <https://www.sciencedirect.com/science/article/pii/S0925231222000054>. 53
- [167] Hatipoglu, Nuh e Gokhan Bilgin: *Histopathological image segmentation using u-net based models*. Em *2021 Medical Technologies Congress (TIPTEKNO)*, páginas 1–4, 2021. 53
- [168] Sudre, Carole H., Wenqi Li, Tom Vercauteren, Sebastien Ourselin e M. Jorge Cardoso: *Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations*, página 240–248. Springer International Publishing, 2017, ISBN 9783319675589. http://dx.doi.org/10.1007/978-3-319-67558-9_28. 54
- [169] Sugino, Takaaki, Toshihiro Kawase, Shinya Onogi, Taichi Kin, Nobuhito Saito e Yoshikazu Nakajima: *Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks*. Healthcare, 9(8):938, 2021. <https://www.mdpi.com/2227-9032/9/8/938>. 54
- [170] Chen, Wei, Kai Yang, Zhiqiang Yu, Yew Soon Ong, Yaochu Jin, Min Wu, Shitong Wang, Ke Tang, Xin Yao, Zhi Hua Zhou, Byoung Tak Zhang, Kay Chen Tan, Kalyanmoy Deb, Jürgen Branke, Carlos A. Coello Coello, Qingfu Zhang, Jacek M. Zurada, Ponnuthurai N. Suganthan, Nikola Kasabov, Xin She Yang, Amir Hussain, Huan Liu, Qiang Yang, Philip S. Yu e Zhi Hua Zhou: *A survey on imbalanced learning: Latest research, applications, and future directions*. Artificial Intelligence Review, 57(6):137, 2024. <https://link.springer.com/article/10.1007/s10462-024-10759-6>. 54
- [171] Taha, Alireza A. e Allan Hanbury: *Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool*. BMC Medical Imaging, 15(1):29, 2015. 54
- [172] Diarra, Mamadou, Ayikpa Kacoutchy Jean, Abou Ballo e Kouassi Médard: *Analysis of the Impact of Color Spaces on Skin Cancer Diagnosis Using Deep Learning Techniques*. Revue d’Intelligence Artificielle, 37:1377–1385, 2023. 56
- [173] Barros., George Oliveira, David Campos Wanderley., Luciano Oliveira RebouÇas., Washington L. C. dos Santos., Angelo A. Duarte. e Flavio de Barros Vidal.: *Podnet: Ensemble-based classification of podocytopathy on kidney glomerular images*. Em *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*, páginas 405–412. INSTICC, SciTePress, 2022, ISBN 978-989-758-555-5. 56
- [174] Yang, Jianbo, Minh Nhut Nguyen, Phyo Phyo San, Xiaoli Li e Shonali Krishnaswamy: *Deep convolutional neural networks on multichannel time series for human activity recognition*. Em *Ijcai*, volume 15, páginas 3995–4001. Buenos Aires, Argentina, 2015. 56

- [175] Quan, Chanqin, Lei Hua, Xiao Sun, Wenjun Bai *et al.*: *Multichannel convolutional neural network for biological relation extraction*. BioMed research international, 2016, 2016. 56
- [176] Bidart, Rene e Alexander Wong: *Triresnet: A deep triple-stream residual network for histopathology grading*. Em *Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part II* 16, páginas 369–382. Springer, 2019. 56
- [177] Zhao, Sijie, Xueliang Zhang, Pengfeng Xiao e Guangjun He: *Exchanging dual-encoder–decoder: A new strategy for change detection with semantic guidance and spatial localization*. IEEE Transactions on Geoscience and Remote Sensing, 61:1–16, 2023, ISSN 1558-0644. <http://dx.doi.org/10.1109/TGRS.2023.3327780>. 56
- [178] Ribeiro, Rafael Oliveira, João C. Neves, Arnout Ruifrok e Flavio de Barros Vidal: *Improving the evidential value of low-quality face images with aggregation of deep neural network embeddings*. Science Justice, 64(5):509–520, 2024, ISSN 1355-0306. <https://www.sciencedirect.com/science/article/pii/S135503062400073X>. 56
- [179] Araujo, P. H. Luz de, A. P. G. S. de Almeida, F. Ataide Braz *et al.*: *Sequence-aware multi-modal page classification of brazilian legal documents*. International Journal on Document Analysis and Recognition (IJDAR), 26:33–49, 2023. 56
- [180] El-Badry, Ashraf Mohammad, Stefan Breitenstein, Wolfram Jochum, Kay Washington, Valérie Paradis, Laura Rubbia-Brandt, Milo A Puhan, Ksenija Slankamenac, Rolf Graf e Pierre Alain Clavien: *Assessment of hepatic steatosis by expert pathologists: the end of a gold standard*. Annals of surgery, 250(5):691–697, nov 2009, ISSN 1528-1140 (ELECTRONIC). 64
- [181] Angelini, Annalisa, Claus Boegelund Andersen, Giovanni Bartoloni, Fiona Black, Paul Bishop, Helen Doran, Marny Fedrigo, Jochen W U Fries, Martin Goddard, Heike Goebel, Desley Neil, Ornella Leone, Andrea Marzullo, Monika Ortmann, Francois Paraf, Samuel Rotman, Nesrin Turhan, Patrick Bruneval, Anna Chiara Frigo, Francesco Grigoletto, Alessio Gasparetto, Roberto Mencarelli, Gaetano Thiene e Margaret Burke: *A web-based pilot study of inter-pathologist reproducibility using the ISHLT 2004 working formulation for biopsy diagnosis of cardiac allograft rejection: the European experience*. The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation, 30(11):1214–1220, nov 2011, ISSN 1557-3117 (ELECTRONIC). 64
- [182] Costantini, Massimo, Stefania Sciallero, Augusto Giannini, Beatrice Gatteschi, Paolo Rinaldi, Giuseppe Lanza Nova, Luigina Bonelli, Tino Casetti, Elisabetta Bertinelli, Orietta Giuliani, Guido Castiglione, Paola Mantellini, Carlo Naldoni e Paolo Bruzzi: *Interobserver agreement in the histologic diagnosis of colorectal polyps. the experience of the multicenter adenoma colorectal study (SMAC)*. Journal of clinical epidemiology, 56(3):209–214, mar 2003, ISSN 0895-4356 (PRINT). 64
- [183] Bordes, Florian, Richard Yuanzhe Pang, Anurag Ajay, Alexander C. Li, Adrien Bardes, Suzanne Petryk, Oscar Mañas, Zhiqiu Lin, Anas Mahmoud, Bargav Jayaraman, Mark Ibrahim, Melissa Hall, Yunyang Xiong, Jonathan Lebensold, Candace Ross, Srihari

Jayakumar, Chuan Guo, Diane Bouchacourt, Haider Al-Tahan, Karthik Padthe, Vasu Sharma, Hu Xu, Xiaoqing Ellen Tan, Megan Richards, Samuel Lavoie, Pietro Astolfi, Reyhane Askari Hemmat, Jun Chen, Kushal Tirumala, Rim Assouel, Mazda Moayeri, Arjang Talattof, Kamalika Chaudhuri, Zechun Liu, Xilun Chen, Quentin Garrido, Karen Ullrich, Aishwarya Agrawal, Kate Saenko, Asli Celikyilmaz e Vikas Chandra: *An introduction to vision-language modeling*, 2024. <https://arxiv.org/abs/2405.17247>. 73

- [184] Devlin, Jacob: *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805, 2018. 73