UNIVERSIDADE FEDERAL DO MARANHÃO

UNIVERSIDADE FEDERAL DO PIAUÍ

Doutorado em Ciência da Computação Associação UFMA/UFPI

Caio Eduardo Falcão Matos

Segmentação de Rins, Cistos e Tumores Renais em Imagens de Tomografia Computadorizadas baseada em Módulos de Pirâmide de Pooling e Redes Convolucionais

Orientador: Prof. Dr. Geraldo Braz Junior

Co-orientador: Prof. Dr. Anselmo Cardoso de Paiva

São Luís - MA
Março, 2025

Caio Eduardo Falcão Matos

Segmentação de Rins, Cistos e Tumores Renais em Imagens de Tomografia Computadorizadas baseada em Módulos de Pirâmide de Pooling e Redes Convolucionais

TESE DE DOUTORADO

Tese apresentada como requisito parcial para obtenção do título de Doutor em Ciência da Computação, ao Doutorado em Ciência da Computação, Associação UFMA/UFPI.

Orientador: Prof. Dr. Geraldo Braz Junior
Co-orientador: Prof. Dr. Anselmo Cardoso de Paiva

São Luís - MA
Março, 2025

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Matos, Caio Eduardo Falcão.
Segmentação de Rins, Cistos e Tumores Renais em Imagens de Tomografia Computadorizadas baseada em Módulos de Pirâmide de Pooling e Redes Convolucionais / Caio Eduardo Falcão Matos. - 2025.
104 p.

Coorientador(a) 1: Anselmo Cardoso de Paiva.
Orientador(a): Geraldo Braz Junior.
Tese (Doutorado) - Programa de Pós-graduação Doutorado em Ciência da Computação - Associação UFMA/UFPI, Universidade Federal do Maranhão, São Luís/ma, 2025.

1. Câncer de Renal. 2. Segmentação de Rins, Cisto e Tumores. 3. Módulos de Pirâmide de Pooling. 4. Redes Neurais Convolucionais. 5. Tomografia Computadorizada. I. Junior, Geraldo Braz. II. Paiva, Anselmo Cardoso de. III.

Caio Eduardo Falcão Matos

Segmentação de Rins, Cistos e Tumores Renais em Imagens de Tomografia Computadorizadas baseada em Módulos de Pirâmide de Pooling e Redes Convolucionais

A presente Tese de Doutorado foi avaliada e aprovada por banca examinadora composta pelos seguintes membros:

Prof. Dr. Geraldo Braz Junior
Orientador
Universidade Federal do Maranhão

Prof. Dr. Anselmo Cardoso de Paiva
Co-orientador
Universidade Federal do Maranhão

Prof. Dr. Herman Martins Gomes
Examinador Externo
Universidade Federal de Campina Grande

Prof. Dr. Leandro Augusto Frata Fernandes
Examinador Externo
Universidade Federal Fluminense

Prof. Dr. Antonio Oseas de Carvalho Filho
Examinador Interno
Universidade Federal do Piauí

Prof. Dr. João Dallyson Sousa de Almeida
Examinador Interno
Universidade Federal do Maranhão

Certificamos que esta é a versão original e final da Tese de Doutorado que foi julgada adequada para obtenção do título de Doutora em Ciência da Computação.

gov.br
Documento assinado digitalmente
GERALDO BRAZ JUNIOR
Data: 24/04/2025 20:25:31-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Geraldo Braz Junior
Orientador

Prof. Dr. Anselmo Cardoso de Paiva
Coordenador

São Luís - MA, 20 de Março de 2025

Aos meus pais e esposa.

Agradecimentos

A Deus, primeiramente, pelas graças e bênçãos sempre derramadas em minha vida.

Aos meus pais, por todos os anos de dedicação, sacrifícios e luta, dando o suporte necessário nesta caminhada.

À minha esposa, Aline Klycia, pelo amor incondicional, dedicação, paciência e cuidado ao longo desta jornada. Seu apoio foi fundamental, pois sempre acreditou em mim e incentivou minha trajetória acadêmica desde o início.

Ao meu orientador, Prof. Dr. Geraldo Braz Junior, pela confiança, credibilidade e, sobretudo, paciência dedicadas a mim desde o início desta caminhada, que se iniciou em 2013.1, quando ingressei no Núcleo de Computação Aplicada (NCA/UFMA) ainda na graduação. Sou imensamente grato por seus conselhos, ensinamentos e orientações, assim como pela amizade construída ao longo desses anos.

Ao meu coorientador, Prof. Dr. Anselmo Paiva, pela oportunidade concedida em 2013 de ingressar no NCA, participando do projeto de pesquisa e extensão AGITO. Essa experiência marcou minha primeira experiência na pesquisa e foi, sem dúvida, um divisor de águas em minha trajetória, contribuindo diretamente para minha formação acadêmica.

Aos amigos do Laboratório de Visão Computacional (VIP Lab) — Marcos, Durand, Lucas, Matheus, Thiago, Arthur, João, Joana e Mackele — pela troca de conhecimento, apoio e suporte ao longo desta jornada. Certamente, os cafezinhos e lanches foram fundamentais para fortalecer a amizade construída ao longo do caminho.

Aos amigos de longa data Jefferson, João Bandeira, Giovanni Lucca, Ricardo Marques, Darlan Quintanilha e Weslley Kelson, pelos conselhos, dicas, amizade e todo apoio dado desde o início desta jornada.

Ao Núcleo de Computação Aplicada (NCA) e ao Programa de Pós-Graduação do Doutorado em Ciência da Computação em Associação UFMA e UFPI (DCCMAPI), pelo suporte.

Ao Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA), pelo suporte e liberação para a finalização desta tese.

Aos professores, membros da banca examinadora, pela disponibilidade para avaliação deste trabalho.

A todos os professores e funcionários do DCCMAPI, por todos os ensinamentos e experiências compartilhadas.

"Pressão é um privilégio" (Billie Jean King).

Resumo

A detecção e o diagnóstico precoce do câncer renal desempenham um papel fundamental no prognóstico e tratamento, elevando significativamente as chances de cura e sobrevivência dos pacientes. Nesse contexto, os avanços nos exames radiológicos permitem ao médico especialista realizar, por meio de imagens, a análise e identificação de lesões suspeitas. A Tomografia Computadorizada (TC) torna-se uma ferramenta amplamente utilizada como auxílio no diagnóstico do câncer renal devido à sua capacidade de gerar um grande volume de imagens detalhadas das estruturas internas do corpo, incluindo cistos e tumores renais. Em contrapartida, a análise desses exames de imagem é realizada manualmente pelo médico, observando cada exame, o que torna esse processo suscetível à fadiga e distração. Surge, portanto, a necessidade de aplicar técnicas computacionais para automatizar e melhorar a precisão da segmentação das lesões suspeitas, contribuindo para o diagnóstico precoce do câncer renal. Nos últimos anos, as Redes Neurais Convolucionais (CNNs) ganharam significativa atenção na tarefa de segmentação automática, oferecendo suporte valioso aos especialistas na detecção e diagnóstico em imagens médicas. Nesta pesquisa, desenvolvemos um modelo convolucional que explora os conceitos de multirresolução das pirâmides de *pooling* (*Pyramid Pooling Module* - PPM e Atrous Spatial Pyramid Pooling - ASPP) combinadas à rede UNet para segmentação de rins, cistos e tumores em imagens de tomografia. A metodologia produziu resultados promissores nos conjuntos de dados do desafio de segmentação de tumores renais e renais KiTS21 e KiTS23, exibindo coeficiente de similaridade Dice de 92,32% e 93,78% para rins e massas, 91,38% e 93,06% para massas renais e 87,84% e 91,44% para Tumores, respectivamente.

Palavras-chave: Câncer de rim, Segmentação de rim; Segmentação de cistos e tumores renais, Redes neurais convolucionais, Aprendizado profundo, Módulos de pirâmide de *pooling*, Tomografia computadorizada, Imagens médicas.

Abstract

The early detection and diagnosis of renal cancer play a crucial role in patient prognosis and treatment, significantly increasing the chances of cure and survival. In this context, advances in radiological imaging enable medical specialists to analyze and identify suspicious lesions through imaging techniques. Computed Tomography (CT) has become a widely used tool in aiding the diagnosis of renal cancer due to its ability to generate a large volume of detailed images of the body's internal structures, including renal cysts and tumors. However, the analysis of these images is manually performed by physicians, who must, review each scan individually, making the process susceptible to fatigue and distraction. This highlights the need for computational techniques to automate and enhance the accuracy of suspicious lesion segmentation, contributing to the early diagnosis of renal cancer. In recent years, Convolutional Neural Networks (CNNs) have gained significant attention for automatic segmentation tasks, providing valuable support to specialists in the detection and diagnosis of medical images. In this research, we developed a convolutional model that leverages the multi-resolution concepts of pooling pyramids (Pyramid Pooling Module - PPM and Atrous Spatial Pyramid Pooling - ASPP) combined with the UNet architecture for the segmentation of kidneys, cysts, and tumors in CT images. The proposed methodology produced promising results on the datasets from the Kidney and Kidney Tumor Segmentation Challenge KiTS21 and KiTS23, achieving Dice scores of 92,32% and 93,78% for kidneys and masses, 91,38% and 93,06% for renal masses, and 87,84% and 91,44% for tumors, respectively.

Keywords: Kidney cancer, Kidney segmentation, Kidney cyst and tumor segmentation, Convolutional neural networks, Deep learning, Pyramid Pooling Module, Computed tomography, Medical Images.

Lista de ilustrações

Figura 1 – Semelhanças entre órgãos e estruturas abdominais encontradas em imagens de tomografia voltadas para a segmentação de patologias renais. As amostras foram extraídas das bases de imagens públicas Kits21 (HELLER et al., 2021) e Kits23 (HELLER et al., 2023b). . . . 18
Figura 2 – Estrutura interna do rim. . . . 33
Figura 3 – Aparelho Tomográfico 35
Figura 4 – Neurônio artificial. . . . 38
Figura 5 – *Perceptron* de múltiplas camadas. . . . 39
Figura 6 – Arquitetura padrão das Redes Neurais Convolucionais. . . . 41
Figura 7 – Camada de Convolução. . . . 42
Figura 8 – Camada de subamostragem. Operação de *max pooling*. . . . 42
Figura 9 – Arquitetura padrão da U-Net. . . . 43
Figura 10 – Arquitetura padrão do PPM. . . . 45
Figura 11 – Exemplo de convolução dilatada: (a-c) Taxas de dilatação iguais a 1, 2 e 5. . . . 46
Figura 12 – A estrutura do módulo de pirâmide de pooling espacial dilatado. O módulo é composto por duas etapas: (a) convolução atrous e (b) pooling da imagem, e gera a saída final por meio de uma camada de convolução após a concatenação dos mapas de características. . . . 47
Figura 13 – Arquitetura padrão da DeepLabV3+. . . . 48
Figura 14 – Convolução padrão e convolução separável em profundidade. . . . 49
Figura 15 – Exemplo de matriz de confusão. . . . 50
Figura 16 – Fluxograma da metodologia. . . . 52
Figura 17 – Etapa inicial da anotação: inserção de alfinetes de orientação pelos estudantes de medicina. Exemplo ilustrativo de anotação de tumor renal. 53
Figura 18 – Segmentação final: contorno desenhado com base nos alfinetes de orientação inseridos pelos estudantes de medicina. Exemplo ilustrativo de segmentação de tumor renal. . . . 54
Figura 19 – Exemplo da base KiTS21. (a) fatia do exame e (b) marcação dos rins, tumor e cistos destacadas nas cores verde, vermelho e azul, respectivamente. . . . 55
Figura 20 – Exemplo da base KiTS23. (a) fatia do exame de TC e (b) marcação do rim, tumor e cisto destacadas nas cores verde, vermelho e azul, respectivamente. . . . 56
Figura 21 – Exemplo da aplicação do CLAHE com os respectivos histogramas. . 58
Figura 22 – Arquitetura do bloco PPM 59

Figura 23 – Arquitetura do bloco ASPP 60
Figura 24 – Exemplo de convolução dilatada: (a-d) Imagens de amostra de TC com taxas de dilatação = 1, 6, 12 e 18 61
Figura 25 – Arquitetura PPM-DeepLabv3+ 61
Figura 26 – Arquitetura CPP-UNet 62
Figura 27 – Curvas da função de perda *Dice Categorical Focal Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS21. 69
Figura 28 – Curvas da função de perda *Weighted Categorical Crossentropy Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS21. 69
Figura 29 – Exemplo de segmentação bem-sucedida dos rins e do tumor renal utilizando o modelo PPM-DeepLabv3+. (a) Fatia original do exame 00025; (b) Segmentação efetuada pelo especialista (*GroundTruth*), (c) Segmentação produzida pelo modelo,(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente. 71
Figura 30 – Exemplo de segmentação correta dos rins e do tumor renal utilizando o modelo CPP-UNet.(a) Fatia original do exame 00099; (b) Segmentação efetuada pelo especialista (*GroundTruth*), (c) Segmentação produzida pelo modelo,(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente. 72
Figura 31 – Exemplo de segmentação correta dos rim e do cisto renal utilizando o modelo CPP-UNet;(a) Fatia original do exame 00194; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente. 74
Figura 32 – Segmentação incorreta do rim presente no Exame 00221 a partir do modelo PPM-Deeplabv3+. (a) Fatia original do exame 00221; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente. 75
Figura 33 – Segmentação incorreta do cisto e tumor renal presentes no Exame 00289 a partir do modelo CPP-UNet. (a) Fatia original do exame 00289; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) - (f) Sobreposições dos mapas de ativação dos rins, tumores e cistos à imagem de tomografia computadorizada, respectivamente. 76

Figura 34 – Curvas da função de perda *Dice Categorical Focal Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS23. 79
Figura 35 – Curvas da função de perda *Weighted Categorical Crossentropy Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS23. 80
Figura 36 – Segmentação correta dos rins e tumor renal presentes no Exame 00406 a partir do modelo PPM-DeepLabv3+. (a) Fatia original do exame 00406; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumor renal à imagem de tomografia computadorizada, respectivamente. 82
Figura 37 – Segmentação correta dos rins e tumor renal aplicando o modelo CPP-UNet no Exame 00152. (a) Fatia original do exame 00152; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumor renal à imagem de tomografia computadorizada, respectivamente. 84
Figura 38 – Exemplo de segmentação bem-sucedida do rim e do cisto renal utilizando o modelo CPP-UNet para o Exame com identificador 00175. (a) Fatia original do exame 00175; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação do rim esquerdo e cisto renal à imagem de tomografia computadorizada, respectivamente. . 85
Figura 39 – Exemplo de segmentação incorreta do tumor renal utilizando o modelo PPM-Deeplabv3+ para o Exame com identificador 00475. (a) Fatia original do exame 00475; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumor renal à imagem de tomografia computadorizada, respectivamente. 86
Figura 40 – Segmentação incorreta do cisto renal aplicando o modelo CPP-UNet para o Exame 00060. (a) Fatia original do exame 00060; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins, cisto e tumor renal à imagem de tomografia computadorizada, respectivamente. 87

Lista de tabelas

Tabela 1 – Resumo dos trabalhos relacionados. 30
Tabela 2 – Classes de Avaliação Hierárquica referente ao desafio KiTS. 54
Tabela 3 – Divisão do conjuto de dados para treinamento, teste e validação na base KiTS21. 66
Tabela 4 – Visão geral dos resultados dos modelos de segmentação desenvolvidos (PPM-Deeplabv3+ e CPP-UNet) em comparação com outros modelos aplicados na base KiTS21. 67
Tabela 5 – Divisão do conjunto de dados para treinamento, teste e validação na base KiTS23. 77
Tabela 6 – Visão geral dos resultados dos modelos de segmentação desenvolvidos (PPM-Deeplabv3+ e CPP-UNet) em comparação com outros modelos aplicados na base KiTS23. 78
Tabela 7 – Comparação dos resultados da metologia proposta com os trabalhos relacionados. 88
Tabela 8 – Comparação dos resultados da metodologia proposta com os trabalhos relacionados. 90

Lista de abreviaturas e siglas

AG	*Attention Gate*
ASPP	*Atrous Spatial Pyramid Pooling*
CA	*Contrast Attention*
CAD	*Detecção Auxiliado por Computador*
CADx	*Diagnóstico Auxiliado por Computador*
CAH	*Classes de Avaliação Hierárquica*
CNN	*Convolutional Neural Networks*
DSC	*Coeficiente de Similaridade Dice*
FCN	*Fully Convolutional Network*
FN	*Falso Negativo*
FP	*Falso Positivo*
GCA	*Atenção de Canal Global*
GSA	*Atenção Espacial Global*
JCC	*Índice de Jaccard*
KiTS	*Kidney Tumor Segmentation Challenge*
MLP	*Multilayer Perceptron*
NAS	*Busca por Arquitetura de Rede*
OMS	*Organização Mundial da Saúde*
PPM	*Pyramid Pooling Module*
RCC	*Carcinoma de Células Renais*
ReLU	*Função de Unidade Linear Retificada*
RNA	*Redes Neurais Artificiais*
TC	*Tomografia Computadorizada*
VN	*Verdadeiro Negativo*
VP	*Verdadeiro Positivo*

Sumário

1 INTRODUÇÃO 16
1.1 Definição do Problema 17
1.2 Questões de Pesquisa 19
1.3 Objetivos 20
1.4 Contribuições 20
1.5 Produção Científica 21
1.6 Organização da Tese 22

2 TRABALHOS RELACIONADOS 23
2.1 Métodos de Estágio Único 23
2.2 Métodos em Cascata 26
2.2.1 Considerações finais 31

3 REFERENCIAL TEÓRICO 32
3.1 Rins e Tumores Renais 32
3.2 Tomografia Computadorizada 34
3.3 Processamento de Imagens 35
3.4 Redes Neurais Artificiais 37
3.4.1 Neurônio Artificial 38
3.4.2 *Perceptron* de Múltiplas Camadas (MLP) 39
3.5 Redes Neurais Convolucionais 40
3.6 U-Net 42
3.7 Módulo de Pirâmide de *Pooling* - PPM 43
3.8 Pirâmide de *Pooling* Espacial com Dilatação - ASPP 44
3.9 DeepLabv3+ 48
3.10 Avaliação dos Modelos 50
3.11 Considerações Finais 51

4 METODOLOGIA 52
4.1 Bases de Imagens 52
4.1.1 Base KiTS21 54
4.1.2 Base KiTS23 55
4.2 Pré-Processamento 56
4.3 Modelos de Segmentação 58
4.3.1 Bloco PPM 59
4.3.2 Bloco ASPP 60

4.3.3 Arquitetura Proposta - PPM-DeepLabv3+ . . . 61
4.3.4 Arquitetura Proposta - CPP-UNet . . . 62
4.3.5 Funções de Perda . . . 63
4.4 Considerações Finais . . . 64

5 RESULTADOS . . . 65
5.1 Configurações Experimentais . . . 65
5.2 Experimentos com a base KiTS21 . . . 66
5.2.1 Preparação da Base . . . 66
5.2.2 Resultados . . . 66
5.2.3 Estudos de Casos . . . 70
5.2.3.1 Casos de Sucesso . . . 70
5.2.3.2 Casos de Erro . . . 73
5.3 Experimentos com a base KiTS23 . . . 75
5.3.1 Preparação da Base . . . 76
5.3.2 Resultados . . . 77
5.3.3 Estudos de Caso . . . 81
5.3.3.1 Casos de Sucesso . . . 81
5.3.3.2 Casos de Erro . . . 83
5.4 Comparação com Trabalhos Relacionados . . . 86
5.4.1 Abordagens aplicadas na base KiTS21 . . . 88
5.4.2 Abordagens aplicadas na base KiTS23 . . . 89
5.5 Limitações da Pesquisa e Metodologia . . . 91
5.6 Considerações Finais . . . 92

6 CONCLUSÃO E TRABALHOS FUTUROS . . . 93

REFERÊNCIAS . . . 95

1 Introdução

O câncer é uma condição patológica que está associada a mais de 100 tipos distintos de doenças, caracterizando-se pela proliferação descontrolada de diferentes tipos de células do organismo, com capacidade de acometer órgãos e tecidos adjacentes. Os danos provocados pelo câncer são irreparáveis, que vão além do aspecto financeiro, incluindo consequências físicas e psicológicas dos pacientes. A incerteza e o sofrimento associados a um diagnóstico positivo e seus possíveis prognósticos tornam o enfrentamento do câncer uma tarefa desafiadora. Por outro lado, a medicina demonstra avanços consideráveis nas áreas de prevenção, detecção e diagnóstico do câncer. Parte destes progressos está associada ao desenvolvimento de diversas modalidades de exames preventivos, que possibilitam ao profissional da saúde a avaliação de áreas (SPROUFFSKE et al., 2013).

Medidas como incidência, morbidade hospitalar e mortalidade são fundamentais para o controle da vigilância epidemiológica, permitindo a análise da ocorrência, distribuição e evolução de doenças específicas (Instituto Nacional do Câncer, 2022b). Nesse contexto, o relatório da Organização Mundial da Saúde (OMS), publicado em 2022 (World Health Organization, 2024), estimou aproximadamente 20 milhões de novos casos de câncer e 9,7 milhões de mortes atribuídas à doença naquele ano. Além disso, o estudo apresenta uma projeção alarmante de novos casos de câncer, com previsão de atingir 35 milhões até 2050, representando um aumento de 77% em relação às estimativas de 2022 (World Health Organization, 2024). No Brasil, o Instituto Nacional do Câncer estima o surgimento de cerca de 704 mil casos de câncer por ano até 2025, segundo estudo publicado em 2022 (Instituto Nacional do Câncer, 2022a).

O câncer de rim, também conhecido como câncer renal ou carcinoma de células renais (RCC), corresponde ao 14º câncer mais comum no mundo, com incidência mundial crescente e estimada em 434 mil novos casos a cada ano (World Cancer Research Fund, 2024). Além disso, o câncer de renal possui uma taxa de mortalidade mundial expressiva, se aproximando de 175 mil mortes por ano, de acordo com estudo epidemiológico global do câncer (CIRILLO; INNOCENTI; BECHERUCCI, 2024).

Esta patologia apresenta-se como o terceiro câncer mais frequente do aparelho geniturinário (órgãos genitais e urinários), representando cerca de 3% dos cânceres em adultos, sendo que sua incidência possui probabilidade de 50% maior em homens em relação às mulheres (Hospital Sírio-Libanês, 2024). A incidência estimada no Brasil foi de 10 mil casos ao ano, com taxa bruta de mortalidade de 1,9%, representando um dos cânceres urológicos com maior letalidade (World Health Organization, 2024).

Pesquisas indicam que o crescimento acelerado da incidência de câncer está relacionado ao envelhecimento da população, ao aumento demográfico e à crescente exposição a fatores de risco (World Health Organization, 2024). O desenvolvimento do câncer, especialmente o câncer renal, está associado a diversos fatores que elevam o risco de sua ocorrência, incluindo obesidade, histórico familiar de câncer, tabagismo, hipertensão, entre outros. Entretanto, observa-se que a maioria das ocorrências desse tipo de câncer é marcada por mutações genéticas (DNA) adquiridas ao longo da vida, em vez de serem herdadas (American Cancer Society, 2024c). Desta forma, tanto o tabagismo ativo quanto o passivo desempenham papel preponderante para a causa das mutações adquiridas devido à deterioração de células renais proveniente da exposição aos produtos químicos contidos no tabaco. A obesidade, da mesma forma, destaca-se como fator preocupante ao alterar o equilíbrio da função hormonal no corpo, podendo prejudicar o crescimento normal de diversos tecidos, incluindo os rins (American Cancer Society, 2024b).

O câncer renal, particularmente o carcinoma de células renais (RCC), é o tipo mais comum de câncer renal em adultos, representando cerca de 90% dos diagnósticos. Dentre seus subtipos, o carcinoma de células claras se destaca como o mais frequente e agressivo, correspondendo a 75% dos casos (American Cancer Society, 2024d). Em relação à sua localização, aproximadamente 5% dos tumores renais estão restritos ao rim, 20% são considerados localmente avançados, afetando gânglios regionais próximos, e 25% apresentam metástases, com a doença se espalhando predominantemente para os pulmões, fígado e ossos (PADALA et al., 2020). A taxa média de crescimento dos tumores renais varia significativamente, com estimativas que vão de 0,2 a 6,5 cm por ano, sendo a média em torno de 0,3 cm (GOFRIT et al., 2015). Os sintomas iniciais do carcinoma de células renais (RCC) costumam ser sutis, resultando em diagnósticos tardios em aproximadamente um terço dos casos.

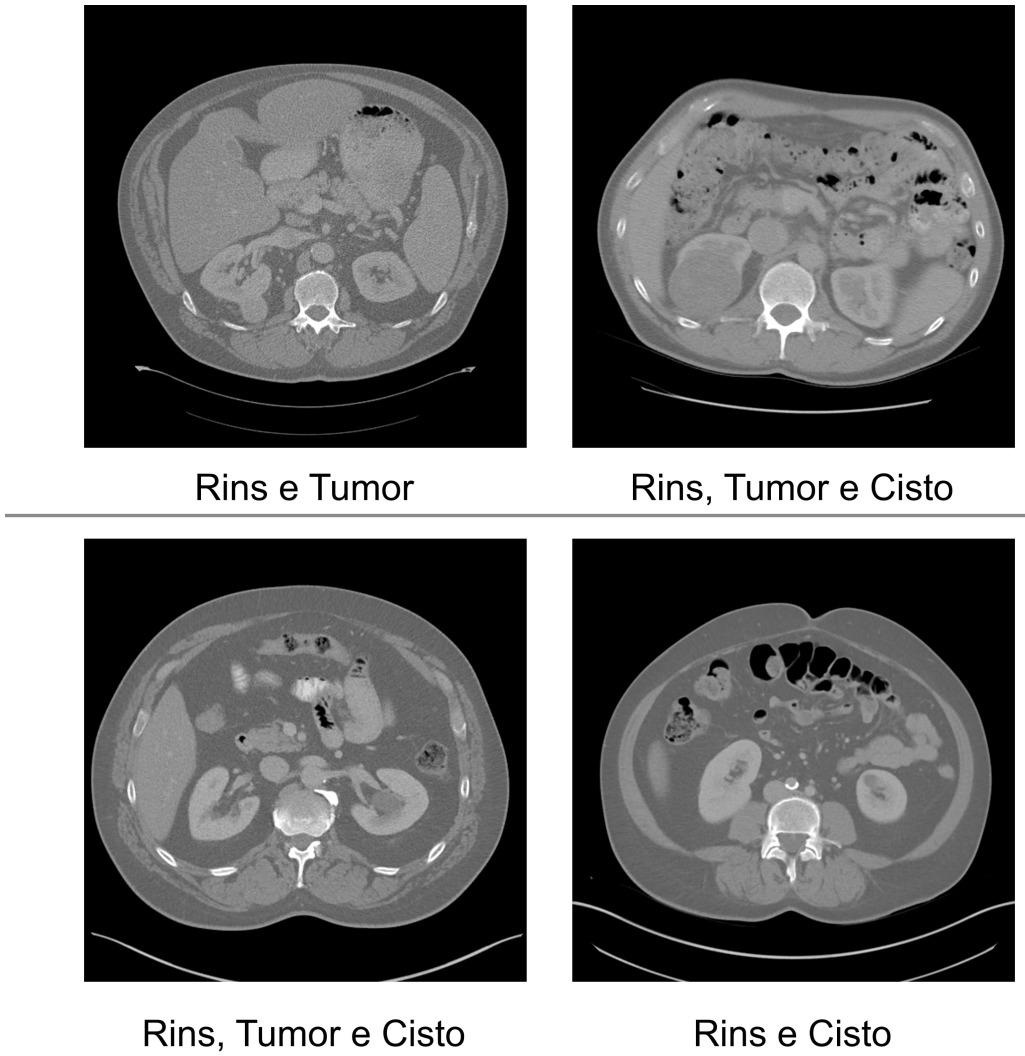
Desta forma, a detecção e o diagnóstico precoce do câncer renal tornam-se essenciais para a construção de estratégias de tratamento eficazes, elevando as chances de cura do paciente. Pacientes diagnosticados em estágio inicial da doença, ou seja, quando não há disseminação da doença, possuem cerca de 93% de taxa de sobrevivência relativa em cinco anos. Por outro lado, quando o diagnóstico é efetuado em estágio avançado (metástase da doença), esta taxa é reduzida para 12%, demonstrando assim a importância deste diagnóstico precoce (American Cancer Society, 2024a).

1.1 Definição do Problema

Os avanços nas técnicas de segmentação automática contribuem para a análise de imagens médicas na identificação dos órgãos e patologias relacionadas. Contudo, desafios persistem e precisam ser explorados para aprimorar a precisão e a

confiabilidade dessas ferramentas. Entre os problemas frequentemente encontrados, destacam-se questões relacionadas ao nível de similaridade obtidos pelas ferramentas, onde baixos valores indicam discrepâncias significativas, evidenciando limitações na identificação de bordas e regiões específicas, especialmente em estruturas com formas irregulares ou contrastes baixos, como tumores e cistos, resultando em elevadas taxas de falsos positivos. Características específicas associadas ao exame de tomografia computadorizada tornam a segmentação semântica ainda mais desafiadora. Dificuldades como a distinção entre estruturas anatômicas semelhantes, a localização dos rins, cistos e tumores em anatomias complexas são fatores que aumentam o risco de erros no diagnóstico. A Figura 1 ilustra exemplos que evidenciam o desafio de diferenciar patologias renais em imagens de tomografia computadorizada.

Figura 1 – Semelhanças entre órgãos e estruturas abdominais encontradas em imagens de tomografia voltadas para a segmentação de patologias renais. As amostras foram extraídas das bases de imagens públicas Kits21 (HELLER et al., 2021) e Kits23 (HELLER et al., 2023b).



Fonte: Elaborado pelo autor.

A análise de grandes volumes de imagens por especialistas é uma tarefa intensiva e repetitiva, que é desafiante devido à possibilidade de fadiga e aumento na suscetibilidade a erros. Nesse contexto, o desenvolvimento de técnicas de segmentação semântica automática surge como uma solução promissora, auxiliando os especialistas na identificação de regiões suspeitas e contribuindo para diagnósticos mais precisos e confiáveis.

Desta forma, o problema abordado nesta tese consiste no desenvolvimento de métodos para a segmentação de estruturas renais, incluindo cistos e tumores, em imagens de tomografia computadorizada (TC). Esses métodos visam aprimorar a identificação dessas estruturas, contribuindo para a maior eficácia no diagnóstico e no planejamento terapêutico do câncer renal.

1.2 Questões de Pesquisa

As questões de pesquisa que direcionaram os objetivos descritos nesta tese são:

- **Q1:** A aplicação de abordagens convolucionais baseadas em pirâmides de pooling são capazes de aumentar a capacidade de extrair características com mais precisão em imagens de CT e tornar a segmentação de áreas nos rins mais precisas?
- **Q2:** O uso combinado de piramides de *pooling* com diferentes tipos de convoluções associados a redes convolucionais pode proporcionar a obtenção de características globais e locais em múltiplas resoluções?
- **Q3:** Uma abordagem que obtém informações em todo o contexto da imagem pode gerar uma segmentação sensível a pequenas variações entre as patologias renais e demais órgãos contidos no exame que possuem características semelhantes?
- **Q4:** O uso de funções de perda com ponderação por classe pode melhorar o desempenho do modelo ao lidar com o desbalanceamento do conjunto de dados do desafio KiTS, caracterizado pela predominância da classe de fundo em relação às demais?

As questões de pesquisa apresentadas nortearam o desenvolvimento desta tese, atuando como guias para a pesquisa na aplicação das abordagens baseadas em pirâmide de *pooling*, na aplicação e avaliação das funções de perda com ponderação por classe para construção da metodologia proposta. Os artefatos implementados buscaram superar os desafios existentes, oferecendo novos contextos e informações mais relevantes para a segmentação semântica de rins, cistos e tumores renais em imagens de tomografia computadorizada (TC).

1.3 Objetivos

O objetivo geral desta tese é desenvolver método computacional capazes de realizar a segmentação automática de rins, cistos e tumores renais em imagens de tomografia computadorizada (TC) de pacientes diagnosticados com câncer renal, buscando alcançar elevados níveis de precisão para viabilizar diagnósticos mais assertivos e confiáveis.

Esta tese concentra-se na investigação das pirâmides de *pooling* como uma ferramenta complementar às redes neurais convolucionais, visando superar as limitações e desafios das redes tradicionais, como: *Multilayer Perceptron*, que apresentam limitações na incorporação de representações globais em diferentes escalas e sub-regiões, voltadas para a segmentação de rins, cistos e tumores renais.

Desta forma, foram estabelecidos os seguintes objetivos específicos:

- Investigar redes convolucionais integradas a módulos de pirâmides de *pooling*, com o intuito de aprimorar a segmentação em imagens de tomografia computadorizada;
- Adaptar técnicas baseadas em pirâmides de *pooling* para melhorar a precisão da segmentação em imagens de tomografia computadorizada;
- Analisar o uso do módulo de pirâmide de *pooling* com dilatação (ASPP) como extrator de características para segmentação de rins, cistos e tumores renais em imagens de tomografia computadorizada, avaliando seu impacto na precisão e robustez da segmentação;
- Explorar estratégias para a combinação de redes convolucionais, amplamente utilizadas na segmentação de imagens médicas, com a abordagem baseada em módulos de pirâmides de *pooling* (PPM e ASPP);
- Desenvolver métodos adaptados de redes neurais convolucionais para realizar a segmentação de rins, cistos e tumores em pacientes com diagnóstico de câncer renal;
- Realizar a validação dos métodos propostos por meio de experimentos aplicados em bases de dados públicas de imagens de tomografia computadorizada amplamente adotadas na literatura;

1.4 Contribuições

De maneira geral, as principais contribuições desta tese são:

- Na construção e aplicação de modelos inovadores de segmentação baseados em redes convolucionais, com foco no contexto do câncer renal. O objetivo principal foi identificar com precisão rins, cistos e tumores renais, utilizando novas arquiteturas convolucionais projetadas especificamente para esse propósito;
- Concepção de metodologia eficiente para a segmentação de rins, cistos e tumores renais, com potencial para integração a sistemas CAD (*Computer-Aided Design*), visando aumentar a produtividade do médico especialista e auxiliar no diagnóstico do câncer renal;
- Na adaptação e combinação de módulos de piramides de *pooling* associados a redes convolucionais, especificamente:
  - Adaptação da:
    * Rede convolucional Deeplabv3+ combinada ao módulo de pirâmide de *pooling* (PPM) com codificador da rede.
    * Rede convolucional U-Net associando diferentes módulos de pirâmide de *pooling*: PPM e *Atrous Spatial Pyramid Pooling* (ASPP) na fase de codificação para extração de características;

1.5 Produção Científica

Durante o desenvolvimento desta pesquisa, foram publicados/aceitos os seguintes artigos científicos relacionados ao tema:

- **Publicado em Revista** - **Caio E. F. Matos**, Geraldo Braz Júnior, João D. Sousa de Almeida, Anselmo Cardoso de Paiva. CPP-UNet: Combined Pyramid Pooling Modules in the U-Net Network for Kidney, Tumor and Cyst Segmentation. IEEE Latin America Transactions, v. 22, p. 642-650, 2024. DOI: 10.1109/TLA.2024.10620387
- **Publicado no SBCAS 2023** - **Caio E. F. Matos**, Marcus Vinicius S. L. Oliveira, João Otávio Bandeira Diniz, Arthur G. S. Fernandes, Geraldo Braz Júnior, Anselmo Cardoso de Paiva. PPM-DeepLab: Modulo de Pirâmide de Pooling como Codificador da rede DeepLabV3+ para Segmentação de Rins, Cistos e Tumores Renais. Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde. SBC, 2023. p. 210-221. DOI: 10.5753/sbcas.2023.229611
- **Publicado em Revista** - Marcus Vinicius S. L. Oliveira, **Caio E. F. Matos**, Geraldo Braz Júnior, Anselmo Cardoso de Paiva, João D. Sousa de Almeida, Gabriel J. S. Costa, Matheus L. L. Bessa & Mario P. Freitas Filho (2023) A PPM-based UNet for Tumour and Kidney Segmentation in CTScans, Computer Methods

in Biomechanics and Biomedical Engineering: Imaging & Visualization, DOI: 10.1080/21681163.2023.2198047

- **Publicado no SBCAS 2024** - João Guilherme Araujo do Vale, Italo Francyles Santos da Silva, **Caio E. F. Matos**, Geraldo Braz Júnior, Marcos Gabriel Mendes Lauande. Redes DenseNet com Mecanismos de Atenção Múltipla aplicadas à Classificação Automática de Câncer Peniano em Imagens Histopatológicas. Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde. Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS). SBC, 2024. p. 495-506. DOI: 10.5753/sbcas.2024.2755

- **Publicado em Evento Internacional** - **Caio E. F. Matos**, João G. A. do Vale, Marcos Melo Ferreira, Geraldo Braz Júnior, João D. Sousa de Almeida (2024). Evaluation of Transfer Learning with a U-Net Architectures for Kidney Segmentation. In: Wireless Mobile Communication and Healthcare. MobiHealth 2023. vol 578. Springer, DOI: 10.1007/978-3-031-60665-6_8

1.6 Organização da Tese

Os demais capítulos desta tese foram estruturados da seguinte forma:

- O Capítulo 2 descreve trabalhos relacionados à problemática da segmentação dos rins, cistos e tumores renais em imagens de tomografia computadorizada.

- O Capítulo 3 trata da fundamentação teórica necessária para a construção da pesquisa proposta. Neste capítulo, são abordados os temas referentes ao câncer renal, tomografias computadorizadas, processamento de imagens, aprendizagem de máquina, métricas de avaliação e arquiteturas de redes neurais utilizadas.

- Os Capítulos 4 e 5 descrevem todas as etapas que compõem a metodologia proposta para esta pesquisa, assim como os resultados obtidos, discussões e comparações com pesquisas correlatas;

- Por fim, o Capítulo 6 apresenta as considerações finais sobre os resultados e propostas de trabalhos futuros.

2 Trabalhos Relacionados

Este capítulo apresenta uma análise da literatura atual voltada para a segmentação de rins, cistos e tumores renais por meio de imagens de tomografia computadorizada. A análise abrange a aplicação de abordagens de aprendizado profundo, não apenas para identificação das estruturas relacionadas ao câncer renal, mas também para adaptações propostas nestas redes para obter informações mais relevantes sobre a patologia.

A segmentação dos rins, cistos e tumores renais em imagens de TC visa auxiliar os especialistas no diagnóstico do câncer renal. Assim, diversas pesquisas investigam a aplicação e desenvolvimento destas tecnologias para a segmentação automática de estruturas renais.

A literatura consultada foi categorizada conforme o tipo de abordagem adotada, podendo ser Métodos de Estágio Único ou em Cascata. Além disso, foram classificados segundo a base de imagem, KiTS21 ou KiTS23, utilizada para validação da metodologia proposta. Um aspecto relevante sobre as bases é o uso de macro classes chamadas de classes de avaliação hierárquica (CAH) que agrupam os resultados para rins, cistos e tumores nas classes Rins e Massas, Massas Renais e Tumores. Estas classes são definidas na Seção 4.1.

2.1 Métodos de Estágio Único

Nesta seção, são apresentados brevemente trabalhos disponíveis na literatura. Que utilizam métodos de estágio único baseado em aprendizado profundo para segmentação de rins, cistos e tumores renais. Esses métodos normalmente usam uma única rede neural que recebe a imagem inteira como entrada e gera as máscaras de segmentação ou rótulos de classe diretamente, sendo essa a vertente explorada por nossa metodologia proposta. Os métodos de estágio único visam simplificar o processo ao prever diretamente os resultados de segmentação sem a necessidade de etapas adicionais ou representações intermediárias. A seguir, tem-se um resumo de alguns desses trabalhos.

No estudo realizado em Shen et al. (2022), foi desenvolvido o modelo COTRNet para a segmentação de patologias renais. Este modelo é baseado na arquitetura de codificador-decodificador, onde o codificador é formado por várias camadas convolucionais entrelaçadas com camadas de *transformer* e operadores de max-pooling. Como contribuição, este trabalho indica que a associação entre uma rede convolucional, pré-treinada com imagenet, em conjunto com a rede transformer produz um modelo capaz de capturar características globais e locais.

Por outro lado, diversas pesquisas utilizam as redes U-Net como base de suas arquiteturas propostas efetuando algum tipo de adaptação. Os trabalhos (HEO, 2022, 2022; YANG et al., 2022; ZANG et al., 2022) utilizam a rede U-Net voltada a imagens tridimensionais (3D) aplicando variações em sua estrutura. A primeira pesquisa aplica a rede U-Net em 3D com conexões adicionais de saltos (*Skip Connections*) baseadas na rede residual. Como contribuição, esta pesquisa indica que a rede U-Net aplicada no contexto em tridimensional e associada a conexões residuais incrementa o desempenho para segmentação no contexto renal.

Além disso, a pesquisa também explora os benefícios das redes residuais na composição da arquitetura da U-Net 3D. No entanto, neste estudo, os blocos residuais são aplicados exclusivamente no codificador da rede, distribuídos em três níveis. A pesquisa propõe a estratégia de sobreposição de blocos (*Overlap-Tile Strategy*) como uma etapa de pré-processamento. A principal contribuição desse estudo encontra-se na aplicação dessa estratégia, visando reduzir as limitações da segmentação em grandes volumes de dados 3D, em conjunto com a U-Net 3D aprimorada por blocos residuais.

Essa contribuição também é observada na quarta pesquisa (YANG et al., 2022), voltada para segmentação utilizando a U-Net 3D. Neste trabalho, os blocos residuais são empregados no codificador da rede, mas com a incorporação da técnica de transferência de aprendizado (*transfer learning*). Para isso, a pesquisa utiliza a Residual 3D U-Net com pesos pré-treinados na base de imagens *Multi-Organ and Tumor Segmentation* (MOTS) (ZHANG et al., 2021) e integra o módulo *Atrous Spatial Pyramid Pooling* ao decodificador. A transferência de aprendizado, por meio da utilização de pesos pré-treinados do conjunto MOTS para as tarefas de segmentação do banco de dados KiTS21 na U-Net 3D, apresenta-se como contribuição desta pesquisa.

Por fim, o último estudo, que utiliza a abordagem da rede U-Net 3D (ZANG et al., 2022), adapta a rede U-Net 3D utilizando o *framework* nnUNet. A contribuição desta pesquisa está compreendida na aplicação deste *framework* para buscar a configuração melhor otimizada da rede U-Net 3D voltada para a segmentação das estruturas e patologias renais. Para tanto, hiperparâmetros como tamanho do *batch*, parâmetros do otimizador, e tamanhos de *patch* e de filtros são otimizados para melhorar a capacidade de generalização da rede.

A rede U-Net tradicional também teve sua arquitetura modificada por alguns trabalhos na literatura. As pesquisas desenvolvidas em (SUN et al., 2022; WU; LIU, 2022) propõem variações da U-Net utilizando mecanismos de atenção para a tarefa de segmentação de rins, cistos e tumores renais. O primeiro trabalho é baseado na variação da rede U-Net denominada DenseU-Net adaptada com o mecanismo de atenção, *Attention Gate* (AG), como *skip connections*. Por outro lado, o segundo trabalho propõe a utilização de outro mecanismo denominado *Contrast Attention* (CA) para efetuar conexões

de salto (*skip connections*) entre codificador e decodificador da rede. Ambos os trabalhos apresentam como principal contribuição a aplicação e adaptação de mecanismos de atenção para potencializar a capacidade da rede na identificação de estruturas renais, incluindo rins, tumores e cistos.

Por fim, a abordagem proposta em Sun et al. (2023) consiste em um modelo denominado 2.5D MFFAU-Net desenvolvido a partir da adaptação da arquitetura U-Net tradicional, inserindo o mecanismo de recursos em múltiplos níveis integrado ao mecânico de atenção. Este mecanismo baseia-se na aplicação do esquema 2.5D onde se obtém as informações tridimensionais das fatias do exame de TC. Como contribuição, esta pesquisa apresenta a associação do mecanismo 2.5D ao bloco ResConv que aplica uma abordagem residual utilizando convoluções com filtros $3x3$.

Dentre os estudos que empregam métodos de estágio único, verifica-se que a literatura também contempla abordagens baseadas na rede U-Net 3D para a segmentação no conjunto de dados KiTS23. Desta forma, os métodos desenvolvidos por Stoica, Breaban e Barbu (2023) e Uhm et al. (2023) exploram a capacidade da rede, efetuando adaptações em sua arquitetura. O primeiro estudo emprega técnicas para a transferência de domínio entre os conjuntos de dados KiPA22 (XIAO; YANG; SONG, 2023) e KiTS23. Para isso, foram aplicadas etapas de pré-processamento baseadas em correspondência de histograma (histogram matching), bem como técnicas de aumento de dados, incluindo rotação, escalonamento, ruído gaussiano e desfoque gaussiano. Essas estratégias foram integradas à arquitetura da U-Net 3D para a tarefa de segmentação. A principal contribuição desta pesquisa consiste na aplicação da transferência de domínio como um mecanismo para aprimorar o desempenho da U-Net 3D, demonstrando seu impacto na melhoria dos resultados obtidos. Por outro lado, a segunda pesquisa baseia-se na aplicação de duas redes U-Net 3D em paralelo, processando a imagem de entrada em diferentes resoluções. Uma das imagens é submetida a uma operação de reamostragem, enquanto a outra mantém seu tamanho original. Ao final do processamento, os autores implementam uma etapa de pós-processamento multiescala, na qual são eliminadas as regiões segmentadas que não apresentam correspondência entre as saídas geradas por ambas as redes. O resultado dessa etapa constitui a segmentação final do modelo proposto. As principais contribuições do estudo incluem o desenvolvimento de abordagens de pós-processamento voltadas para a melhoria do desempenho ao combinar de maneira eficiente as previsões de modelos treinados em diferentes configurações, além da implementação paralela da rede U-Net 3D.

As redes de autoatenção (self-attention) também são exploradas na literatura atual com foco na segmentação de estruturas renais. Os estudos conduzidos por Qian et al. (2023) e Myronenko et al. (2023) utilizam essa abordagem como um componente da rede U-Net tradicional, resultando em uma arquitetura especializada para a tarefa. Na

primeira, uma estrutura híbrida de redes é aplicada associando as redes nnU-Net e Swin Transformer visando capturar características locais e globais das imagens. Por outro lado, a segunda desenvolve a adaptação das redes SegResNet e SwinUNETR integradas à técnica de busca por arquitetura de rede (*Neural Architecture Search* - NAS) denominada de DiNTS. Segundo os estudos, a associação de redes de autoatenção com redes convolucionais oferece benefícios significativos, especialmente em tarefas de segmentação de imagens médicas. Segundo os estudos, a principal contribuição está na associação de redes de autoatenção com redes convolucionais, oferecendo benefícios, especialmente em tarefas de segmentação de imagens médicas. Essa combinação integra a capacidade das redes convolucionais de capturar características locais e estruturais com a habilidade das redes de autoatenção de modelar contextos mais amplos, sem a necessidade de múltiplas camadas convolucionais.

2.2 Métodos em Cascata

Nos métodos em cascata, há múltiplos estágios ou níveis de segmentação aplicados sequencialmente para refinar os resultados de segmentação. Nessa abordagem, a saída de um estágio é usada como entrada para o próximo estágio, permitindo um aprimoramento progressivo na precisão da segmentação. Normalmente são aplicadas sucessivas redes de segmentação semântica para a localização de objetos em imagens. Os trabalhos a seguir desenvolvem esta abordagem para a segmentação de rins, tumores e/ou cistos.

Semelhantemente às pesquisas anteriormente analisadas na seção anterior (2.1 Métodos de Estágio Único), observa-se que, considerando método em cascata, a literatura apresenta diversas pesquisas baseadas na rede U-Net e suas variações para segmentação nas bases KiTS21 e KiTS23. Portanto, os estudos elaborados por Li et al. (2022) e Chen e Liu (2022) aplicam modificações na arquitetura da rede e as aplicam em cascata para segmentar rins, cistos e tumores renais. Especificamente, a primeira pesquisa combina a rede U-Net 3D com a rede ResNet, resultando na arquitetura ResU-Net3D. Em seguida, essa arquitetura é aplicada em etapas sequenciais, no formato de cascata, sendo a primeira dedicada à segmentação dos rins e a segunda voltada para a segmentação de cistos e tumores renais.

Por outro lado, a segunda pesquisa explora o esquema 2.5D aplicado a duas etapas de segmentação. A primeira etapa, principal, utiliza a rede ResSENormUnet, formada pela incorporação de blocos residuais associados ao mecanismo de atenção na arquitetura da U-Net. A segunda etapa, voltada para o refinamento, é composta pela rede DenseTransUnet, resultante de uma modificação da U-Net com a adição de blocos da DenseNet e do Transformer. As principais contribuições de ambas as pesquisas estão nas adaptações das redes estudadas e na aplicação em cascata, de modo que a

segmentação das estruturas renais e suas patologias é realizada em etapas sequenciais, com cada etapa dependendo da anterior.

Abordagens que utilizam a U-Net 3D também são amplamente exploradas na literatura, demonstrando progresso significativos na segmentação das estruturas renais. Nesse contexto, os estudos realizados em (GOLTS et al., 2022), (ZHAO; CHEN; WANG, 2022) e (GEORGE, 2022) investigam o potencial dessa rede, aplicando-a em formato de cascata. Em detalhes, o primeiro estudo aplica a rede em duas etapas para capturar características contextuais espaciais em diferentes resoluções. Na primeira etapa, a rede recebe como entrada uma imagem com resolução reduzida. Em seguida, realiza-se uma operação de aumento de resolução na saída, seguida pela concatenação da imagem no tamanho original. Na segunda etapa, a rede U-Net 3D é aplicada novamente, gerando a segmentação final. As principais contribuições deste estudo incluem a adaptação e aplicação da rede em sequência para segmentação, bem como o uso dessas redes com pesos pré-treinados em uma base de imagens médicas com um objetivo diferente, especificamente a base *Liver Tumor Segmentation* (BILIC et al., 2023).

De maneira semelhante, o segundo estudo utiliza a rede U-Net 3D para desenvolver um método em cascata, realizando a segmentação sequencial dos rins, cistos e tumores renais. Neste estudo, um total de quatro redes U-Net 3D são aplicadas para a segmentação. Inicialmente, a imagem de tomografia computadorizada (TC) é submetida à primeira rede, que realiza a segmentação aproximada da localização dos rins. Em seguida, um recorte da área segmentada é realizado para reduzir o espaço de busca. Posteriormente, uma nova rede é aplicada para refinar a segmentação dos rins na área recortada. A segmentação refinada, juntamente com o recorte, é então submetida a uma terceira rede, que realiza a segmentação dos tumores renais. Por fim, a saída dessa rede é refinada pela aplicação de uma última rede U-Net 3D, resultando na segmentação final da arquitetura.

Esta abordagem foi aplicada de forma análoga no segundo estudo (GEORGE, 2022), no qual são definidos dois estágios. No primeiro estágio, a rede U-Net 3D é treinada com imagens que tiveram sua resolução espacial reduzida, visando delinear as regiões de interesse do rim. No segundo estágio, a rede U-Net 3D é novamente treinada para refinar a segmentação, visando obter detalhes mais precisos sobre os rins, cistos e tumores renais. Os avanços científicos indicados nestes estudos baseiam-se no emprego da rede U-Net 3D no formato em cascata, efetuando o refinamento das regiões de interesse para rins, cistos e tumores renais.

Modificações na arquitetura padrão da rede U-Net tradicional são investigadas por Zheng et al. (2024). Este estudo propõe a adaptação da rede por meio da incorporação de aprendizado competitivo (Adversarial Machine Learning) aplicado em múltiplas escalas, resultando na rede MSALDS-UNet. A principal motivação dessa modificação é

melhorar o desempenho da rede tradicional na segmentação de rins, cistos e tumores, utilizando aprendizado em diferentes escalas. Especificamente, a rede discriminadora (*Discriminator Network*) é aplicada ao decodificador da rede U-Net, promovendo aprendizado competitivo. Adicionalmente, a rede MSALDS-UNet é empregada em cascata para realizar a segmentação sequencial dos rins, cistos e tumores. Como contribuição, o estudo aponta que a combinação do aprendizado competitivo com o decodificador da U-Net permite o refinamento das características antes da operação de aumento de resolução (*Upsampling*), resultando em uma segmentação mais precisa.

Abordagens em cascata também são exploradas na literatura para a tarefa de segmentação na base KiTS23. Nesse contexto, os estudos conduzidos por Wang et al. (2023), Li, Peng e Zhang (2023) e Salahuddin et al. (2023) investigam adaptações da rede nnU-Net aplicadas em um esquema de segmentação em múltiplas etapas. Em detalhes, o primeiro estudo divide o processo de segmentação em dois estágios, onde o primeiro a rede nnU-Net é aplicada a um contexto inicial para extração de regiões de interesse na imagem, com foco na segmentação dos rins. No segundo estágio, a rede é reaplicada com base nos resultados obtidos na etapa anterior, visando refinar a segmentação e aprimorar a identificação de tumores e cistos renais.

Por outro lado, o estudo desenvolvido por Li, Peng e Zhang (2023) propõe uma adaptação da rede nnU-Net, incorporando um mecanismo de atenção residual. Essa modificação é destacada como a principal contribuição do estudo, uma vez que os autores argumentam que esse mecanismo pode aprimorar o treinamento da rede, permitindo a extração de características mais relevantes para a segmentação de rins, cistos e tumores. O estudo sugere que a aplicação da rede em um formato de cascata favorece a identificação progressiva das estruturas renais, especializando a nnU-Net na segmentação de cada um dos objetos de interesse.

O último estudo, Salahuddin et al. (2023), também emprega a rede nnU-Net em uma abordagem em cascata, estruturada em duas etapas. Inicialmente, a rede é utilizada para localizar e realizar a segmentação preliminar dos rins. Em seguida, uma bounding box é aplicada para abranger ambos os rins, direito e esquerdo. Essas regiões segmentadas são então recortadas e processadas novamente pela rede nnU-Net, que executa a segmentação multiclasse. O treinamento adota uma estratégia de taxa de aprendizado cíclica, na qual os pesos finais do modelo são extraídos em torno de múltiplos máximos locais, permitindo a criação de um *ensemble* de predições para cada uma das classes identificadas. A redução da dimensionalidade dos dados, estimativa de incerteza e modelo em ensemble de treinamento são destacados neste estudo como principais contribuições. Além disso, destaca-se que a aplicação dessas técnicas em formato de cascata contribui de maneira significativa para a segmentação, especialmente no que se refere a tumores e cistos renais. Outro ponto importante ressaltado no estudo é

o potencial da abordagem para a detecção de falsos positivos.

Seguindo o levantamento das pesquisas relacionadas, é possível identificar que as estratégias em cascata propostas por Hu e Peng (2023), Vispi (2023) e Chen e Zhang (2023) modificam da arquitetura padrão da rede U-Net, substituindo seu codificador convencional por outras redes convolucionais. No primeiro estudo, a proposta inclui a aplicação de blocos residuais associados à rede U-Net 3D, para identificar e filtrar as regiões de interesse que contêm os rins. Posteriormente, o estudo apresenta a arquitetura denominada Global Spatial Channel Attention Network (GSCA-Net), que incorpora mecanismos de atenção espacial global (GSA) e atenção de canal global (GCA), visando à segmentação dos cistos e tumores renais. O avanço científico principal desta pesquisa reside na combinação em cascata das redes U-Net 3D e GSCA-Net. Segundo os autores, essa associação proporciona uma maior robustez à metodologia, aprimorando a capacidade de diferenciação das características das estruturas renais. Isso se destaca especialmente na segmentação de tumores e cistos, onde as variações intra-classe são consideráveis e exigem uma abordagem mais refinada para garantir uma segmentação eficiente.

A segunda pesquisa, Vispi (2023), adota a rede EfficientNet-B5 como codificador da rede U-Net tradicional. Neste contexto, a rede foi treinada por meio de etapas estruturadas em cascata, com a segmentação dos rins, cistos e tumores realizada sequencialmente. Além disso, a otimização de parâmetros foi realizada utilizando a abordagem GridSearch, a fim de identificar as melhores configurações para a rede. Como principais contribuições, são destacadas tanto a adaptação da arquitetura quanto a aplicação da rede em cascata. O estudo sugere ainda que a capacidade da rede convolucional U-Net pode ser ampliada quando associada à EfficientNet-B5, a qual demonstra competência no processamento de imagens em alta resolução. Essa característica é particularmente relevante para a segmentação de imagens médicas, que frequentemente demandam um nível elevado de detalhamento para a identificação de estruturas irregulares, como tumores e cistos.

A última pesquisa analisada, Chen e Zhang (2023), propõe a adaptação da rede U-Net por meio da inserção de blocos residuais em seu codificador, resultando na construção da rede ResUNet. Além disso, etapas de pré-processamento e pós-processamento são implementadas, incluindo estratégias de identificação e recorte das regiões de interesse, realizadas após a aplicação sucessiva da rede. A abordagem em cascata apresentada neste estudo é estruturada em duas etapas: a primeira realiza a segmentação inicial, enquanto a segunda executa a segmentação refinada, abrangendo rins, cistos e tumores renais. Dentre as contribuições destacadas, está a aplicação das estratégias de pré e pós-processamento em conjunto com a rede ResUNet. O estudo ressalta que os blocos residuais facilitam a extração de características mais sensíveis a

variações, especialmente quando combinados com a rede U-Net e aplicados a regiões de interesse previamente identificadas.

A Tabela 1 contém uma síntese das pesquisas relacionadas. Ao analisá-la, observa-se que a maioria dos métodos propostos utiliza variações na arquitetura da rede convolucional U-Net (RONNEBERGER; FISCHER; BROX, 2015), evidenciando o potencial das redes neurais convolucionais baseadas em estruturas de codificador-decodificador. Contudo, esses métodos, ao utilizar convoluções convencionais, apresentam limitações na identificação de características globais das imagens. Nesse sentido, a adaptação dessas redes a novos mecanismos e estruturas convolucionais, como: blocos residuais, blocos pirâmides de *pooling* e mecanismos de atenção, surge como uma alternativa promissora para ampliar sua eficácia na tarefa de segmentação semântica em imagens médicas. Analisando o estado da arte, percebe-se que essas modificações incorporam características de redes utilizadas em contextos não médicos, integrando-as às redes de segmentação. Essa abordagem aprimora a identificação de estruturas e órgãos em imagens de tomografia computadorizada, contribuindo para o diagnóstico mais eficiente do câncer renal.

Tabela 1 – Resumo dos trabalhos relacionados.

	Trabalho	Técnica(s)	Pacientes	Coeficiente Dice			
				Rins e Massas	Massas	Tumor	Composição
Métodos de Estágio Único		**Base KiTS21**					
	Shen et al. (2022)	COTRNet	300	92,30%	55,30%	50,6%	66,07%
	Heo (2022)	3D U-Net	300	93,69%	78.83%	75,03%	84,36%
	Adam et al. (2022)	3D U-ResNet	300	95,10%	79,80%	78,80%	84,57%
	Zang et al. (2022)	3D U-Net	300	97,00%	85,10%	81,90%	88,00%
	Sun et al. (2022)	DenseU-Net com AG	300	97,10%	88,30%	81,50%	88,97%
	Yang et al. (2022)	Residual 3D U-Net com ASPP	210	97,30%	87,40%	82,20%	88,97%
	Wu e Liu (2022)	DenseU-Net com CA	300	96,93%	87,60%	85,09%	89,87%
	Sun et al. (2023)	2.5D MFFAU-Net	300	97,30%	88,70%	87,30%	91,10%
		Base KiTS23					
	Qian et al. (2023)	nnU-Net e Swin Transformer	489	-	-	68,70%	80,10%
	Stoica, Breaban e Barbu (2023)	3D U-Net e Data Augmentation	599	94,70%	76,00%	71,30%	80,67%
	Myronenko et al. (2023)	SegResNet e SwinUNETR	599	95,60%	79,20%	75,80%	83,53%
	Uhm et al. (2023)	3D U-Net e Pós-processamento	599	97,90%	82,6%	85,70%	88,73%
Métodos em Cascata		**Base KiTS21**					
	Chen e Liu (2022)	ResSENormUnet e DenseTransUnet	300	-	-	77,79%	-
	Li et al. (2022)	Residual e 3D U-Net	300	-	-	81,00%	-
	Golts et al. (2022)	3D U-Net	300	97,15%	87,90%	86,38%	90,48%
	Zhao, Chen e Wang (2022)	nnUNet	300	97,52%	88,51%	86,93%	90,99%
	George (2022)	3D U-Net	300	97,48%	88,13%	87,10%	90,90%
	Zheng et al. (2024)	MSALDS-UNet	300	96,10%	84,40%	87,40%	89,30%
		Base KiTS23					
	Wang et al. (2023)	nnU-Net	599	86,60%	54,50%	49,00%	63,37%
	Li, Peng e Zhang (2023)	nnU-Net + Residual Attention	599	93,60%	-	67,00%	-
	Hu e Peng (2023)	Residual e 3D U-Net	599	93,30%	74,40%	67,90%	78,53%
	Vispi (2023)	EfficientNet-B5 + UNet	599	97,71%	81,39%	73,81%	84,30%
	Salahuddin et al. (2023)	nnUNet com CLR	599	94,00%	86,50%	83,50%	88,00%
	Chen e Zhang (2023)	2.5D ResUnet	489	89,44%	85,85%	85,91%	87,07%

A integração de pirâmides de *pooling* às redes convolucionais, especialmente nas arquiteturas DeepLabv3+ (CHEN et al., 2018) e U-Net, surge como uma solução

promissora, capaz de superar as limitações dessas redes. Essa abordagem aprimora a capacidade das redes em identificar informações tanto globais quanto locais, particularmente no contexto da segmentação de rins, cistos e tumores.

2.2.1 Considerações finais

Neste capítulo, realizou-se uma revisão de literatura relacionada à segmentação de rins, cistos e tumores renais em imagens de TC. Apresentaram-se concisamente os métodos e resultados obtidos nesses estudos. Além disso, foram exploradas as diferenças entre as abordagens apresentadas, a fim de proporcionar uma visão abrangente do progresso realizado pela comunidade científica nesse campo. No próximo capítulo, os conceitos teóricos fundamentais necessários para o desenvolvimento desta pesquisa serão abordados.

3 Referencial Teórico

Neste capítulo, é fornecida a base teórica que fundamenta este trabalho. Desta forma, um enfoque geral é apresentado, de modo a tornar compreensível a teoria relacionada ao objeto de estudo, rins, cistos e tumores renais, bem como, as técnicas utilizadas na metodologia proposta para segmentação destas informações em imagens de tomografia computadorizada.

3.1 Rins e Tumores Renais

Os rins são pequenos órgãos retroperitoneais[1] com formato de feijão (oval), coloração marrom-avermelhada, localizados um de cada lado da coluna vertebral, próximos à 10ª e 12ª costelas torácicas. Eles pesam em média 125-170 gramas nos homens e 115-155 gramas nas mulheres, tendo aproximadamente 10 cm de comprimento, 5 cm de largura e 2,5 de espessura. Além disso, os rins se relacionam com os demais órgãos abdominais como estômago, baço, pâncreas, jejuno e cólon descendente e fígado. Devido à sua posição anatômica, o rim direito geralmente está localizado aproximadamente 2,5 cm abaixo do rim esquerdo, uma característica relevante a ser considerada em procedimentos cirúrgicos (DANTAS, 2023).

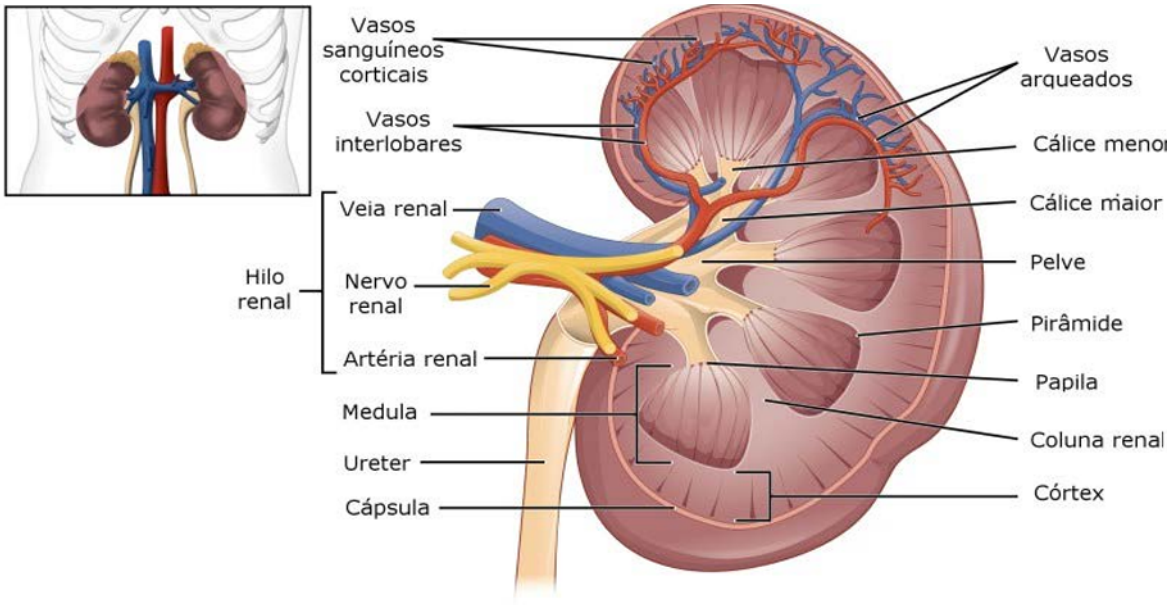
Anatomicamente, os rins possuem estrutura única, conforme ilustrado na Figura 2. Os rins são compostos por duas regiões internas distintas: o córtex e a medula renal. O córtex renal é a camada mais externa e é nela que se encontram os néfrons, as unidades funcionais dos rins responsáveis pela formação da urina. Estima-se que um rim humano possua cerca de 1 milhão de néfrons, cada um deles formado por inúmeros glomérulos que desempenham um papel ativo na filtração do sangue (NETTER; FRANK, 2000).

Por outro lado, a medula renal é composta por 8-18 estruturas em formato de cunha chamadas de pirâmides renais. Essas pirâmides renais possuem ápices que convergem, formando as papilas renais. As papilas renais projetam-se em cálices menores por meio dos ductos papilares, transportando a urina produzida até um cálice renal maior. Esses cálices renais, por sua vez, conectam-se na pelve renal, uma porção dilatada do ureter, responsável por conduzir a urina até o seu destino através das vias urinárias (NETTER; FRANK, 2000; SILVERTHORN, 2010).

Os rins possuem duas margens: uma côncava e uma convexa. Na margem côncava, encontra-se o hilo renal, região responsável pela entrada e saída de vasos sanguíneos, nervos e o ureter. O sangue chega aos rins através da Artéria Renal, que se

[1] Espaço anatômico localizado atrás da cavidade abdominal.

Figura 2 – Estrutura interna do rim.



Fonte: Adaptado de (DANTAS, 2023).

ramifica para fornecer sangue aos capilares glomerulares, onde ocorre o processo de filtração. Após a filtração, o sangue converge para as arteríolas eferentes, que se transformam em capilares peri-tubulares e, por fim, em veias pequenas que se unem para formar as veias renais. Essas veias deixam o rim através da região do hilo renal (GUYTON; J.E., 2017; SanarMED, 2023).

Os rins são órgãos excretores e reguladores que desempenham um papel fundamental na fisiologia humana. Sua principal função é eliminar o excesso de água e metabólitos do organismo, além de controlar o volume de líquidos corporais, contribuindo para a manutenção da homeostase, ou seja, o equilíbrio interno do corpo. Além disso, desempenham múltiplas funções, tais como: filtragem e descarte de substâncias indesejadas, como resíduos metabólicos (ureia, creatinina, ácido úrico) e substâncias exógenas, como medicamentos e outros produtos químicos não originados do metabolismo corporal, regulação do equilíbrio de água e eletrólitos, regulação do pH do corpo, regulação da pressão arterial e produção de hormônios (NefroClínicas, 2023; GUYTON; J.E., 2017; SBN - Sociedade Brasileira de Nefrologia, 2023).

Diversas enfermidades podem comprometer a função dos rins, incluindo a nefrite (aguda ou crônica), cálculos renais, insuficiência renal, pielonefrite crônica e nefropatia diabética. Além disso, quando células renais sofrem uma transformação maligna e se multiplicam descontroladamente, pode ocorrer a formação de tumores renais. Os tumores renais podem ser classificados segundo o tecido de origem ou patologia, podendo ser benignos ou malignos. Os benignos apresentam comportamento mais controlado, tendo sua proliferação mais lenta, enquanto os malignos normalmente são mais agressivos,

podendo até mesmo acometer outros órgãos através da corrente sanguínea, ocasionando metástase rápida. O câncer de rim mais comum é denominado de carcinoma de células renais (ARAP; COELHO, 2023).

3.2 Tomografia Computadorizada

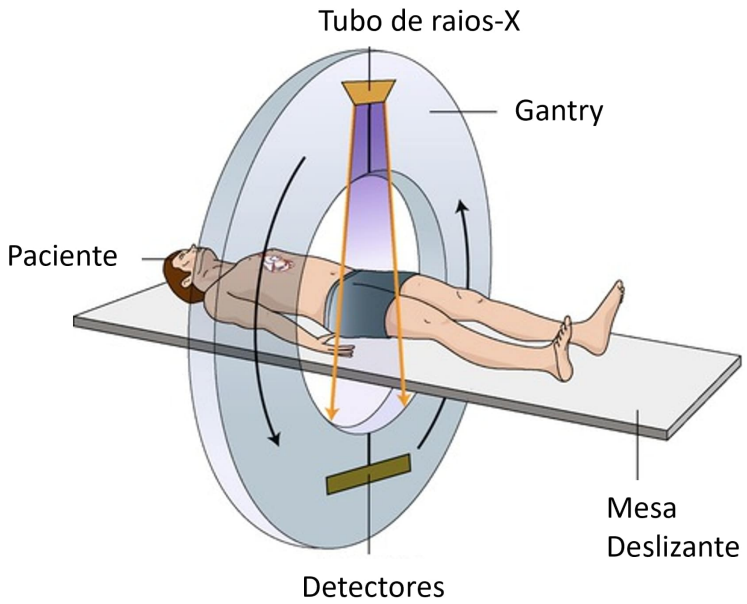
A tomografia computadorizada (TC) é um método não invasivo e indolor que permite aos médicos visualizarem estruturas internas do paciente com maior precisão em relação às radiografias convencionais. A primeira máquina de tomografia computadorizada, denominada tomógrafo, foi desenvolvida em 1972 pelos físicos Allan M. Cormack e Godfrey Hounsfield tornando-se um marco para a medicina moderna. Por esta inovação, foi concedido o Prêmio Nobel de Fisiologia ou Medicina em 1982. O primeiro tomógrafo computadorizado do mundo foi utilizado no Hospital Atkinson Morley, em Londres (PETRIK et al., 2006).

O tomógrafo, ou aparelho tomográfico, é constituído por vários componentes, incluindo o *Gantry*, tubo de raios-x, Detectores e Mesa deslizante, como pode-se observar a Figura 3a. O *Gantry* é a estrutura circular onde são abrigados os componentes eletro-mecânicos que farão a coleta dos dados durante o exame. Normalmente, sua abertura possui a variação de 70 a 90 cm de diâmetro (PELBERG; PELBERG, 2015). Durante o exame, o paciente é posicionado na mesa deslizante que se move através do *Gantry*. Em seguida, ocorre a rotação do tubo de raios-x e detectores em torno do paciente, quando, simultaneamente, são disparados feixes de raios-x capturados pelos detectores. Conforme o aparelho efetua a rotação, diversas imagens são geradas, resultando em imagem de seção transversal completa do paciente. Por fim, a densidade dos tecidos atenua o feixe de raio-x que, por sua vez, atinge com diferentes intensidades os detectores. Após esse processo, as informações obtidas são transformadas para o sistema computacional, produzindo assim a imagem de TC (NIH - Nacional Institute of Biomedical Imaging and Bioengenineering, 2025).

O exame de TC possibilita a obtenção de imagens de ”fatias” ou ”cortes” do corpo do paciente, tornando-se uma das principais ferramentas para o diagnóstico por imagem. Comparada à radiografia tradicional, a TC demonstra maior precisão e sensibilidade, pois utiliza mecanismo de captura dinâmico para a produção das imagens, além de obter imagens sem sobreposição de tecidos (NIH - Nacional Institute of Biomedical Imaging and Bioengenineering, 2025; Johns Hopkins Medicine, 2023). Além disso, este exame é excelente para detecção de anomalias renais, incluindo angiomiolipomas[2] e adenocarcinomas renais. A maioria dos cânceres renais é detectada incidentalmente com imagens de TC (SILVERMAN et al., 2018).

[2] Tipo de tumor renal benigno

Figura 3 – Aparelho Tomográfico



Fonte: Adaptado de Pelberg e Pelberg (2015), Univen (2025).

Os tecidos densos, como os ossos, têm maior capacidade de absorver raios-x, o que resulta em uma aparência mais clara nas imagens de TC. Por outro lado, tecidos menos densos, como a gordura e os órgãos internos, absorvem menos radiação e, consequentemente, aparecem em tons mais escuros na imagem. Como forma de padronização, foi desenvolvida a escala numérica expressa em Unidade de Hounsfield (HU) que, por sua vez, relaciona os diferentes níveis de cinza adquiridos no processo de emissão dos raios-x, em valores numéricos. Esta escala possui variação acompanhando a densidade dos tecidos, sendo -1000 HU para o ar, 30 HU para os rins, podendo chegar a +3000 HU para os ossos (CIERNIAK, 2011).

3.3 Processamento de Imagens

O processamento digital de imagens é uma disciplina que engloba um conjunto de técnicas computacionais utilizadas para modificar, analisar e interpretar imagens digitais. Nesse contexto, uma imagem de entrada é processada de acordo com algoritmos específicos, resultando em uma imagem de saída, que pode passar por melhorias visuais,

correção de imperfeições, extração de informações relevantes ou transformações geométricas. Além de abordar aspectos visuais e estruturais da imagem, o processamento digital de imagens busca fornecer elementos que permitam uma interpretação tanto visual quanto computacional da mesma (GONZALEZ; WOODS, 2010).

Segundo Gonzalez e Woods (2010), pode-se definir uma imagem em duas dimensões (2D) como função bidimensional, $f(x, y)$, em que x e y são coordenadas espaciais (plano), e a amplitude de f em qualquer par de coordenadas (x, y) é chamada de intensidade ou nível de cinza da imagem nesse ponto. Portanto, denomina-se de imagem digital quando x,y e os valores de intensidade de f são quantidades finitas e discretas. As imagens digitais são compostas por um número finito de elementos, onde cada um possui localização e valor específicos. Esses elementos são denominados de elementos pictóricos ou *pixels*.

Os avanços tecnológicos ocorridos nas últimas décadas, tanto em termos de hardware quanto de software, têm possibilitado um amplo espectro de aplicações na área de processamento digital de imagens. Como resultado, inúmeras pesquisas têm sido desenvolvidas visando aprimorar e expandir ainda mais as ferramentas que fundamentam o processamento de imagens. Além disso, conceitos comumente utilizados no processamento de imagens têm sido amplamente adotados em diversos domínios, como indústria, segurança, física, arqueologia, biologia entre outros, contribuindo para a análise, compreensão e manipulação de dados visuais de forma eficiente e precisa (FILHO; NETO, 1999).

A área médica é um dos campos que tem sido amplamente beneficiado pelo crescente número de pesquisas no âmbito do processamento de imagens médicas, especialmente no desenvolvimento de sistemas de auxílio ao diagnóstico (CAD/CADx). Estes avanços, foram impulsionados pela modernização, redução de custos e acessibilidade dos exames radiológicos, sobretudo a tomografia computadorizada (TC) (FILHO; NETO, 1999; GONZALEZ; WOODS, 2010; PEDRINI; SCHWARTZ, 2008).

Um sistema clássico de processamento digital de imagens pode ser estruturado em cinco etapas distintas segundo Gonzalez e Woods (2010): aquisição de imagens, pré-processamento, segmentação, representação e descrição e reconhecimento de padrões.

Na primeira etapa, ocorre a aquisição das imagens, em que estas são capturadas por dispositivos ou sensores e convertidas em formato digital para posterior processamento. Na etapa seguinte, o pré-processamento visa aplicar técnicas e algoritmos que melhorem os aspectos visuais e estruturais das imagens. Como resultado dessa etapa, são obtidos benefícios como a eliminação de ruídos, o aumento do contraste e o realce de estruturas, entre outros.

Na etapa de segmentação, ocorre a divisão da imagem em objeto(s) ou regiões distintas e fundo com base em características de cor, textura, intensidade, entre outros. Em seguida, a etapa de representação e descrição, também denominada por extração de características, visa a identificação de informações específicas da imagem, ou seja, representar por meio de valores uma imagem ou partes dela.

Por fim, tem-se a etapa de reconhecimento de padrões. Nesta etapa, ocorre o processo de identificação e classificação de objetos ou padrões da imagem com base em característica previamente extraídas na etapa anterior. Técnicas de aprendizado de máquina, sobretudo, aprendizagem profunda são amplamente utilizadas para discernir entre possíveis padrões contidos nas imagens.

3.4 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs), ou simplesmente redes neurais, são sistemas e métodos computacionais avançados para aprendizado de máquina, representação de conhecimento e otimização das respostas de sistemas complexos (CHEN et al., 2019). As redes neurais são capazes de adquirir conhecimento a partir de exemplos e dados, permitindo que aprendam padrões e relações complexas de forma automatizada. Essa capacidade de aprendizado permite que as redes neurais sejam aplicadas em diversas áreas, maximizando as saídas de sistemas complexos e promovendo avanços em diversas áreas de pesquisa e desenvolvimento (HAYKIN, 2001).

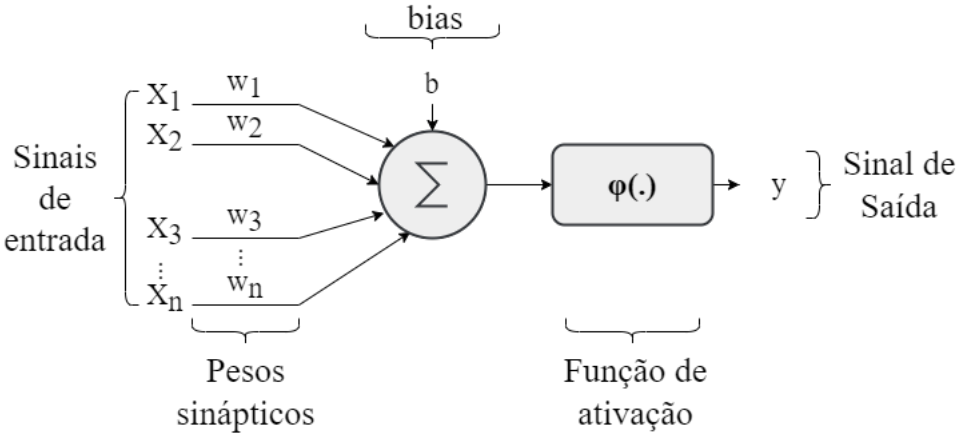
A Rede Neural Artificial (RNA) é um modelo de processamento de dados inspirado no funcionamento do sistema nervoso biológico, como o cérebro (LIVINGSTONE, 2008). A ideia central das redes neurais artificiais, inspirada na dinâmica do cérebro, é realizar o processamento de dados e informações, possibilitando o aprendizado e a indução de conhecimento a partir de dados. Esse conceito é fundamentado no funcionamento do sistema neural biológico, que, por meio da interconexão de bilhões de neurônios, processa informações de entrada e saída. Esse processo complexo permite a interpretação e o processamento eficiente de dados, de forma semelhante ao que ocorre no cérebro humano (GERVEN; BOHTE, 2017).

Redes neurais podem ser descritas como conjuntos de unidades de processamento chamadas neurônios, conectados por interconexões conhecidas como sinapses artificiais (LIVINGSTONE, 2008). As redes neurais artificiais (RNA) têm sido amplamente empregadas em diversas aplicações de reconhecimento de padrões, incluindo reconhecimento de voz, detecção de objetos, identificação de tumores e muito mais (KOVÁCS, 2002).

3.4.1 Neurônio Artificial

Em 1943, McCulloch e Pitts foram responsáveis pela elaboração de um trabalho pioneiro que resultou na primeira aplicação das Redes Neurais Artificiais (RNA), introduzindo a concepção do neurônio artificial (MCCULLOCH, 1943). O neurônio artificial consiste em uma representação matemática simplificada de um neurônio biológico no cérebro humano sendo este fundamental para a construção de modelos mais complexos. A Figura 4 ilustra um neurônio artificial simplificado.

Figura 4 – Neurônio artificial.



Fonte: Adaptado de Kovács (2002)

Matematicamente, um neurônio artificial é descrito de acordo com Equação 3.1,

$$f(x) = \varphi\left(\sum_{i=1}^{n} x_i w_i + b\right), \tag{3.1}$$

onde $x_1, x_2, ..., x_n$ representam os sinais de entrada; $w_1, w_2, ..., w_n$ são os pesos sinápticos associados à entrada i, b é o termo bias e φ é a função de ativação. Desta forma, cada neurônio combina suas entradas e, em seguida, passa por uma função de ativação, que pode ser função identidade, sigmoide ou tangente hiperbólica. Além destas, as funções não-lineares *Rectified Linear Units* (ReLU) e *Leaky Rectified Linear Unit* (Leaky ReLU) demonstram em experimentos convergência em até 6 vezes mais rápida em relação às outras funções anteriormente citadas (MAAS et al., 2013; HINTON et al., 2012). Estas funções são expressas conforme as Equações 3.2 e 3.3:

$$\sigma(x) = max(0, x), \tag{3.2}$$

$$\sigma(x) = max(0.1, x). \tag{3.3}$$

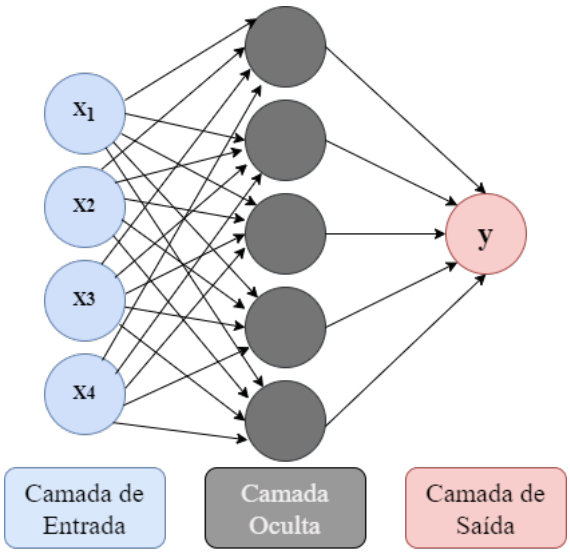
3.4.2 *Perceptron* de Múltiplas Camadas (MLP)

Dentre as diversas arquiteturas de Redes Neurais Artificiais (RNA) existentes, as MLPs (*Perceptron* de Múltiplas Camadas) são, indiscutivelmente, as mais amplamente exploradas na literatura (BANSAL, 2006).

As redes MLPs consistem em um tipo de arquitetura de rede neural artificial em múltiplas camadas de neurônios, incluindo uma camada de entrada, uma ou mais camadas intermediárias (também conhecidas como camadas ocultas) e uma camada de saída.

Os neurônios presentes na camada oculta permitem que a rede neural aprenda tarefas complexas, pois eles extraem gradualmente as características mais relevantes dos padrões de entrada. A Figura 5 exemplifica uma arquitetura básica de MLP com apenas uma camada oculta.

Figura 5 – *Perceptron* de múltiplas camadas.



Fonte: Elaborado pelo autor.

O algoritmo de retropropagação (*backpropagation*) é amplamente utilizado para o treinamento de Redes Neurais Multilayer Perceptron (MLP) (VORA; YAGNIK; SCHOLAR, 2014). Este algoritmo é composto por duas etapas. Na primeira, apresenta-se um padrão à camada de entrada da rede que é repassado camada a camada, até a produção da resposta na camada de saída. A segunda etapa, por sua vez, compara a saída obtida à saída esperada para o padrão informado. Caso não esteja correta, calcula-se o erro. Em sequência, o erro é propagado a partir da camada de saída em direção à camada de entrada, sendo os pesos das conexões das unidades das camadas ocultas atualizados conforme o erro é retropropagado (LI et al., 2012).

O processo de aprendizagem do algoritmo *backpropagation* envolve os seguintes

passos:

1. Inicialização dos pesos sinápticos da rede com pequenos valores aleatórios;
2. Propagação direta (*feedforward*):
   a) Apresentar exemplos de treinamento à entrada da rede, ou seja, um vetor $x_1, x_1, ..., x_N$ de características e especificar um vetor $d_1, d_2, ..., d_N$, de saídas esperadas;
   b) Calcular as saídas reais da rede $y_1, y_2, ..., y_N$, baseadas na Equação 3.1;
3. A atualização dos pesos na rede neural é efetuada pelo algoritmo recursivo iniciando pelos neurônios da camada de saída, em direção à camada de entrada, conforme a Equação 3.4,

$$w_{ij}(t+1) = w_{ij}(t) + \eta\delta_j x_i, \tag{3.4}$$

   onde w_{ij} é o peso do neurônio j em uma iteração t,x_i corresponde a um neurônio de saída ou de entrada, η representa a taxa de aprendizagem e δ_j é um erro de gradiente para o neurônio j. Caso j seja um neurônio de saída, então δ_j é definido pela Equação 3.5,

$$\delta = y_j(1 - y_j)(d_j - y_j), \tag{3.5}$$

   onde d_j denota a saída desejada e y_j é a saída obtida pela rede. Caso o neurônio j seja do tipo oculto, então δ_j é definido pela Equação 3.6

$$\delta_j = x_j(1 - x_j)\sum_k \delta_k w_{jk}, \tag{3.6}$$

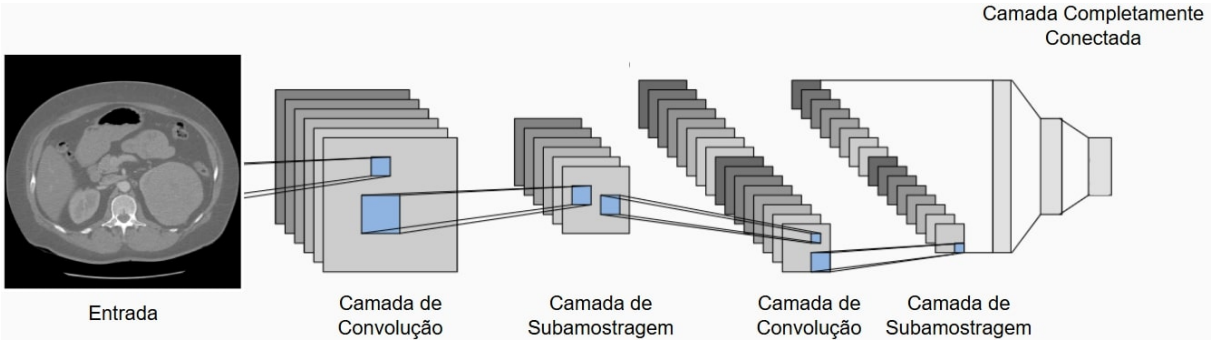
   onde k denota todos os neurônios da camada após a camada do neurônio j;
4. Repetição do processo: os passos 2 a 4 são repetidos para todos os exemplos de treinamento até que um critério de parada seja atingido, como um número máximo de iterações ou uma medida de acerto ou de erro desejada.

3.5 Redes Neurais Convolucionais

As redes neurais convolucionais, do inglês *Convolutional Neural Networks* (CNN) consistem em modelos biologicamente inspirados que buscam aprender representações e características de forma hierárquica (LECUN et al., 1998). A arquitetura da CNN baseia-se, normalmente, em três tipos de camadas, convolução, subamostragem e completamente conectada (KANG; WANG, 2014).

O processo de extração de características efetuado pelas redes CNNs normalmente ocorre nas camadas de convolução e subamostragem localizadas nas etapas intermediárias da rede. A camada completamente conectada, localizada no final da CNN, é responsável pela classificação dos padrões de entrada. A Figura 6 apresenta um fluxo geral da rede.

Figura 6 – Arquitetura padrão das Redes Neurais Convolucionais.



Fonte: Adaptado de (LECUN; KAVUKCUOGLU; FARABET, 2010).

A CNN possui uma arquitetura de rede neural *feedforward* (BEBIS; GEORGIOPOULOS, 1994) onde cada camada provê a entrada para a camada posterior. Desta forma, as características são extraídas sequencialmente de uma camada para outra da rede, onde um neurônio conecta-se a um campo receptivo local na camada anterior. A partir da combinação das características provenientes de cada uma das camadas são geradas características de mais alto nível (FERNANDES, 2013).

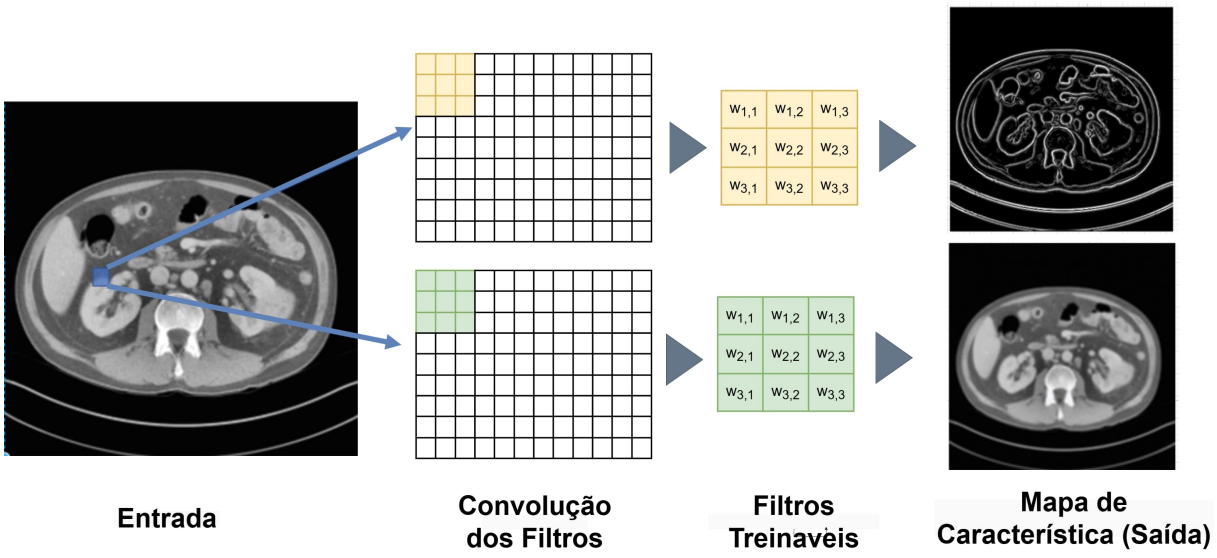
As camadas de convolução, ilustradas na Figura 7, possuem filtros treináveis aplicados individualmente à imagem de entrada, formando assim, mapas de características (LECUN; KAVUKCUOGLU; FARABET, 2010). Com o treinamento da rede, cada um dos filtros contidos na rede possuem a responsabilidade de identificar uma característica em particular que ocorre em alguma parte da imagem (HAFEMANN, 2014). Uma vez aprendida, a característica torna-se invariante, podendo ser encontrada independente de sua localização na imagem (LECUN et al., 1998).

A Equação 3.7, define o cálculo dos mapas de características, onde k representa o índice do neurônio, h_k é o mapa de característica de saída k, ω_k é o filtro treinável k, x é o mapa de característica de entrada, b_k indica o termo bias k e $*$ representa o operador de convolução espacial bidimensional.

$$(h_k)_{ij} = (\omega_k * x)_{ij} + b_k \tag{3.7}$$

As camadas de subamostragem reduzem a resolução espacial dos mapas de características ao mesmo tempo que selecionam características invariantes às distorções e deslocamento (LECUN et al., 1998). O processo de sumarização realizado por esta

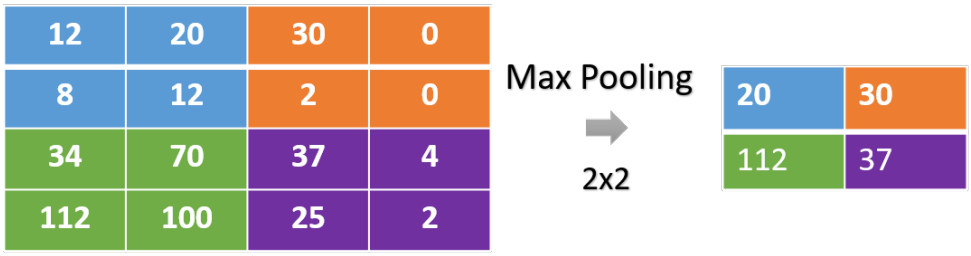
Figura 7 – Camada de Convolução.



Fonte: Adaptado de (LECUN; KAVUKCUOGLU; FARABET, 2010).

camada pode ser visualizado na Figura 8, onde é aplicada a abordagem máxima de subamostragem ou *max pooling*. Nesta abordagem, apenas o *pixel* de maior intensidade é passado para a saída. Essa sumarização de dados serve para diminuir a quantidade de pesos a serem aprendidos e também para evitar *overfitting*.

Figura 8 – Camada de subamostragem. Operação de *max pooling*.



Fonte: Desenvolvida pelo Autor

Após a finalização do processo relativos às camadas de convolução e subamostragem, os *pixels* de todos os mapas de características extraídos da última camada de subamostragem são submetidos à camada completamente conectada, onde a classificação dos padrões é efetuada (FERNANDES, 2013). A operação desta camada ocorre de maneira semelhante às redes neurais artificiais com arquitetura MLP - *Multilayer Perceptron* (ALMEIDA, 1997).

3.6 U-Net

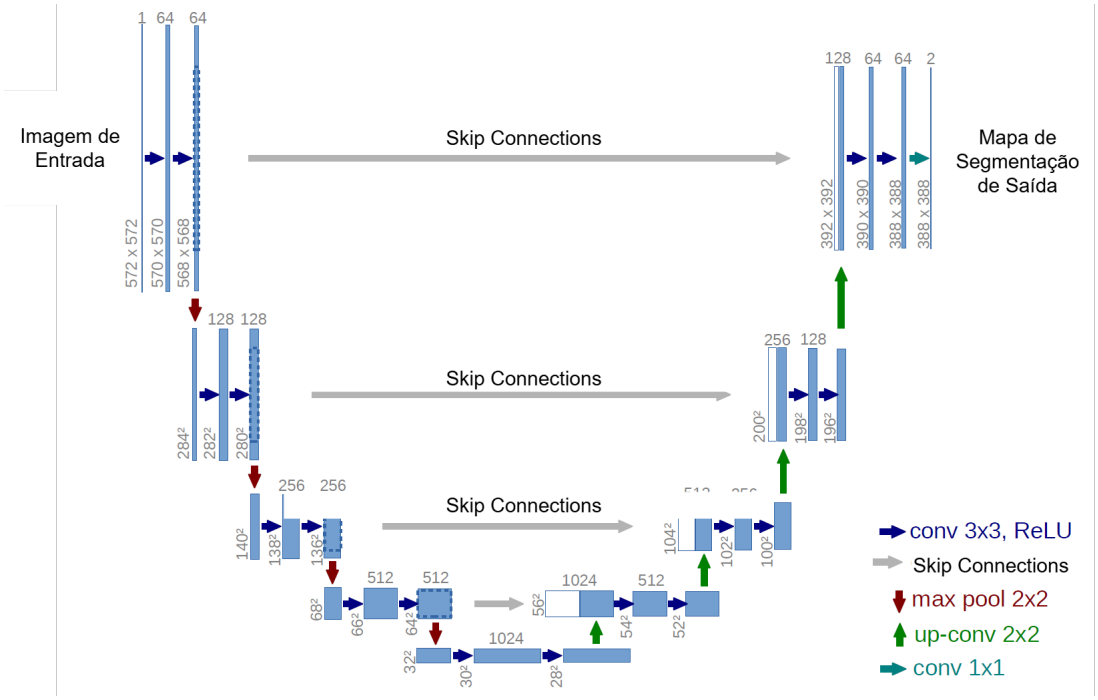
U-Net é uma rede baseada nas redes neurais convolucionais (CNN) voltada para segmentação semântica de imagens. Sua arquitetura é baseada em codificadores-

decodificadores, em que a parte de codificação extrai características e reduz a resolução da imagem de entrada, enquanto a parte do decodificação realiza a reconstrução detalhada da imagem segmentada. Essa estrutura permite que a rede aprenda características em diferentes escalas e capture informações contextuais em vários níveis de abstração (RONNEBERGER; FISCHER; BROX, 2015).

A fase de codificação é composta por uma sequência de camadas convolucionais com filtros (*kernel*) 3x3, seguidas pela aplicação de uma função de unidade linear retificada (ReLU). Ao final de cada etapa, é realizada uma operação de subamostragem máxima (*max pooling*) de tamanho 2×2 com passo 2, que reduz a resolução espacial da entrada enquanto preserva as características aprendidas. Essa redução progressiva da resolução permite que a rede capture informações em diferentes níveis de detalhe.

A fase de decodificação é composta por camadas de convolução transposta, responsáveis por aumentar a resolução espacial das características. Em seguida, essas características são combinadas com as da fase de codificação por meio das conexões de salto (skip connections). A arquitetura padrão da rede U-Net é ilustrada na Figura 9 (RONNEBERGER; FISCHER; BROX, 2015).

Figura 9 – Arquitetura padrão da U-Net.



Fonte: (RONNEBERGER; FISCHER; BROX, 2015).

3.7 Módulo de Pirâmide de *Pooling* - PPM

O Módulo de Pirâmide de *Pooling* (PPM) consiste em uma rede de análise de imagem em pirâmide para incorporar características contextuais complexas em um

framework de previsão de *pixels* baseado em redes completamente convolucionais (FCN - *Fully Convolutional Network*) (ZHAO et al., 2017).

A arquitetura base foi inicialmente proposta como componente central da rede PSPNet (ZHAO et al., 2017) e pode ser observada na Figura 10. Nesta arquitetura, inicialmente para cada imagem de entrada são obtidos mapas de características através da aplicação da rede convolucional ResNet (HE et al., 2016) com pesos pretreinados na base Imagenet (DENG et al., 2009). Em seguida, é aplicado o módulo PPM que agrega características em quatro escalas ou níveis de pirâmide diferentes. O nível mais amplo, destacado em vermelho, consiste em um agrupamento global que gera uma única saída binária. O nível seguinte, separa o mapa de características em diferentes sub-regiões e cria representações agrupadas para cada local. As saídas dos diferentes níveis no módulo de agrupamento em pirâmide têm tamanhos variados de mapa de características.

É importante destacar que uma camada de convolução $1x1$ é aplicada após cada nível da pirâmide para reduzir a dimensionalidade, seguida de uma interpolação bilinear para aumentar as características de baixa dimensão, obtendo assim um mapa de características com o mesmo tamanho do original. Por fim, as características de diferentes níveis são concatenadas para formar a característica global final do agrupamento em pirâmide.

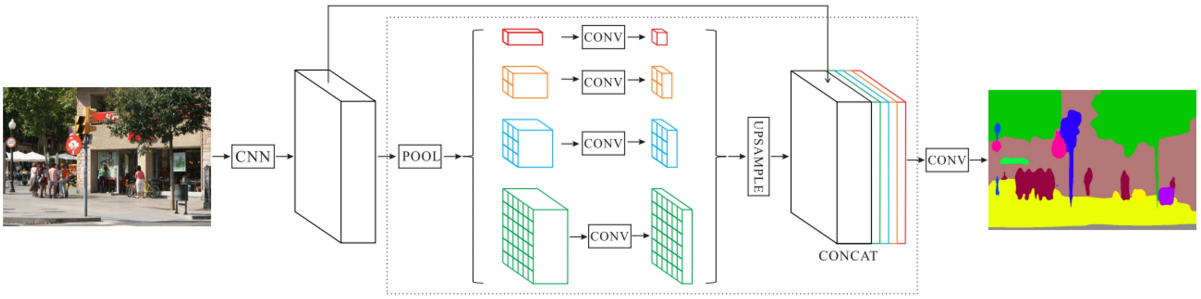
É relevante salientar que é possível ajustar o número de níveis na pirâmide e o tamanho de cada nível. Esses ajustes estão diretamente relacionados ao tamanho do mapa de características de entrada do módulo de agrupamento em pirâmide.

A configuração do módulo de agrupamento em pirâmide permite a captura de diferentes sub-regiões por meio do uso de filtros convolucionais (*kernels*) com tamanhos variados e um número limitado de passos. Conforme mencionado por Zhao et al. (2017), é crucial manter uma diferença adequada na representação entre os filtros em estágios múltiplos. O módulo de agrupamento em pirâmide definido por Zhao et al. (2017) consiste em quatro níveis, com tamanhos de $1x1$, $2x2$, $3x3$ e $6x6$, respectivamente. A arquitetura PPM oferece flexibilidade quanto ao tipo de operação de agrupamento, permitindo a utilização do agrupamento máximo (*maxpooling*) ou médio (*average pooling*).

3.8 Pirâmide de *Pooling* Espacial com Dilatação - ASPP

A convolução Atrous, também chamada de convolução dilatada, é uma técnica utilizada em redes neurais convolucionais profundas (DCNN) para controlar explicitamente a resolução das características extraídas durante o processamento de imagens. Essa técnica funciona ajustando o campo de visão do filtro convolucional, o que permite capturar informações em múltiplas escalas e resoluções, essenciais para tarefas que exigem uma compreensão detalhada de diferentes níveis de granularidade nos dados.

Figura 10 – Arquitetura padrão do PPM.



(a) Imagem de Entrada (b) Mapa de Característica (c) Módulo de Pirâmide de *Pooling* (d) Predição Final

Fonte: Adaptado de Zhao et al. (2017)

O princípio fundamental da convolução dilatada, é a aplicação de uma taxa de dilatação (r), que define o espaçamento entre os *pixels* aplicado no filtro convolucional. Esse mecanismo permite expandir o campo de visão da convolução sem aumentar o tamanho do filtro ou adicionar mais parâmetros ao modelo. Como resultado, a convolução atrous é capaz de capturar informações de contexto mais amplas, mantendo o mesmo número de parâmetros e a mesma complexidade computacional que uma convolução tradicional.

Considerando sinais bidimensionais, para cada localização i no mapa de características de saída y e um filtro de convolução w, a convolução atrous é aplicada ao mapa de características de entrada x da seguinte forma:

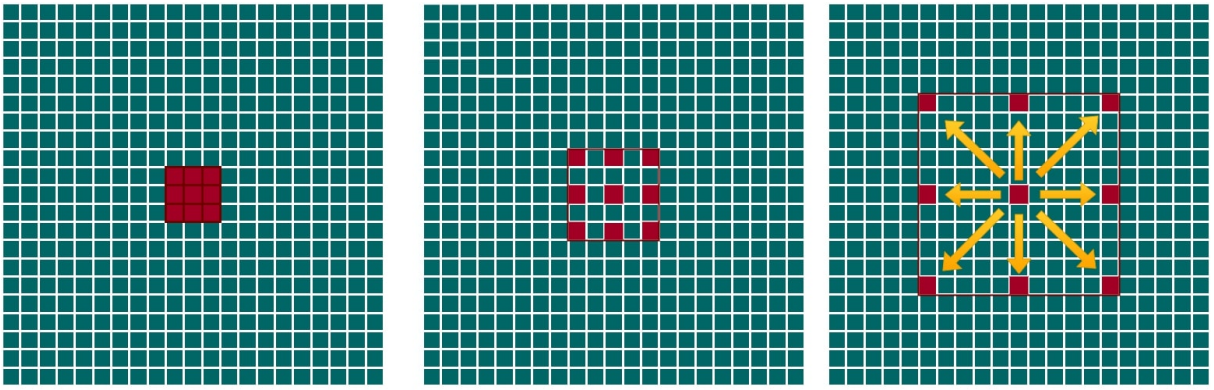
$$y\left[i\right] = \sum_{k} x\left[i + r \cdot k\right] w\left[k\right] \tag{3.8}$$

onde a taxa de dilatação r determina o espaçamento do filtro que será aplicado ao sinal de entrada.

O comportamento das convoluções dilatadas é diretamente influenciado pela taxa de dilatação (r) aplicada, como discutido anteriormente. A Figura 11 ilustra o impacto dessa taxa sobre o filtro de convolução em uma imagem. Quando a taxa de dilatação é igual a 1, obtém-se o filtro de convolução tradicional, como mostrado na Figura 11(a). À medida que a taxa de dilatação aumenta, conforme ilustrado nas Figuras 11(b) e (c), observa-se uma expansão no campo de visão do filtro.

O módulo de Pirâmide de *Pooling* Espacial com Dilatação (*Atrous Spatial Pyramid Pooling* - ASPP) baseia-se na aplicação de convoluções dilatadas estruturadas em formato de pirâmide seguida por uma camada de *pooling*. Esta estruturação objetiva incrementar a capacidade de captura de informações em múltiplas resoluções das redes convolucionais tradicionais. Esta técnica combina convoluções dilatadas com diferentes

Figura 11 – Exemplo de convolução dilatada: (a-c) Taxas de dilatação iguais a 1, 2 e 5.



(a) *Kernel* de tamanho 3x3 taxa de dilatação 1

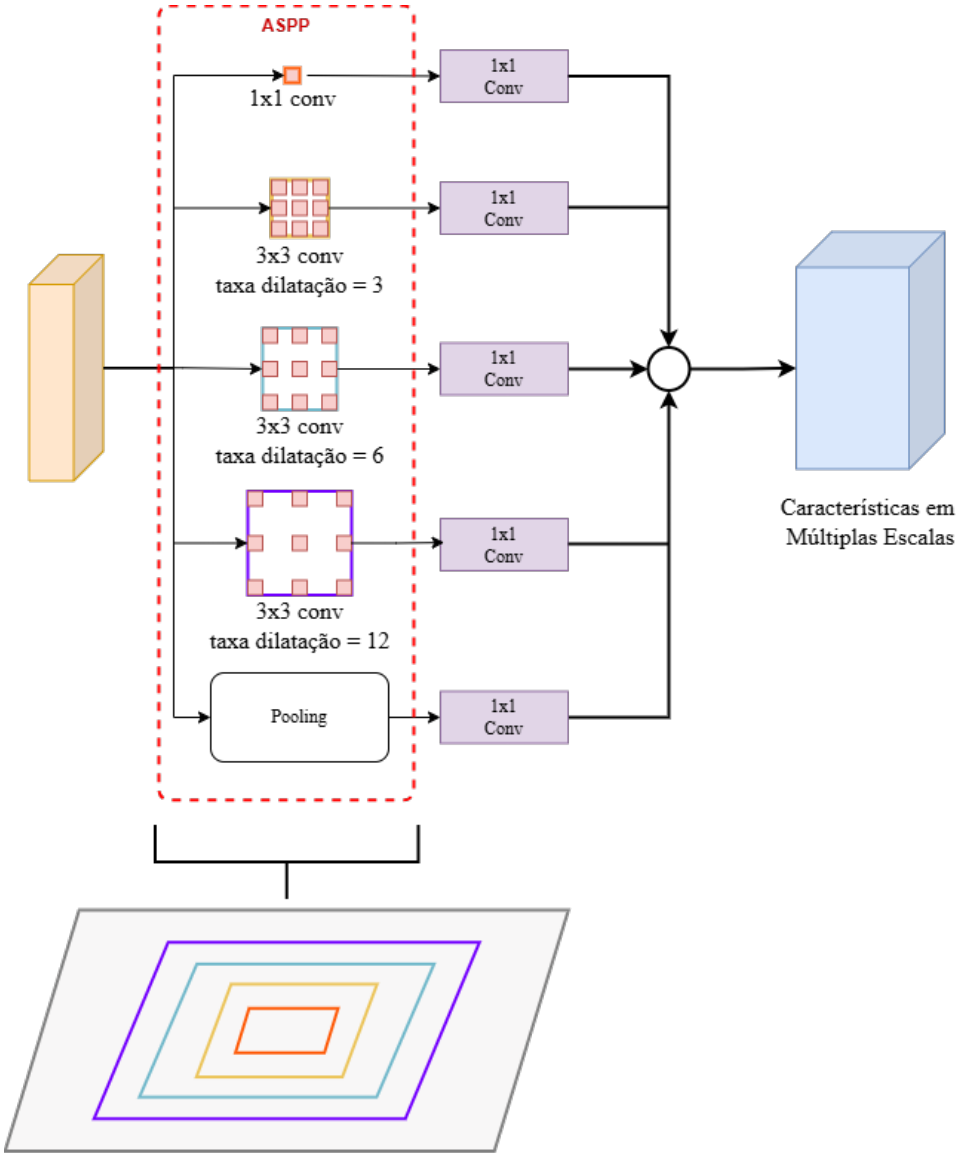
(b) *Kernel* de tamanho 3x3 taxa de dilatação 2

(c) *Kernel* de tamanho 3x3 taxa de dilatação 5

Fonte: Adaptado de Chen et al. (2017).

taxas de dilatação para extrair informações de características em múltiplos níveis de granularidade, sem perder detalhes de resolução espacial ou aumentar significativamente a complexidade computacional (CHEN et al., 2017).

Figura 12 – A estrutura do módulo de pirâmide de pooling espacial dilatado. O módulo é composto por duas etapas: (a) convolução atrous e (b) pooling da imagem, e gera a saída final por meio de uma camada de convolução após a concatenação dos mapas de características.



Fonte: Adaptado de Ding et al. (2020).

A Figura 12 apresenta a estrutura arquitetural do módulo ASPP, no qual o mapa de características de entrada é processado por uma pirâmide de *pooling*. Nesse processo, convoluções dilatadas são aplicadas em diferentes níveis, utilizando distintas combinações de taxas de dilatação e tamanhos dos filtros, permitindo a captação de informações em múltiplas escalas. O primeiro nível da pirâmide utiliza uma convolução com filtro de tamanho $1x1$ e taxa de dilatação igual a 1. Nos níveis subsequentes, são aplicadas convoluções $3x3$ e taxas de dilatação de $13, 6$ e 12, respectivamente, do segundo ao quarto nível.

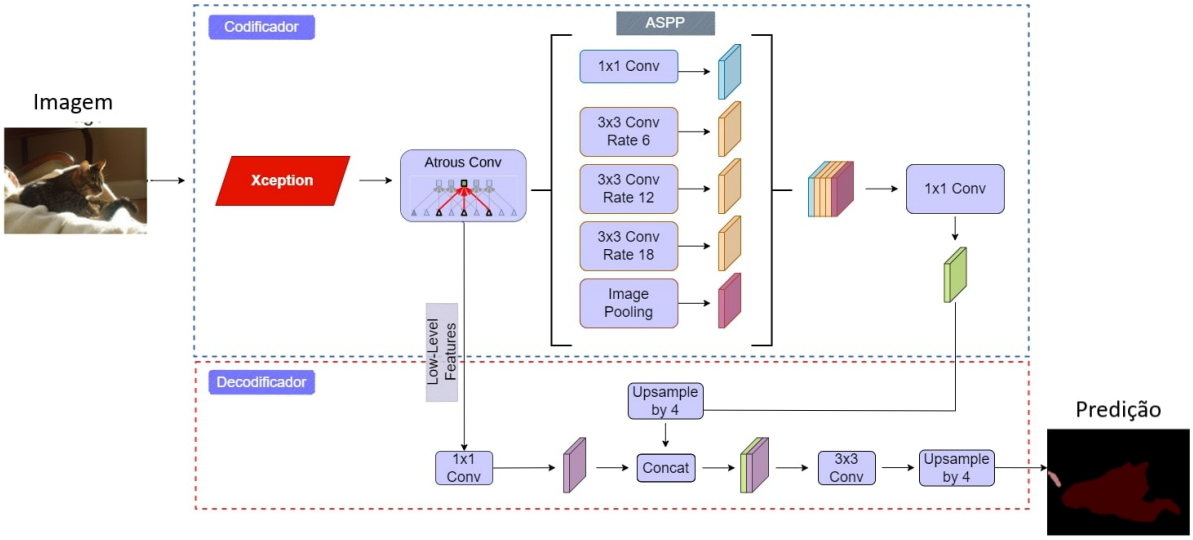
Ao final do módulo de pirâmide, a operação de *pooling* máximo é aplicada

no último nível. Por fim, todas as saídas são concatenadas para gerar o mapa de características final da arquitetura.

3.9 DeepLabv3+

Para a segmentação dos rins, cistos e tumores renais em imagens de TC, foi adotada a arquitetura DeepLabv3+ (CHEN et al., 2018) como base para a construção de um dos modelos de segmentação desenvolvidos na metodologia desta tese. Essa arquitetura é reconhecida por sua capacidade de codificar informações contextuais e aprimorar os resultados da segmentação, especialmente nas regiões de contorno dos objetos. Conforme mencionado anteriormente, as redes neurais profundas adotam a estrutura de codificador-decodificador para realizar tarefas de segmentação semântica. O codificador é responsável por extrair informações contextuais em diferentes escalas, enquanto o decodificador é encarregado de capturar os contornos dos objetos de maneira mais precisa, recuperando gradualmente as informações espaciais para reconstruir a saída (CHEN et al., 2018). Uma visão da arquitetura DeepLabv3+ pode ser observada na Figura 13.

Figura 13 – Arquitetura padrão da DeepLabV3+.



Fonte: Adaptado de Chen et al. (2018).

A fase de codificação é composta por três etapas principais: (1) extração de características, (2) módulo ASPP e (3) saída do codificador. Na primeira etapa, é utilizada uma versão modificada da rede convolucional *Xception* (*Extreme Inception*) (CHOLLET, 2017). Essa versão incorpora convoluções separáveis em profundidade (*depthwise separable convolution*), normalização em lote (*batch normalization*) e ativação ReLU.

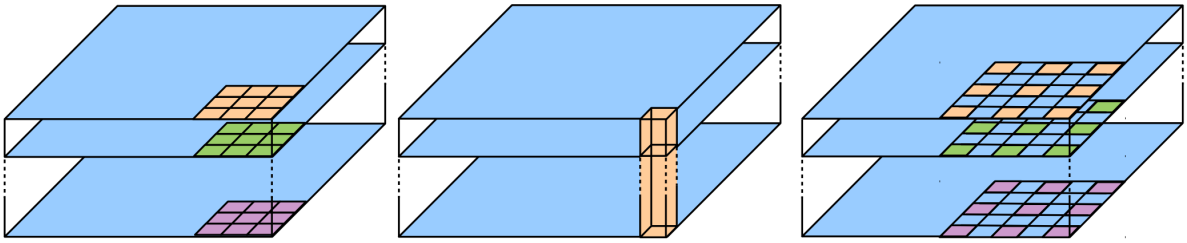
Diferentemente da convolução convencional, que realiza simultaneamente os cálculos espaciais e por canal, a convolução separável em profundidade divide esse

processo em duas fases. Inicialmente, um filtro convolucional específico é aplicado a cada canal de entrada. Em seguida, a convolução ponto a ponto combina linearmente as saídas geradas.

A rede Deeplabv3+ aprimora essa abordagem ao empregar uma variação denominada convolução separável em profundidade com dilatação (*Atrous Depthwise Convolution*), permitindo capturar informações em diferentes escalas de forma mais eficiente. A Figura 14 apresenta uma comparação entre os diferentes tipos de convolução: convencional, separável em profundidade com e sem dilatação.

Na segunda etapa do codificador, o módulo ASPP é aplicado conforme as definições descritas na Seção 3.8, utilizando taxas de dilatação iguais a 6, 12 e 16. Por fim, na terceira etapa, a saída do codificador é obtida por meio da concatenação dos níveis provenientes do módulo ASPP e de uma operação de convolução $1x1$.

Figura 14 – Convolução padrão e convolução separável em profundidade.



(a) Convolução Separável com Profundidade

(b) Convolução Pontual

(c) Convolução Separável com Profundidade e Dilatação

Fonte: Chen et al. (2018).

A fase de decodificação é responsável pela reconstrução das informações extraídas na codificação para gerar a segmentação final. Portanto, operações de aumento, concatenações e convoluções são aplicadas nesta fase. As características de saída do codificador são inicialmente ampliadas por interpolação bilinear em um fator de 4. Em seguida, as características de baixo nível geradas pela rede Xception são submetidas a uma convolução $1x1$ para reduzir o número de canais, uma vez que essas características geralmente possuem muitos canais. Posteriormente, as saídas correspondentes às operações de convolução e ampliação são concatenadas e, em seguida, convoluções $3x3$ são aplicadas para refinar as características. Por fim, nova operação de ampliação bilinear de fator 4 é aplicada para então gerar a predição final contendo a segmentação proposta pela rede Chen et al. (2018).

3.10 Avaliação dos Modelos

Em problemas voltados à área da saúde, a composição básica dos testes de diagnósticos define a capacidade do modelo de discriminar as estruturas presentes em exames de imagem para a identificação de possíveis patologias.

Desta forma, para validar os resultados da metodologia desta tese, foram utilizadas métricas de avaliação amplamente empregadas na segmentação de imagens médicas. Nessas aplicações, o cálculo dessas métricas baseia-se na matriz de confusão, ilustrada na Figura 15.

Figura 15 – Exemplo de matriz de confusão.

		Valor Predito	
		Sim	**Não**
Real	**Sim**	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: Adaptado de (TING; VIJAYAKUMAR; SCHAAL, 2010).

Esta matriz organiza as previsões do modelo em relação às classes reais dos dados. Em uma classificação binária, a matriz de confusão é composta por quatro variáveis principais: Verdadeiro Positivo (VP) que indica os casos em que o modelo previu corretamente a classe positiva, ou seja, correta segmentação do objeto de estudo (rins, cistos e tumores renais); Falso Positivo (FP) registra a ocorrência de classificações equivocadas, onde os *pixels* da classe negativa são erroneamente classificados como pertencentes à classe positiva; Verdadeiro Negativo (VN) indica a classificação correta dos *pixels* da classe negativa; e Falso Negativo (FN) denota a classificação incorreta dos *pixels* da classe positiva como negativa.

Nesta pesquisa, são empregadas métricas comumente usadas em abordagens de segmentação semântica como o coeficiente de similaridade Dice e o índice de Jaccard (XU, 2017; TAHA; HANBURY, 2015; EELBODE et al., 2020).

O coeficiente Dice, também conhecido como *F-Score*, é considerado a métrica de avaliação mais utilizada para a segmentação de imagens médicas. Esta métrica pode ser calculada através da razão entre o dobro da interseção entre as duas regiões segmentadas e a soma das áreas das duas regiões. Matematicamente, pode ser

expressa pela Equação 3.9,

$$DSC = \frac{2VP}{2VP + FP + FN}. \tag{3.9}$$

O índice de Jaccard, também conhecido como coeficiente de Jaccard ou *Intersection over Union* (IoU), assim como o coeficiente Dice, é uma métrica aplicada para verificar a similaridade entre dois conjuntos. Pode ser calculado pela razão entre o tamanho da interseção dos conjuntos e o tamanho da união dos conjuntos. Pode ser denotada pela Equação 3.10,

$$JCC = \frac{VP}{VP + FP + FN}. \tag{3.10}$$

As métricas mencionadas são essenciais para avaliar o desempenho da metodologia, permitindo uma análise comparativa com pesquisas correlatas. Dessa forma, é possível posicionar o método proposto quanto ao seu desempenho, identificando se é satisfatório e destacando pontos positivos e limitações. Essas informações são valiosas não apenas para validar a abordagem, mas também para orientar melhorias e fornecer subsídios para pesquisas futuras.

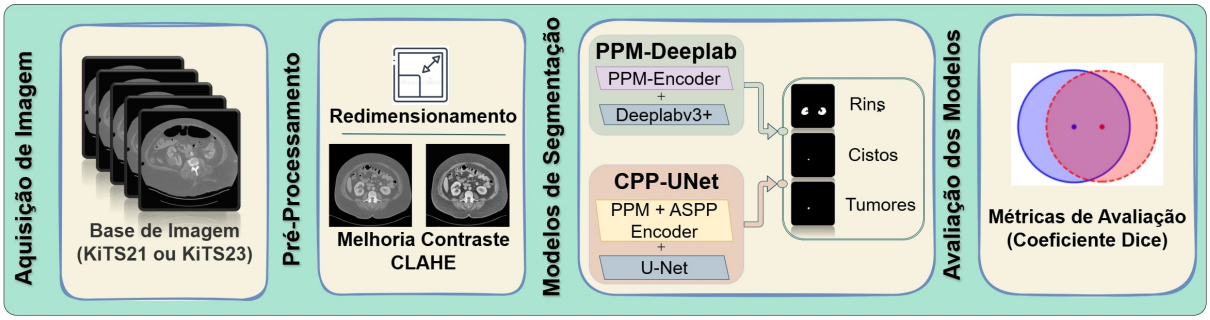
3.11 Considerações Finais

Este capítulo apresentou a fundamentação teórica usada no desenvolvimento deste trabalho. Desta forma, os conceitos e teorias apresentados são de suma importância para a compressão da metodologia desenvolvida. Para tanto, foram apresentados conceitos sobre a fisiologia e o câncer renal, redes neurais artificiais, redes completamente convolucionais (FCN), bem como arquitetura baseadas em módulos de pirâmide de *pooling*. No próximo Capítulo (4 - Metodologia), as etapas da metodologia desenvolvida nesta tese são apresentadas e discutidas.

4 Metodologia

Este capítulo descreve os procedimentos adotados para o desenvolvimento do método proposto para segmentação dos rins, cistos e tumores renais em imagens de tomografia computadorizada (TC). As quatro etapas desenvolvidas neste trabalho são descritas na Figura 16.

Figura 16 – Fluxograma da metodologia.



Fonte: Desenvolvida pelo Autor

Inicialmente, na etapa de aquisição de imagem, foram obtidas duas versões, KiTS21 e KiTS23, da base de imagem pública disponibilizada pelo desafio de segmentação (HELLER et al., 2021; HELLER et al., 2023b), como descrito na Seção 4.1. Posteriormente, foi realizada a etapa de pré-processamento em todos os volumes de tomografia computadorizada (TC), na qual ocorreu o redimensionamento e aprimoramento das imagens. Na etapa seguinte, são aplicados os modelos para a segmentação de câncer renal propostos na metodologia desta tese. Para tanto, foram desenvolvidos dois novos modelos arquiteturais, denominados PPM-Deeplab e CPP-UNet, ambos baseados na integração de blocos de Pirâmide de *Pooling* (PPM e ASPP) como componentes do codificador das redes convolucionais Deeplabv3+ e U-Net. Por fim, os resultados obtidos são avaliados na etapa final da metodologia.

4.1 Bases de Imagens

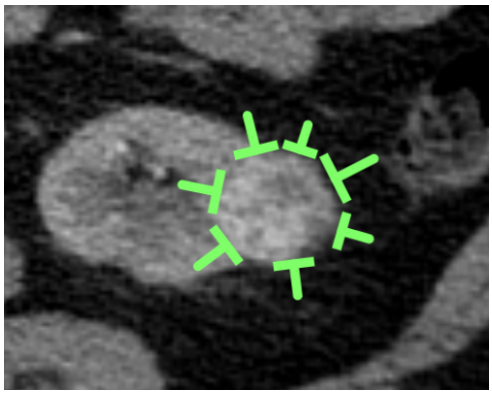
Neste trabalho, foram utilizadas as duas versões mais recentes da base de imagens disponibilizada pelo desafio de segmentação KiTS (*Kidney Tumor Segmentation Challenge*) (HELLER et al., 2021; HELLER et al., 2023a), lançadas nos anos de 2021 e 2023. O objetivo desse desafio é fomentar avanços na pesquisa relacionada à segmentação de tumores renais, contribuindo para o aprimoramento do diagnóstico e tratamento de pacientes com câncer renal. O conjunto de dados fornecido inclui

informações de pacientes submetidos a crioablação[1], nefrectomia parcial ou nefrectomia radical devido à suspeita de malignidade renal.

Esses dados foram coletados entre 2010 e 2022 no centro médico *M Health Fairview* (M Health Fairview, 2024), estabelecido por meio de uma parceria entre o centro médico da Universidade de Minnesota (EUA) e a empresa de assistência médica *Fairview Health Services* (Fairview Health Services, 2024). Todas as versões do desafio de segmentação KiTS foram desenvolvidas a partir dessa colaboração. Em termos gerais, a competição oferece um amplo conjunto de dados de pacientes nos quais pelo menos um tumor renal foi identificado.

Além disso, ambos os conjuntos de dados, KiTS21 e KiTS23, incluem anotações manuais das regiões correspondentes aos rins e das patologias identificadas nas imagens de tomografia computadorizada. O processo de anotação é realizado em duas etapas. Na primeira, estudantes de medicina marcaram as patologias inserindo pinos de orientação em cada imagem do exame, conforme ilustrado na Figura 17.

Figura 17 – Etapa inicial da anotação: inserção de alfinetes de orientação pelos estudantes de medicina. Exemplo ilustrativo de anotação de tumor renal.



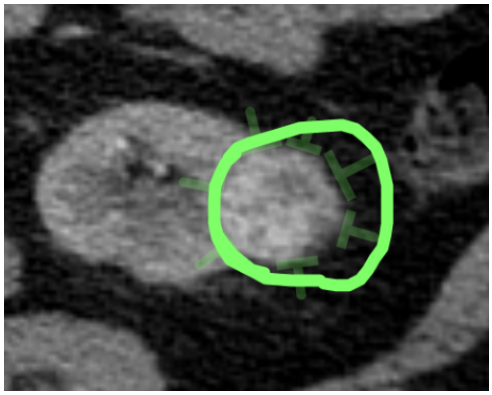
Fonte: Adaptado de (HELLER et al., 2021)

As anotações foram revisadas por um grupo de médicos especialistas e, em seguida, retornadas aos estudantes para correção. Após o término das revisões, o processo é concluído, gerando a segmentação final, como apresentado na Figura 18. É importante destacar que esses estudantes passaram por várias horas de treinamento conduzidas por médicos especialistas.

O desafio KiTS define quatro macroclasses denominadas de Classes de Avaliação Hierárquica (CAH) para avaliação e comparação dos resultados apresentados pelos participantes. Segundo o desafio, estas classes são baseadas na métrica de coeficiente Dice relacionada ao resultados obtidos dos rins, cistos e tumores renais. Portanto, estas

[1] Procedimento minimamente invasivo que congela e destrói as células cancerígenas. Pode ser utilizado no tratamento do câncer renal em estágio inicial (MORKOS et al., 2020).

Figura 18 – Segmentação final: contorno desenhado com base nos alfinetes de orientação inseridos pelos estudantes de medicina. Exemplo ilustrativo de segmentação de tumor renal.



Fonte: Adaptado de (HELLER et al., 2021)

classes são combinadas da seguinte forma: Rins e Massas (Rim + Tumor + Cisto), Massas Renais (Tumor + Cisto) e Tumores renais. A Tabela 2 ilustra esta composição.

Tabela 2 – Classes de Avaliação Hierárquica referente ao desafio KiTS.

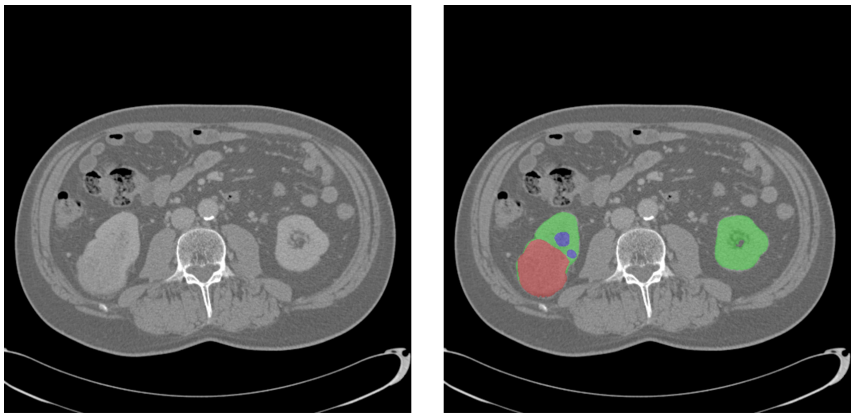
	Rins	**Tumores**	**Cistos**
Rins e Massas	X	X	X
Massas Renais		X	X
Tumores		X	

4.1.1 Base KiTS21

A segunda edição do desafio, KiTS21 (HELLER et al., 2021) teve como objetivo promover o desenvolvimento de sistemas de segmentação automática de rins, tumores e cistos renais. A base de dados é composta por um total de 300 pacientes que passaram por nefrectomia devido a tumores renais no período de 2010 a 2020.

As anotações dos especialistas e imagens obtidas nos exames de tomografia computadorizada foram fornecidas pelo desafio no formato NIFTI (GROUP et al., 2023) para cada volume. Em seguida, os volumes foram convertidos em imagens (fatias) em escala de cinza com resolução de $512x512$ *pixels*. Devido às variações nos protocolos e na duração dos exames, a quantidade de imagens por paciente pode variar significativamente. A Figura 19 exemplifica uma fatia e sua respectiva marcação disponibilizada pela base de dados KiTS21. Nessa ilustração, as cores verde, vermelho e azul correspondem às marcações dos rins, tumores e cistos renais, respectivamente.

Figura 19 – Exemplo da base KiTS21. (a) fatia do exame e (b) marcação dos rins, tumor e cistos destacadas nas cores verde, vermelho e azul, respectivamente.



a) b)

Fonte: (HELLER et al., 2021)

4.1.2 Base KiTS23

Na sua versão mais recente, KiTS23 (HELLER et al., 2023a), o desafio introduziu como novidade a inclusão de casos na fase de contraste nefrogênico[2], enquanto as versões anteriores apresentavam apenas pacientes na fase de arterial tardia[3]. Para esta edição, foram disponibilizados 599 exames de tomografia computadorizada, dos quais 489 foram destinados ao treinamento e validação dos modelos, enquanto 110 exames foram reservados para teste e ranqueamento das soluções submetidas pelos participantes.

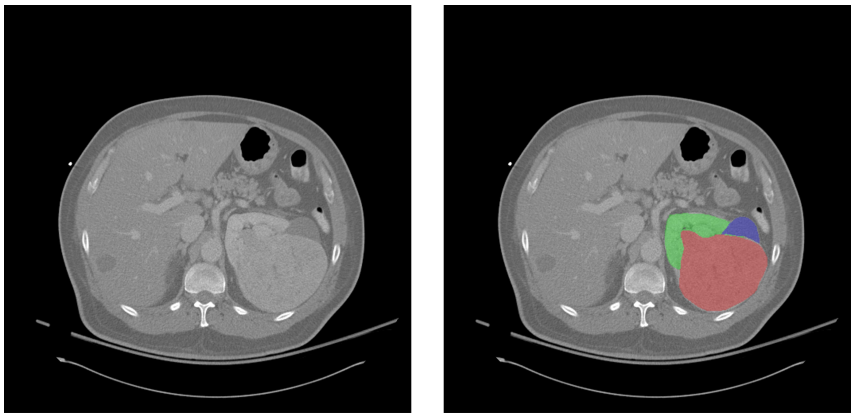
Assim como na versão anterior, KiTS21, o formato NIFTI (GROUP et al., 2023) foi utilizado pela competição para representação dos volumes referentes às imagens de tomografia computadorizada (TC) e suas respectivas anotações (segmentações fornecidas). As imagens que compõem o exame foram convertidas em representações em escala de cinza com resolução de $512x512$ *pixels*.

Por representar uma extensão da base anterior, variações nos protocolos de aquisição e na duração dos exames podem influenciar na quantidade de imagens produzidas por paciente no exame de TC. A Figura 20 ilustra um exemplo de fatia obtida em um exame de TC e sua respectiva marcação manual. A Figura 20(a) apresenta a imagem em escala de cinza da região abdominal, enquanto a Figura 20(b) destaca, nas cores verde, vermelho e azul, a segmentação manual feita por especialistas para identificar rins, tumores e cistos renais.

[2] Intervalo de tempo pós-injeção de contraste no qual há um realce ótimo do parênquima renal

[3] Tomografia computadorizada com contraste, na qual há um realce ideal das estruturas que recebem seu suprimento sanguíneo diretamente do sistema arterial (KULKARNI et al., 2021)

Figura 20 – Exemplo da base KiTS23. (a) fatia do exame de TC e (b) marcação do rim, tumor e cisto destacadas nas cores verde, vermelho e azul, respectivamente.



a) b)

Fonte: (HELLER et al., 2023b)

4.2 Pré-Processamento

O objetivo desta etapa consiste na redução da dimensionalidade das fatias e no aumento do contraste dos objetos de interesse (rins, cistos e tumores) em relação aos demais órgãos presentes no exame. Com isso, é possível fornecer melhores características às demais etapas da metodologia.

Devido a limitações computacionais encontradas no treinamento dos modelos propostos, foi necessário efetuar o redimensionamento das fatias (imagens) de todos os exames fornecidos da base. As imagens foram reduzidas de $512x512$ *pixels* para $256x256$ *pixels*, tornando assim possível a aplicação do modelo arquitetural proposto neste trabalho.

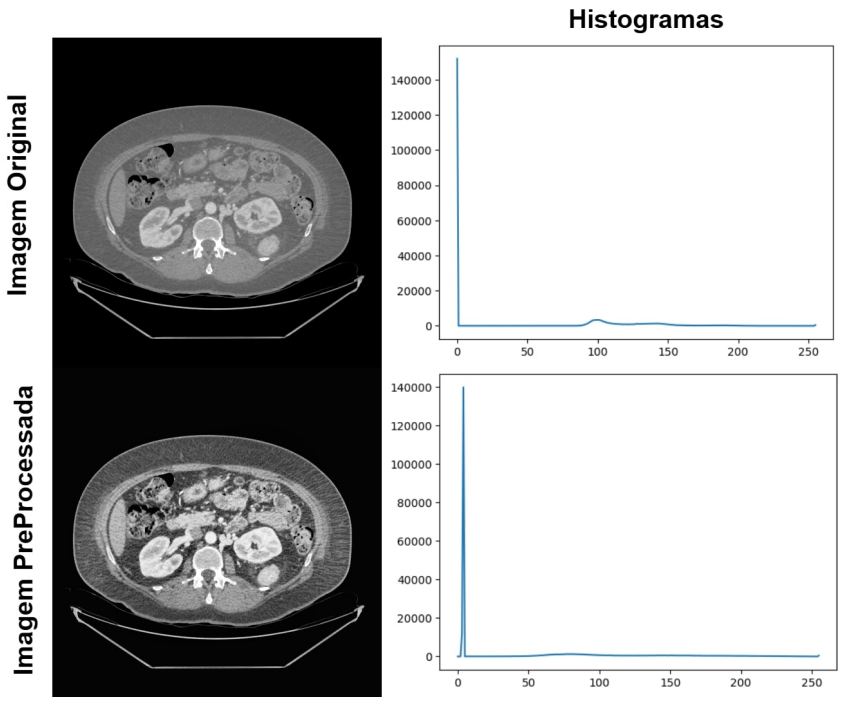
Em seguida, cada fatia passou por realce de contraste usando CLAHE (PIZER et al., 1987). A técnica de preprocessamento Equalização do Histograma Adaptativo Limitada ao Contraste (do inglês, Contrast Limited Adaptive Histogram Equalization - CLAHE) consiste em uma variante da equalização de histograma adaptável (*Adaptive histogram equalization* - AHE) que regula a amplificação excessiva do contraste. Esta técnica opera em pequenas regiões da imagem, onde cada *pixel* da imagem original está no centro da região contextual. Como resultado deste processo, o histograma gerado pela nova imagem apresenta melhores contrastes (PIZER et al., 1987).

Nesta metodologia, o CLAHE foi aplicado visando melhorar o contraste das imagens, preservando os detalhes locais e reduzindo a amplificação de ruídos. O limite de contraste foi configurado para 3.0, o que controla a clipagem do histograma e evita a superexposição em áreas homogêneas da imagem. O tamanho da grade para equalização do histograma foi definido como $4x4$, segmentando a imagem em pequenas

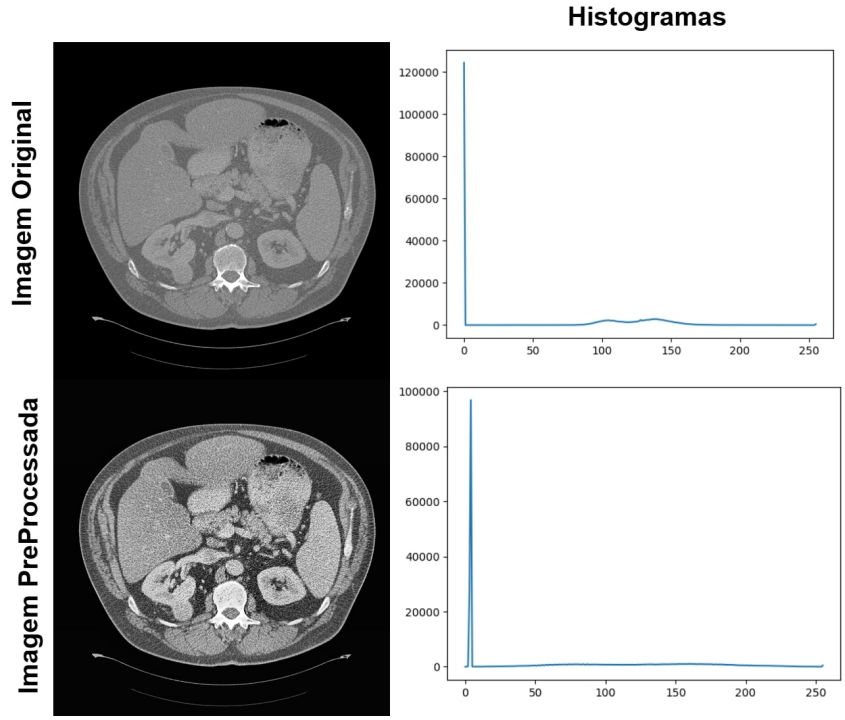
regiões chamadas de *tiles*, permitindo o ajuste local do contraste eficientemente. Esses parâmetros foram escolhidos para equilibrar o aumento de contraste e a preservação de detalhes em imagens médicas, onde a interpretação precisa é crucial. A Figura 21 ilustra a aplicação da técnica nas bases KiTS21 e KiTS23, demonstrando a eficácia do CLAHE em realçar características relevantes para a segmentação de estruturas renais.

Figura 21 – Exemplo da aplicação do CLAHE com os respectivos histogramas.

[Fatia da base KiTS21 e imagem preprocessada.]



[Fatia da base KiTS23 e imagem preprocessada.]



Fonte: Elaborado pelo Autor.

4.3 Modelos de Segmentação

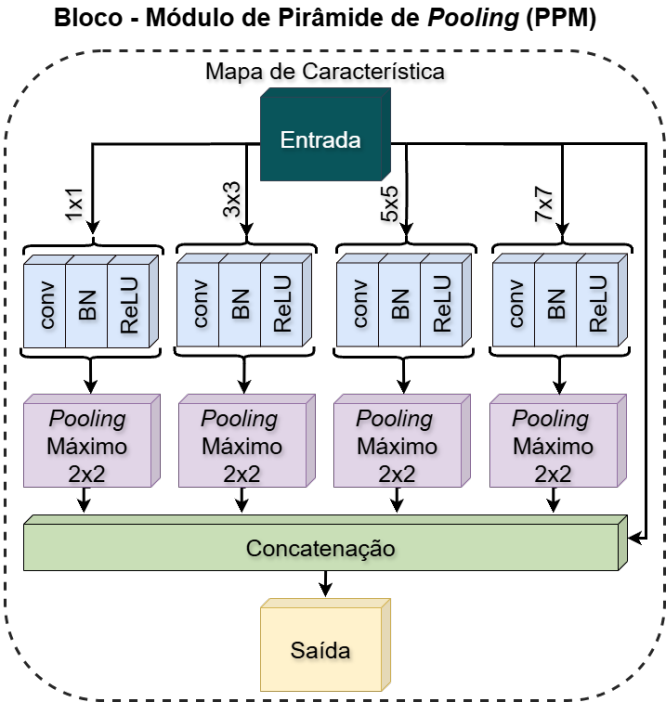
Nesta seção, apresentamos o desenvolvimento de arquiteturas de redes convolucionais projetadas para a segmentação de rins, cistos e tumores renais em

imagens de tomografia computadorizada (TC), com base no uso de módulos de pirâmides de *pooling*. Foram desenvolvidos dois modelos arquiteturais: o PPM-Deeplabv3+ e o CPP-UNet. O primeiro integra o módulo de pirâmide PPM (ZHAO et al., 2017) com a arquitetura DeepLabv3+ (CHEN et al., 2018). Por outro lado, o segundo modelo, denominado de combinação de pirâmides de *pooling* ou CPP-UNet, baseia-se na combinação dos módulos PPM e ASPP (Piramide de *Pooling* Espacial com Dilatação) incorporados à rede U-Net. Nas próximas subseções, a estrutura arquitetural e o funcionamento desses modelos são descritos em detalhes.

4.3.1 Bloco PPM

Neste trabalho, utilizou-se a arquitetura do Módulo de Piramide de *Pooling* (PPM), conforme ilustrado na Figura 22. No início do processo, o mapa de características de entrada é submetido aos quatro níveis da pirâmide de *pooling*. Para cada um dos níveis foram aplicados os filtros de tamanhos $1x1$, $3x3$, $5x5$ e $7x7$ representando os níveis da pirâmide do primeiro ao último. Para cada uma dessas resoluções, são aplicadas operações de convolução seguidas por *Batch Normalization* e ativação ReLU, para refinar as características extraídas em cada escala.

Figura 22 – Arquitetura do bloco PPM



Fonte: Adaptado de Zhao et al. (2017)

Em seguida, aplica-se a operação de *pooling* máximo, com um *filtro* de tamanho $2x2$, uniformemente em todos os níveis da pirâmide, para reduzir a dimensão espacial enquanto preserva informações contextuais relevantes. Ao final deste processo, as

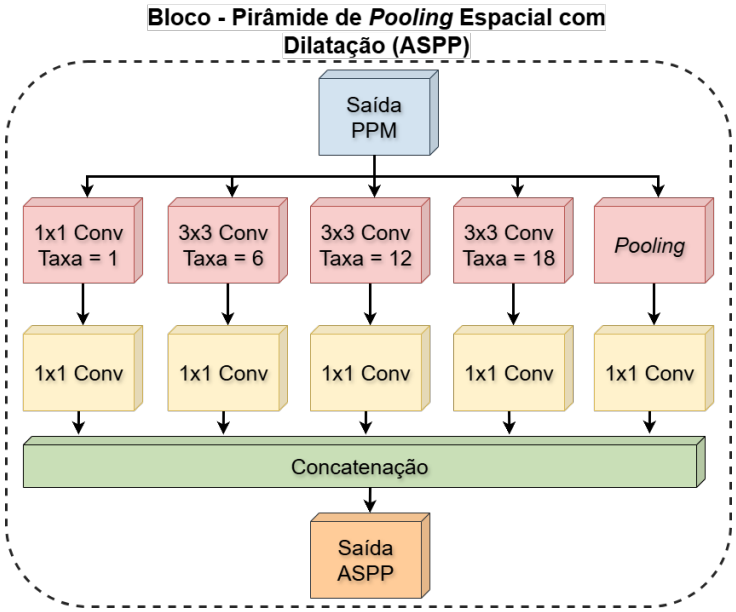
saídas das diferentes resoluções são concatenadas, sendo integradas ao mapa de características originais do bloco PPM.

É importante ressaltar que o tamanho dos filtros convolucionais aplicados e a quantidade de níveis da pirâmide foram definidos de forma empírica, com base em testes que resultaram nos melhores desempenhos para a segmentação de rins, cistos e tumores renais.

4.3.2 Bloco ASPP

O módulo de pirâmide de pooling espacial com dilatação foi aplicado neste trabalho para compor a etapa de extração de características. Para isso, o módulo foi integrado ao codificador dos modelos de segmentação desenvolvidos nesta tese (PPM-DeepLabv3+ e CPP-UNet). Em linhas gerais, a arquitetura ASPP utilizada é composta por um mapa de características de entrada seguido pela pirâmide de pooling com convoluções dilatadas. Na etapa seguinte, aplica-se um bloco de convolução de tamanho $1x1$ ao final de cada nível da pirâmide. Por fim, todas as saídas são concatenadas para gerar as características de saída do bloco ASPP. A composição desta arquitetura é apresentada na Figura 23.

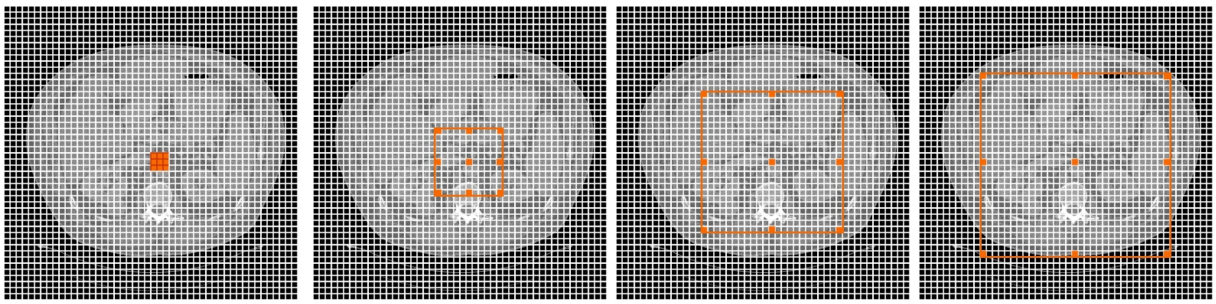
Figura 23 – Arquitetura do bloco ASPP



Fonte: Adaptado de Chen et al. (2017)

Nesta arquitetura, os filtros aplicados foram de tamanho $1x1$ e $3x3$. As taxas de dilatação empregadas foram $r = 1, 6, 12$ e 18. A Figura 24 ilustra o impacto dessas taxas sobre o filtro de convolução em uma fatia de tomografia computadorizada (TC).

Figura 24 – Exemplo de convolução dilatada: (a-d) Imagens de amostra de TC com taxas de dilatação = 1, 6, 12 e 18



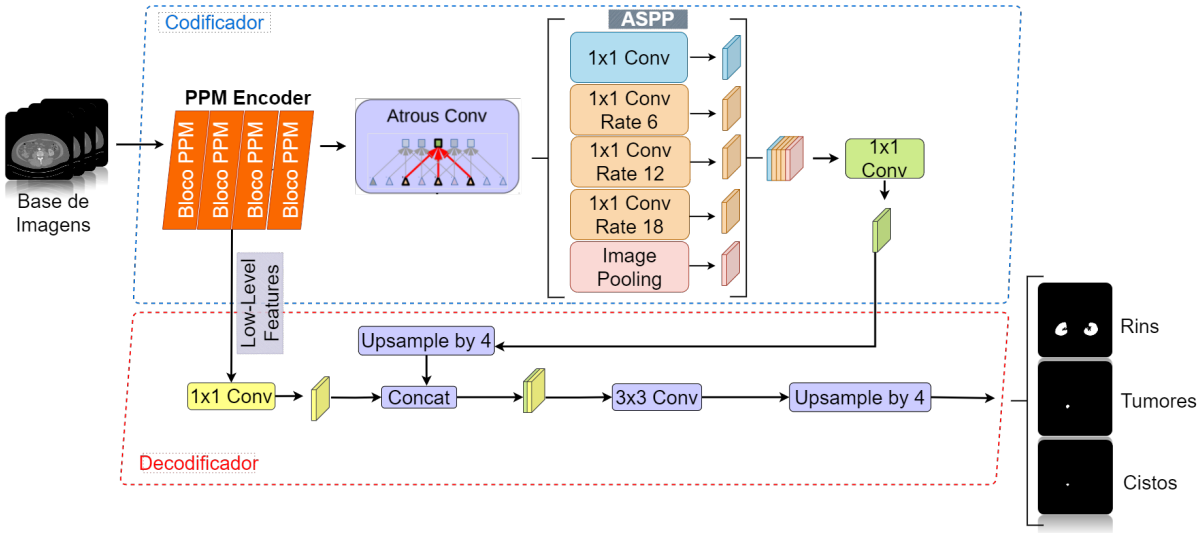
(a) *Kernel* de tamanho 3x3 taxa de dilatação 1 | (b) *Kernel* de tamanho 3x3 taxa de dilatação 6 | (c) *Kernel* de tamanho 3x3 taxa de dilatação 12 | (d) *Kernel* de tamanho 3x3 taxa de dilatação 18

Fonte: Elaborado pelo Autor

4.3.3 Arquitetura Proposta - PPM-DeepLabv3+

Neste trabalho, foi proposto um novo modelo arquitetural denominado PPM-DeepLabV3+. Este modelo consiste em uma adaptação da arquitetura padrão DeepLabv3+ (Seção 3.9), onde se efetuou substituição da rede extratora de característica, normalmente uma rede Xception, presente no codificador padrão da DeepLabv3+, pela arquitetura de Pirâmides de *Pooling* (PPM). Para tanto, foi adicionado um total de quatro blocos de PPM à composição do PPM *Encoder* conforme ilustrado na Figura 25. Os blocos são aplicados sequencialmente, ou seja, a saída de um bloco corresponde a entrada do bloco seguinte.

Figura 25 – Arquitetura PPM-DeepLabv3+



Fonte: Adaptado de Chen et al. (2017)

Após a execução do PPM *Encoder*, o mapa de características resultantes é passado para o módulo ASPP. Esse módulo aplica uma pirâmide de *pooling*,

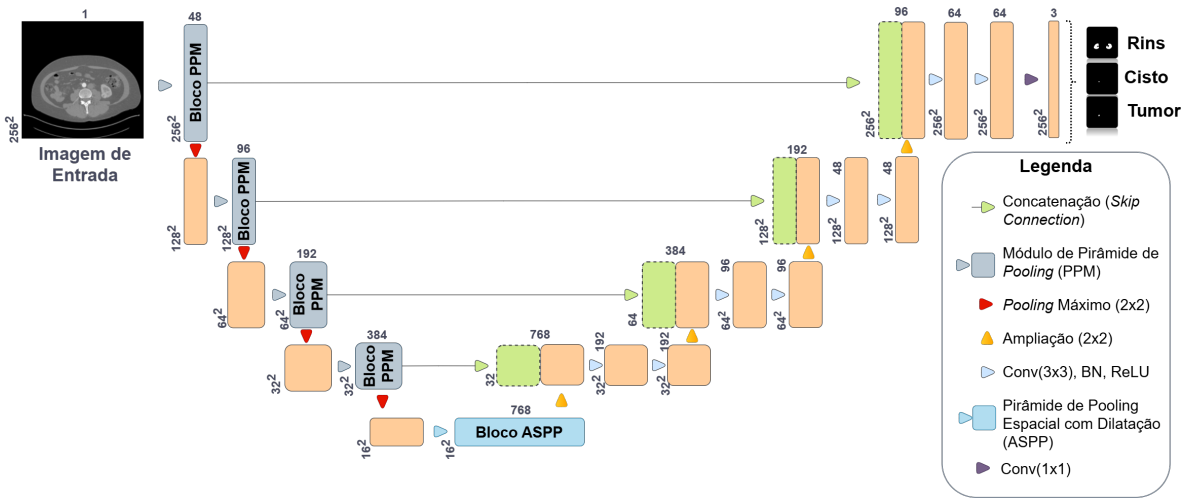
contudo, utilizando convoluções dilatadas (*atrous convolutions*) permitindo a captura de características e a observação de objetos em várias escalas. O módulo ASPP é especialmente projetado para identificar objetos em contextos globais nas imagens (CHEN et al., 2018).

Neste trabalho, a estrutura arquitetural do módulo ASPP e do decodificador da rede foram mantidas, enquanto as adaptações foram realizadas no início do codificador. Dessa forma, as taxas de dilatação e os níveis do módulo ASPP seguiram os parâmetros padrão da rede DeepLabv3+, definidos como $r = 1, 6, 12$ e 18.

4.3.4 Arquitetura Proposta - CPP-UNet

Semelhante à rede U-Net tradicional, o segundo modelo de segmentação proposto, denominado CPP-UNet, incorpora módulos piramidais de *pooling* à sua estrutura. O modelo é composto por duas fases principais: o caminho de contração, localizado no lado esquerdo, e o caminho de expansão, no lado direito. A integração dos módulos piramidais de *pooling* ocorre na etapa de codificação da rede convolucional, aprimorando a extração de características em múltiplas escalas. A Figura 26 apresenta uma visão geral dessa arquitetura.

Figura 26 – Arquitetura CPP-UNet



Fonte: Desenvolvido pelo Autor.

A arquitetura CPP-UNet é uma adaptação da estrutura padrão da rede UNet, substituindo a rede extratora de características do codificador (encoder) convencional pela combinação de Pooling Piramidal. O codificador segue a configuração típica de uma rede convolucional, onde o bloco PPM é aplicado sucessivamente em cada nível. Após cada aplicação do bloco, uma operação de agrupamento máximo (*max-pooling*) com filtro 2x2 e stride 2 é utilizada para reduzir as dimensões do mapa de características. A cada etapa de downsampling, o número de mapas de características é dobrado, aumentando a

capacidade de extração de informações do modelo.

Neste modelo de segmentação, o bloco PPM (*Pyramid Pooling Module*) é implementado conforme descrito na Seção 4.3.1 e ilustrado na Figura 22. Esse bloco é composto por quatro convoluções realizadas simultaneamente no mapa de características de entrada, utilizando os filtros de tamanhos $1x1, 3x3, 5x5$ e $7x7$. Após cada convolução, são aplicados os mecanismos de ativação ReLU e normalização em Lote (BatchNorm). As saídas resultantes das quatro convoluções são então concatenadas, formando mapa de características enriquecido para as etapas subsequentes.

Após a execução do codificador, o mapa de características resultante é processado pelo bloco ASPP *(Atrous Spatial Pyramid Pooling*), que atua como o ponto de compressão ou *bottleneck* da arquitetura U-Net. Esse bloco é detalhado na Seção 4.3.2 e ilustrado na Figura 23. Ele utiliza convoluções dilatadas (*atrous convolutions*) para realizar uma pirâmide de agrupamento, permitindo capturar informações em múltiplas escalas. Essa abordagem é fundamental para identificar objetos em diferentes contextos globais na imagem, como descrito na Seção 3.8.

Na fase de decodificação, o modelo realiza upsampling recorrente no mapa de características, restaurando as dimensões reduzidas na fase de codificação. Cada etapa de upsampling é seguida pela concatenação com o mapa de características correspondente à etapa de codificação, ajustado por recorte. Em seguida, são aplicadas duas convoluções com filtro $3x3$, cada uma seguida pelos mecanismos de ReLU e normalização em lote (*BatchNorm*), garantindo a reconstrução detalhada do mapa de saída e preservando a consistência das características extraídas ao longo do processo.

4.3.5 Funções de Perda

As funções de perda são métricas fundamentais para avaliar o desempenho de um modelo, quantificando a discrepância entre as predições e os rótulos verdadeiros presentes no conjunto de dados. No contexto desta tese, a segmentação de múltiplos artefatos, como rins, tumores e cistos, apresenta desafios, incluindo o desbalanceamento de classes e a dificuldade na diferenciação entre os rótulos.

Para mitigar esses problemas, foram adotadas duas funções de custo: *Dice Categorical Focal Loss* e *Weighted Categorical Crossentropy* (MOSTAFA; ABDEL-GALIL; BELAL, 2021). A primeira função resulta da combinação das funções *Categorical Focal Loss* (LIN et al., 2020) e *Dice Loss* (MILLETARI; NAVAB; AHMADI, 2016), ambas amplamente utilizadas em tarefas de segmentação (MU et al., 2024; YEUNG et al., 2022; WAKAMATSU; ONO, 2023; ISENSEE et al., 2021).

Por outro lado, a função *Weighted Categorical Crossentropy* consiste em uma variação da tradicional *Categorical Crossentropy*, incorporando pesos diferenciados para

cada classe. Essa abordagem visa equilibrar o impacto de classes desbalanceadas no treinamento do modelo, favorecendo a segmentação de estruturas menos representadas.

No conjunto de dados KiTS, observa-se um elevado desbalanceamento, com a classe rins sendo significativamente mais predominante em comparação às classes tumores e cistos. Para lidar com essa discrepância, foi aplicada uma ponderação específica para cada classe em ambas as funções de perda utilizadas: $[0.2, 0.5, 1, 1]$, correspondendo respectivamente às classes fundo, rins, tumores e cistos. Esses valores, variando de 0 a 1, indicam o grau de influência de cada classe no processo de aprendizado, onde valores mais próximos de 1 atribuem maior importância à classe e valores menores reduzem sua contribuição. Essa estratégia visa reduzir a influência da classe majoritária e aumentar a contribuição das classes minoritárias durante o treinamento, promovendo um aprendizado mais equilibrado e sensível às estruturas de menor representatividade.

4.4 Considerações Finais

Este capítulo apresentou em detalhes a metodologia desenvolvida para a segmentação de rins, cistos e tumores em imagens de TC. Cada uma das etapas foram descritas, destacando as adaptações realizadas especificamente para a segmentação destas estruturas renais nesta pesquisa. Entre as principais contribuições, destacam-se: a adaptação das redes DeepLabv3+ e U-Net incoporando os módulos PPM e ASPP como codificador, desenvolvimento assim uma nova arquitetura profunda para segmentação de rins, cistos e tumores renais em imagens de tomografia computadorizadas.

No próximo capítulo, são apresentados os resultados obtidos pela metodologia desta tese. Além disso, são apresentados detalhes sobre a configuração experimental dos modelos de segmentação empregados, experimentos realizados nas bases KiTS21 e KiTS23, e uma análise de estudos de casos para validar as etapas da metodologia.

5 Resultados

Este capítulo apresenta os detalhes da experimentação e avaliação da metodologia proposta. Primeiramente, são descritas as configurações experimentais adotadas, abrangendo a infraestrutura utilizada, os hiperparâmetros definidos, as métricas de validação de desempenho e os cenários de experimentação. Em seguida, são fornecidos os resultados obtidos pela metodologia para a segmentação de rins, cistos e tumores, com base nos conjuntos de dados KiTS21 e KiTS23. Por fim, é realizada uma análise comparativa com trabalhos relacionados.

5.1 Configurações Experimentais

Infraestrututura: A metodologia desta tese foi implementada utilizando a linguagem de programação Python e principalmente a biblioteca de aprendizado profundo Keras (GULLI; PAL, 2017). Os experimentos foram realizados em um computador equipado com uma CPU Intel Core i7-7700K de 4.20 GHz, 16 GB de RAM e uma placa de vídeo Nvidia GeForce GTX 1080-Ti, 12 GB de VRAM, executando o sistema operacional Ubuntu 20.

Hiperparâmetros: Os hiperparâmetros utilizados nos modelos de segmentação foram: *batch* igual a 8, otimizador estocástico Adam com taxa de aprendizagem inicial de 10^{-4}. Aplica-se um total de 50 épocas utilizando a técnica *Early Stop* (BROWNLEE, 2018). Este mecanismo possibilita interromper o treinamento da rede quando a métrica monitorada (valor da função de perda nos dados de validação) não apresentar incremento por 15 épocas consecutivas, sendo este valor estabelecido empiricamente.

Métricas de Avaliação: As métricas utilizadas para avaliar o desempenho do método proposto e dos modelos reimplementados foram o coeficiente de similaridade Dice (DSC), o índice de Jaccard (JCC), conforme descrito na Seção de Métricas (Seção 3.10).

Cénario de Experimentação: Os experimentos conduzidos nesta tese foram realizados utilizando os conjuntos de dados KiTS21 e KiTS23. O estudo de ablação foi desenvolvido por meio da aplicação de diversos modelos de segmentação, incluindo CPP-UNet, PPM-Deeplabv3+, U-Net, PPM-UNet, Deeplabv3+ e Deeplab sem ASPP (SUNANDINI et al., 2023), sendo os dois primeiros propostos nesta pesquisa. Os resultados obtidos são apresentados detalhadamente, considerando as diferentes tarefas de segmentação, que abrangem rins, cistos e tumores renais. Além disso, casos de erros e acerto são ilustrados referente a aplicação dos modelos desenvolvidos.

5.2 Experimentos com a base KiTS21

Esta seção apresenta em detalhes os resultados obtidos com a aplicação dos métodos de segmentação desenvolvidos nesta pesquisa para a tarefa de segmentação de rins, cistos e tumores renais, utilizando o conjunto de dados KiTS21. Em seguida, são exibidos estudos de caso que possibilitam uma análise qualitativa por meio da visualização das segmentações geradas pelos modelos, destacando tanto casos de sucesso quanto de falhas.

5.2.1 Preparação da Base

Segundo a Seção 4.1.1, o conjunto de dados KiTS21 possui um total de 300 exames, abrangendo aproximadamente 65.164 fatias de imagem. Para realizar os experimentos, os volumes de TC foram divididos nas proporções $70\%/20\%/10\%$ para treinamento, teste e validação, respectivamente. A Tabela 3 destaca a quantidade de fatias para cada um dos conjuntos de dados.

Tabela 3 – Divisão do conjuto de dados para treinamento, teste e validação na base KiTS21.

	Divisão Conjunto de Dados KiTS21		
	Proporção	Pacientes	Fatias
Treinamento	70%	210	45.820
Teste	20%	60	13.805
Validação	10%	30	5.539
Total	100%	300	65.164

É importante destacar que a seleção dos exames para os conjuntos de treinamento, teste e validação, em ambas as bases de dados, foi realizada de forma aleatória, visando garantir a imparcialidade na avaliação dos modelos. Além disso, o critério adotado para a divisão dos conjuntos baseou-se na separação por paciente, ou seja, todas as imagens pertencentes a um mesmo paciente foram alocadas exclusivamente em um único conjunto (treinamento, teste ou validação). Esse procedimento visa assegurar que os modelos treinados não sejam expostos a amostras do mesmo exame em diferentes etapas do processo, evitando assim possíveis vieses e garantindo uma avaliação mais fidedigna do desempenho do modelo.

5.2.2 Resultados

Os resultados obtidos para a segmentação de estruturas renais e patologias associadas foram baseados na aplicação dos modelos de segmentação CPP-UNet e PPM-Deeplabv3+, desenvolvidos nesta tese, ao conjunto de dados KiTS21. Com o intuito de avaliar a eficácia dos módulos baseados em pirâmide de *pooling* na extração de

características relevantes para o câncer renal, foram também considerados outros modelos que incorporam esses módulos, como o PPM-UNet e o Deeplabv3+.

As métricas de avaliação dos modelos, representadas pelos coeficientes de similaridade Dice (DSC) e Jaccard (JCC), foram utilizadas conforme descrito na Seção de Métricas (Seção 3.10). Adicionalmente, as classes avaliadas incluem rins, cistos, tumores e fundo (*background* — BG).

A Tabela 4 ilustra uma visão geral dos resultados alcançados pelos modelos de segmentação propostos nesta tese para segmentação de rins, cistos e tumores renais, apresentando os coeficientes de Dice e Jaccard obtidos em diferentes cenários experimentais. Esses resultados permitem uma comparação direta do desempenho entre os modelos desenvolvidos e outras abordagens de referência, evidenciando a eficácia da metodologia.

Tabela 4 – Visão geral dos resultados dos modelos de segmentação desenvolvidos (PPM-Deeplabv3+ e CPP-UNet) em comparação com outros modelos aplicados na base KiTS21.

Método	**Dice Categorical Focal Loss**									
	Dice (DSC)					**Jaccard (JCC)**				
	BG	**Rim**	**Tumor**	**Cisto**	**Médio**	**BG**	**Rim**	**Tumor**	**Cisto**	**Médio**	**Loss**
U-Net Padrão	99,86%	92,60%	83,59%	91,34%	91,82%	99,73%	90,74%	83,59%	86,88%	90,24%	0,069711
Deeplabv3+ Sem ASPP	99,60%	94,88%	84,68%	64,19%	85,84%	99,21%	94,88%	84,68%	64,19%	85,74%	0,14960
Deeplabv3+ Padrão	99,92%	93,31%	85,11%	94,69%	93,26%	99,84%	90,95%	84,73%	94,67%	92,54%	0,071176
PPM-UNet	99,94%	94,59%	86,95%	95,55%	94,26%	99,88%	92,59%	86,22%	95,55%	93,56%	0,060287
PPM-Deeplabv3+	99,94%	90,27%	87,65%	95,84%	93,42%	99,87%	87,85%	86,94%	95,74%	92,60%	0,003896
CPP-UNet	99,93%	94,20%	87,84%	94,91%	94,22%	99,86%	92,35%	87,28%	94,91%	93,60%	0,059685
Método	**Weighted Categorical Crossentropy**									
	Dice (DSC)					**Jaccard (JCC)**				
	BG	**Rim**	**Tumor**	**Cisto**	**Médio**	**BG**	**Rim**	**Tumor**	**Cisto**	**Médio**	**Loss**
Deeplabv3+ Sem ASPP	99,73%	96,03%	83,61%	81,38%	90,19%	99,46%	96,03%	83,61%	78,71%	89,45%	0,103430
U-Net Padrão	99,79%	94,89%	84,63%	82,21%	90,38%	99,59%	94,89%	84,63%	79,43%	89,63%	0,10044
Deeplabv3+ Padrão	99,91%	93,31%	85,11%	94,68%	93,25%	99,83%	90,94%	84,72%	94,66%	92,54%	0,071176
PPM-UNet	99,94%	92,12%	86,75%	94,85%	93,42%	99,88%	89,72%	85,82%	94,72%	92,53%	0,003342
PPM-Deeplabv3+	99,94%	93,53%	87,06%	94,80%	93,83%	99,87%	91,21%	86,14%	94,64%	92,97%	0,003358
CPP-UNet	99,93%	91,23%	87,24%	95,31%	93,43%	99,86%	88,82%	86,49%	95,17%	92,59%	0,003587

De maneira geral, ao analisar da Tabela 4 observa-se que os modelos que não incorporam mecanismos de pirâmide de pooling, U-Net e Deeplabv3+ sem o módulo ASPP, apresentaram médias inferiores nos valores do coeficiente Dice e do índice de Jaccard em comparação com os modelos baseados em pirâmide de *pooling*. Esse comportamento é consistente em ambas as funções de perda avaliadas: *Dice Categorical Focal Loss* e *Weighted Categorical Crossentropy Loss*.

Em detalhes, os modelos U-Net e Deeplabv3+ sem o módulo ASPP apresentaram coeficientes médios de Dice e Jaccard de 91,82% e 91,18% para o U-Net, e 85,84% e 85,74% para o Deeplabv3+ sem ASPP, respectivamente. Por sua vez, os modelos baseados em pirâmide de *pooling* obtiveram coeficientes médios de Dice variando entre 93,25%, alcançados pela Deeplabv3+ Padrão, e 94,22%, obtidos pelo CPP-UNet desenvolvido nesta pesquisa. Esses resultados demonstram o potencial das abordagens

convolucionais baseadas em pirâmides de pooling em aprimorar a capacidade de extração de características em imagens de tomografia computadorizada.

Em contraste, na tarefa de segmentação dos rins, os modelos que não utilizam mecanismos de pirâmide de *pooling* apresentaram desempenho superior em comparação aos demais, alcançando coeficientes de Dice de 94,88% e 96,03% para os modelos U-Net padrão e Deeplabv3+ sem o módulo ASPP, respectivamente. Esse resultado pode ser atribuído ao fato de que os rins apresentam características morfológicas mais homogêneas em exames de tomografia, com variações mínimas em termos de tamanho e localização, favorecendo modelos mais simples e diretos. Por outro lado, os módulos de pirâmide de pooling, que utilizam múltiplas resoluções e convoluções dilatadas (como no ASPP), demonstraram desempenho inferior, possivelmente devido à sua maior sensibilidade a variações estruturais que não são tão relevantes para a segmentação de órgãos com morfologia previsível. Ainda assim, o modelo CPP-UNet obteve resultados próximos aos do U-Net padrão, alcançando 94,20% de Dice ao utilizar a função de perda *Dice Categorical Focal Loss*.

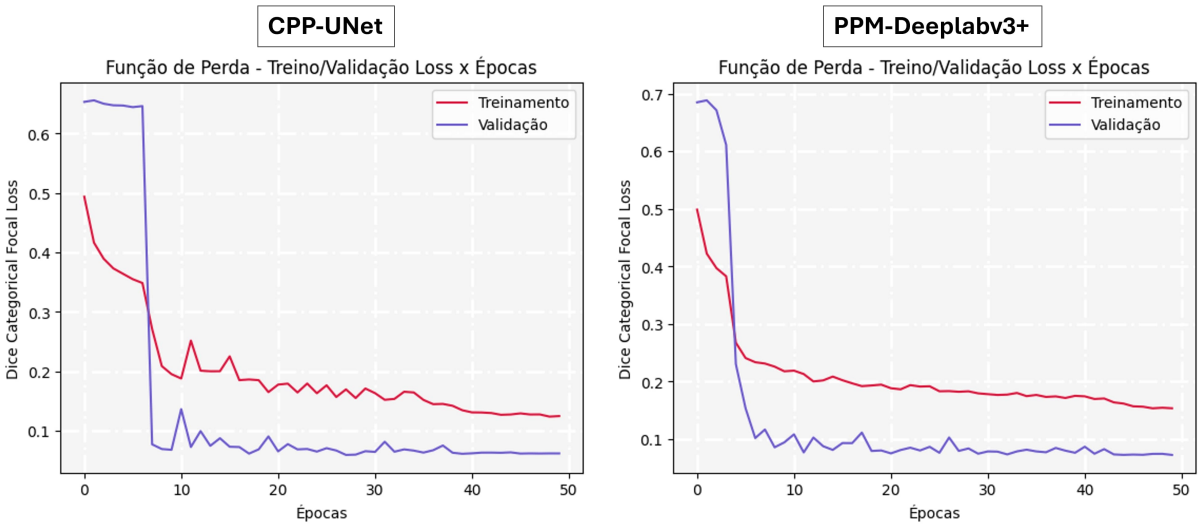
Na segmentação de cistos renais, observa-se um aumento nos coeficientes de Dice e Jaccard com a utilização de modelos baseados em pirâmides de pooling. Especificamente, o modelo U-Net padrão apresentou o melhor desempenho entre as abordagens que não incorporam módulos de pirâmide, alcançando 91,82% de Dice ao utilizar a função de perda *Dice Categorical Focal Loss*. Em comparação, o modelo CPP-UNet obteve um desempenho superior, com um aumento de aproximadamente $4,0$ pontos percentuais em relação ao U-Net, atingindo 95,31% de Dice quando aplicada a função de perda *Weighted Categorical Crossentropy Loss*. Esses resultados destacam o potencial dos modelos baseados em pirâmides de *pooling* na detecção e segmentação de estruturas renais.

A segmentação de tumores renais mostrou-se desafiadora em comparação às demais estruturas analisadas, devido à ampla variabilidade em termos de tamanho e forma. Especificamente, os modelos Deeplabv3+ Padrão, PPM-UNet e PPM-Deeplabv3+ apresentaram desempenho superior em relação aos modelos U-Net e Deeplabv3+ sem o módulo ASPP. Considerando a função de *Dice Categorical Focal Loss* os modelos obtiveram coeficientes de Dice de 85,11%, 86,95% e 87,65%, e índices de Jaccard de 84,73%, 86,22% e 86,94%, respectivamente. Por outro lado, o modelo CPP-UNet demonstrou resultados ainda mais expressivos na segmentação de tumores renais, atingindo 87,84% para o coeficiente Dice e 87,28% para o índice de Jaccard. Esses valores representam um ganho de aproximadamente $2,8$ pontos percentuais em relação ao modelo PPM-UNet, considerando ambas as funções de perda avaliadas.

As curvas de perda fornecem informações valiosas sobre o processo de otimização, ajudando a identificar o comportamento do modelo no processo de aprendizagem. Desta

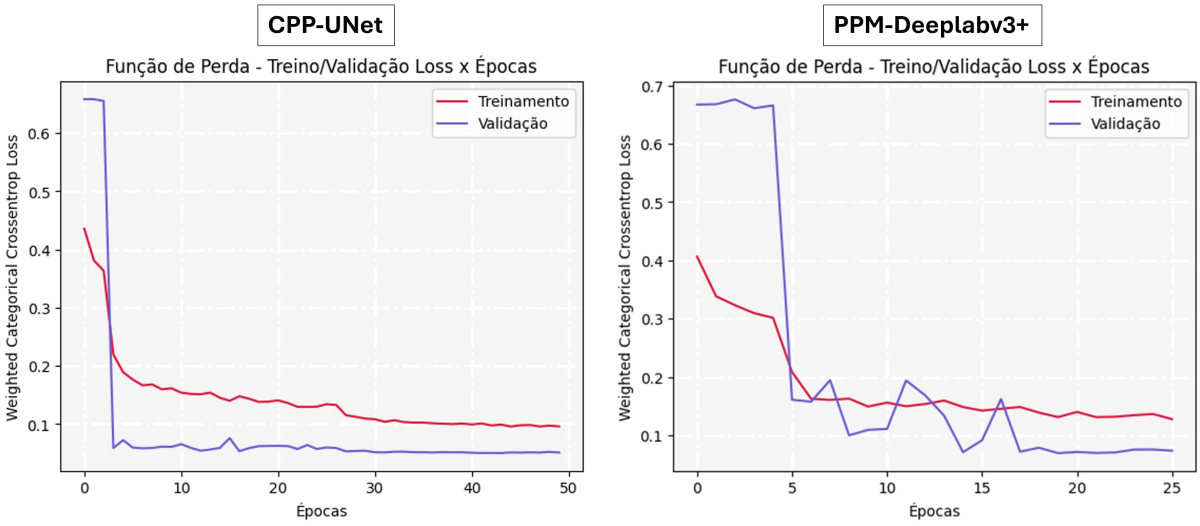
forma, as Figuras 27 e 28 ilustram as curvas de perda para as funções *Dice Categorical Focal Loss* e *Weighted Categorical Crossentropy Loss* voltadas ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS21.

Figura 27 – Curvas da função de perda *Dice Categorical Focal Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS21.



Fonte: Desenvolvido pelo Autor

Figura 28 – Curvas da função de perda *Weighted Categorical Crossentropy Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS21.



Fonte: Desenvolvido pelo Autor

Os resultados obtidos pelos modelos desenvolvidos nesta tese, CPP-UNet e PPM-Deeplabv3+, indicam que a combinação e adaptação dos módulos de pirâmide de *pooling* ASPP e PPM quando associados ao codificador da rede convolucional consegue extrair características de estruturas associadas aos cistos e tumores renais em múltiplas

resoluções, possibilitando assim a identificação de tumores com tamanhos variados. Além disso, esses resultados demonstram o potencial das abordagens convolucionais baseadas em pirâmides de *pooling* em aprimorar a capacidade de extração de características em imagens de tomografia computadorizada, validando a Questão de Pesquisa **Q1** apresentada na Seção 1.2.

5.2.3 Estudos de Casos

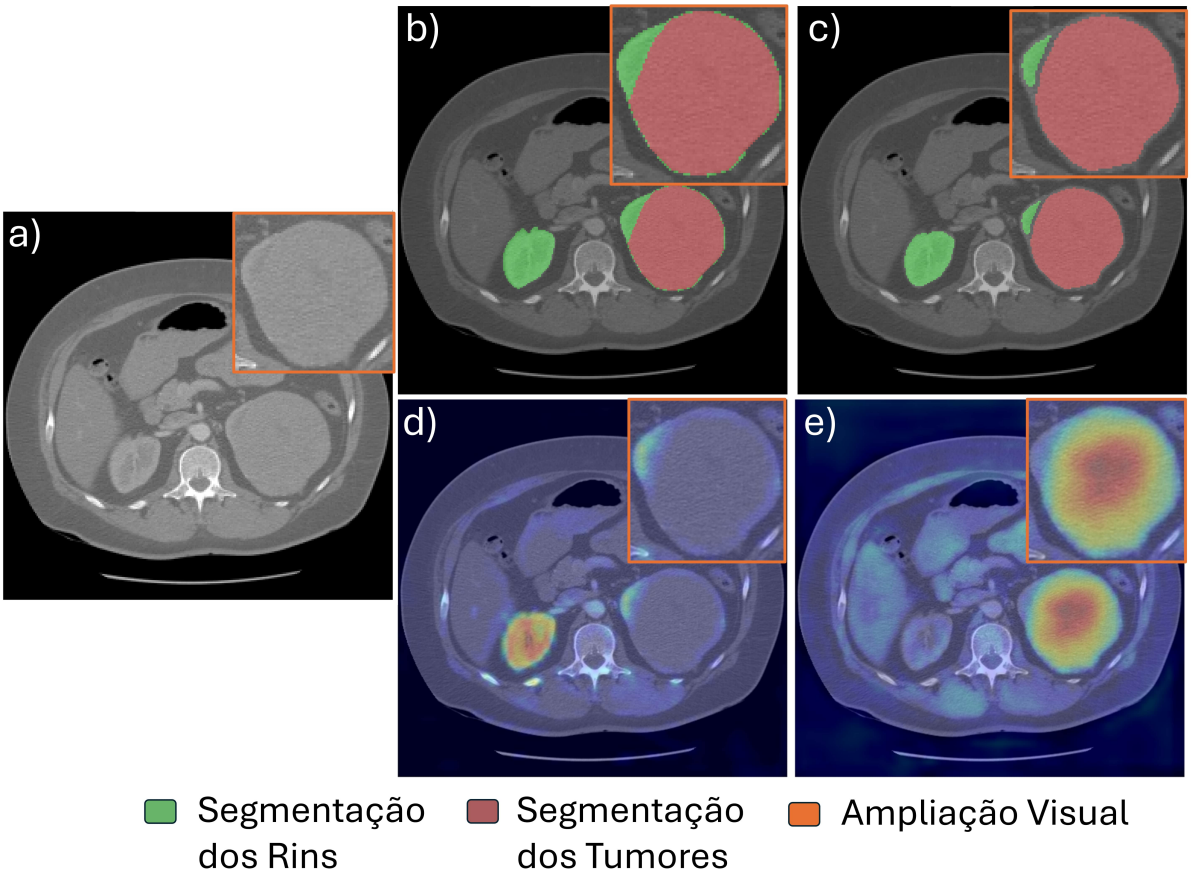
Esta subseção apresenta o estudo de casos voltados para a segmentação de rins, cistos e tumores renais, visando analisar qualitativamente o desempenho dos modelos desenvolvidos nesta tese. A partir da visualização de exemplos representativos e aplicação do método de explicabilidade, como Grad-CAM (SELVARAJU et al., 2017), são discutidos aspectos relevantes relacionados ao desempenho da segmentação, destacando tanto os acertos quanto as limitações observadas em diferentes cenários. Essa abordagem complementa a avaliação quantitativa previamente apresentada, permitindo uma compreensão mais aprofundada da capacidade dos modelos em lidar com variações morfológicas e complexidades presentes nas imagens de tomografia computadorizada.

Em detalhes, a técnica de explicabilidade Grad-CAM foi aplicada aos modelos PPM-Deeplabv3+ e CPP-UNet. Essa abordagem auxilia a análise e interpretação do comportamento das redes neurais, permitindo a identificação das regiões mais relevantes para as predições do modelo. Para o modelo PPM-Deeplabv3+, a saída do codificador foi utilizada para geração do mapa de calor (*heatmap*), especificamente, a camada subsequente à etapa de concatenação, correspondente a uma convolução 3×3, conforme descrito na Seção 4.3.3. Por outro lado, no modelo CPP-UNet, o mapa de ativação foi gerado a partir da saída do bloco ASPP, que representa a etapa final do codificador da arquitetura, conforme detalhado na Seção 4.3.4. A escolha dessas camadas visa avaliar se as características extraídas durante a fase de codificação dos modelos foram eficazes na identificação das estruturas renais, incluindo rins, cistos e tumores.

5.2.3.1 Casos de Sucesso

Exame 00025 — Neste exame de tomografia computadorizada, disponível na base KiTS21, contém marcações dos rins e tumores renais efetuadas pelo médico especialista. A Figura 29 ilustra as regiões segmentadas destacadas em verde para os rins e em vermelho para os tumores renais. Além disso, é apresentado um exemplo de segmentações bem-sucedidas de rins e tumores renais realizadas pelo modelo PPM-DeepLabv3+ para o exame 00025. Observa-se que o modelo pôde segmentar de maneira correta as estruturas dos rins e tumor renal aproximando-se das anotações realizadas pelos especialistas, conforme evidenciado na fatia de imagem do exame.

Figura 29 – Exemplo de segmentação bem-sucedida dos rins e do tumor renal utilizando o modelo PPM-DeepLabv3+. (a) Fatia original do exame 00025; (b) Segmentação efetuada pelo especialista (*GroundTruth*), (c) Segmentação produzida pelo modelo,(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente.



Fonte: Desenvolvido pelo Autor

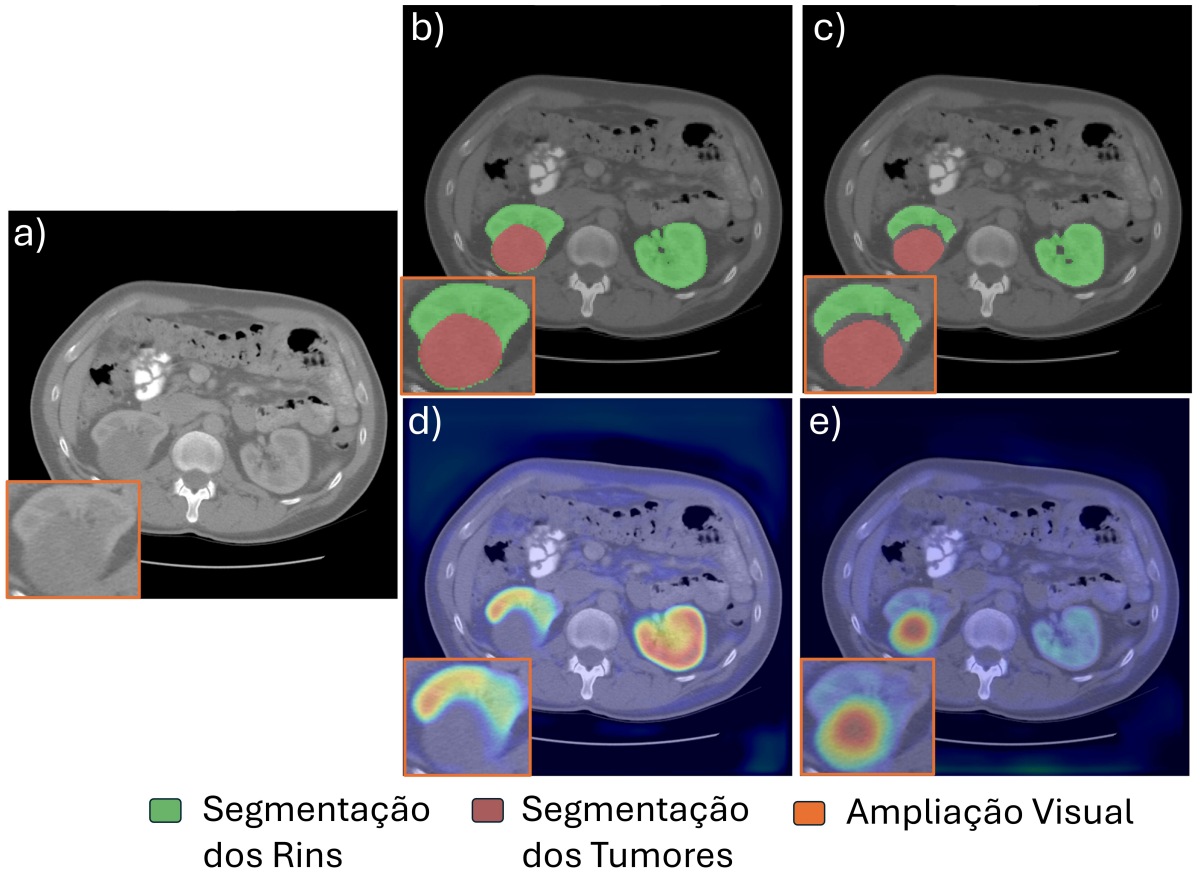
Considerando ainda o Exame 00025, a Figura 29 apresenta a aplicação da técnica Grad-CAM ao modelo PPM-Deeplabv3+, ilustrando a sobreposição do mapa de ativação gerado para a segmentação dos rins e do tumor renal sobre uma fatia de tomografia computadorizada (TC) do paciente 00025. Observa-se que as ativações do modelo estão fortemente concentradas na região dos rins e tumor renal, evidenciando uma alta correspondência entre as áreas destacadas no mapa de ativação e a segmentação esperada.

É relevante ressaltar que o modelo conseguiu distinguir as características do rim mesmo em uma estrutura onde o tumor renal apresenta um tamanho significativamente maior. Apesar da predominância do tumor na região, o modelo conseguiu identificar e segmentar corretamente o rim, evidenciando sua capacidade de diferenciar essas estruturas, mesmo em cenários desafiadores. Desta forma, os resultados demonstram a

capacidade do modelo em identificar a estrutura renal e tumores associados, reforçando a eficácia da abordagem na extração de características relevantes para a segmentação.

Exame 00099 — Por outro lado, o modelo CPP-UNet apresentou resultados relevantes na segmentação das estruturas renais e patologias associadas, tumores e cistos para o paciente com exame 00099. A Figura 30 evidencia, como forma de análise qualitativa, uma comparação entre a segmentação gerada pelo modelo para rins e tumores e a segmentação realizada pelo especialista. Observa-se que a segmentação final produzida pelo modelo possui correspondência significativa com a segmentação do especialista. Além disso, o tumor renal identificado no exame 00099 foi segmentado quase em sua totalidade, ressaltando a eficácia da abordagem na identificação e delimitação de estruturas tumorais.

Figura 30 – Exemplo de segmentação correta dos rins e do tumor renal utilizando o modelo CPP-UNet.(a) Fatia original do exame 00099; (b) Segmentação efetuada pelo especialista (*GroundTruth*), (c) Segmentação produzida pelo modelo,(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente.



Fonte: Desenvolvido pelo Autor

Aplicação da técnica Grad-CAM ao modelo CPP-UNet é ilustrada na Figura 30. Os

resultados obtidos demonstram que o modelo CPP-UNet conseguiu ativar corretamente as regiões correspondentes aos rins e tumores. Além disso, indica que o modelo de segmentação desenvolvido nesta tese possui a capacidade de identificar as áreas que podem estar relacionadas às estruturas renais e tumorais. É importante ressaltar que, na segmentação do tumor, as áreas de maior ativação estão concentradas na massa tumoral, sugerindo que o modelo tem o poder de diferenciar o tumor de outros órgãos presentes na imagem. Essa habilidade contribui para maior precisão na segmentação e análise clínica.

Exame - 00194 — A tarefa de segmentação de cistos renais é de grande relevância no contexto clínico por permitir a identificação dessas lesões, facilitando seu acompanhamento e diagnóstico diferencial em relação a outras patologias renais, como tumores malignos. Desta forma, a Figura 31 ilustra um exemplo de segmentação de rim e cisto renal efetuada pelo modelo CPP-UNet para o exame 00194. Os resultados evidenciam que, em comparação com a segmentação realizada pelo especialista (*Ground Truth*), o modelo conseguiu identificar os contornos das estruturas correspondentes ao rim e ao cisto, realizando uma segmentação correta.

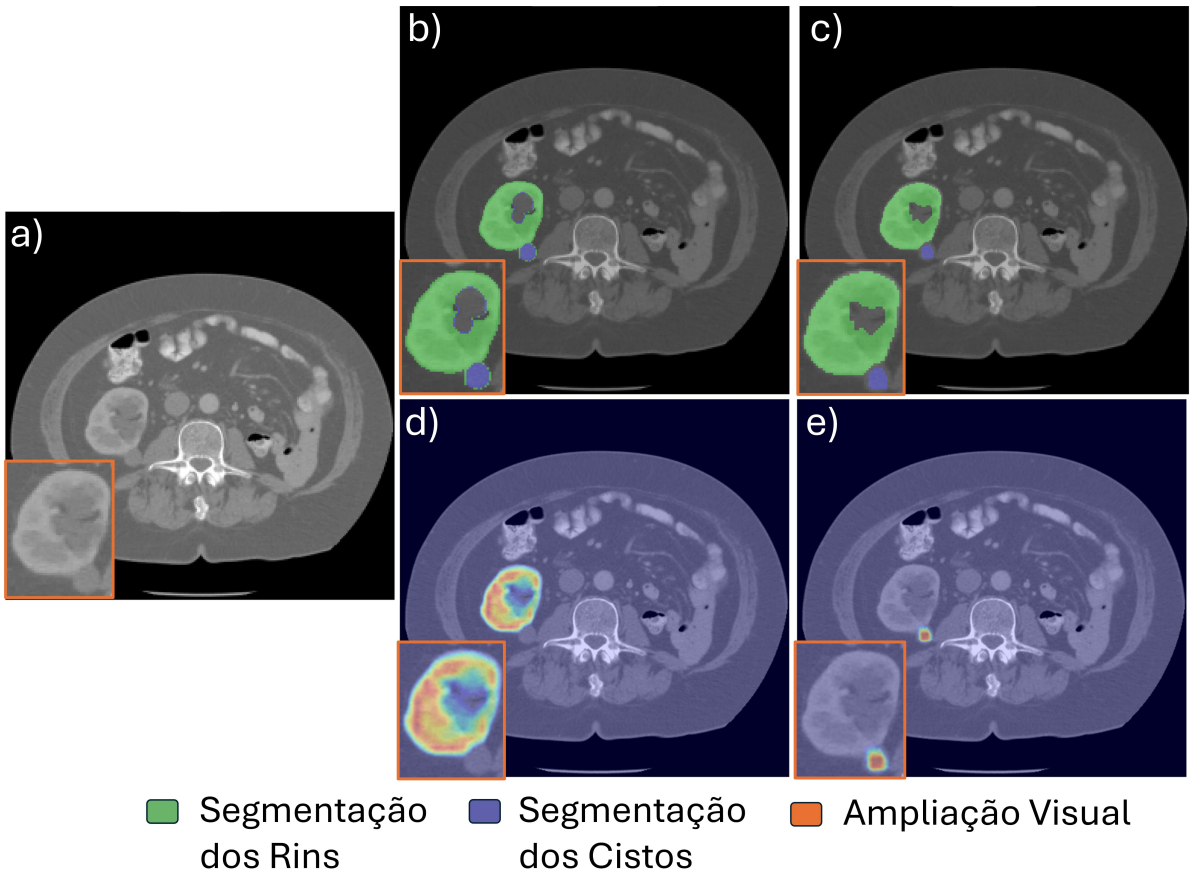
Como forma de avaliar as informações consideradas para efetuar a segmentação do rim e cisto no exame 000194 a técnica de explicabilidade Grad-CAM também foi aplicada neste caso. Analisando o resultado obtido, destacado na Figura 31, observa-se que as regiões do rim esquerdo e cisto associado foram corretamente ativadas pelo modelo CPP-UNet contribuindo assim para a correta segmentação. Vale destacar também que as regiões de ativação com maior intensidade estão concentradas nas regiões do rim e cisto renal, evidenciando a capacidade de segmentação destas estruturas.

5.2.3.2 Casos de Erro

Exame - 00221 - Por outro lado, em alguns casos, o modelo proposto encontrou certas dificuldades na identificação e segmentação dos rins, cistos e tumores. Ao analisar o exame 00221 apresentado na Figura 32, pode-se observar o modelo PPM-Deeplabv3+ não obteve sucesso em segmentar os dois rins presentes na fatia de imagem, identificando apenas o rim esquerdo do paciente. Por outro lado, em termos de tumor, para este paciente, o modelo obteve uma segmentação correta, identificando o tumor localizado próximo ao rim direito. Estas observações podem ser analisadas de maneira mais aprofundada através do mapa de ativação obtido para este paciente ilustrado na Figura 32. Observa-se que, embora o modelo não tenha apresentado uma ativação intensa para o rim à esquerda no exame, ele destacou essa estrutura com maior relevância em comparação às demais regiões e órgãos. Esse resultado sugere que mais épocas de treinamento poderiam aprimorar a capacidade do modelo em identificar essa estrutura com maior precisão.

Exame — 00289 — O modelo CPP-UNet também apresentou limitações na

Figura 31 – Exemplo de segmentação correta dos rim e do cisto renal utilizando o modelo CPP-UNet;(a) Fatia original do exame 00194; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente.



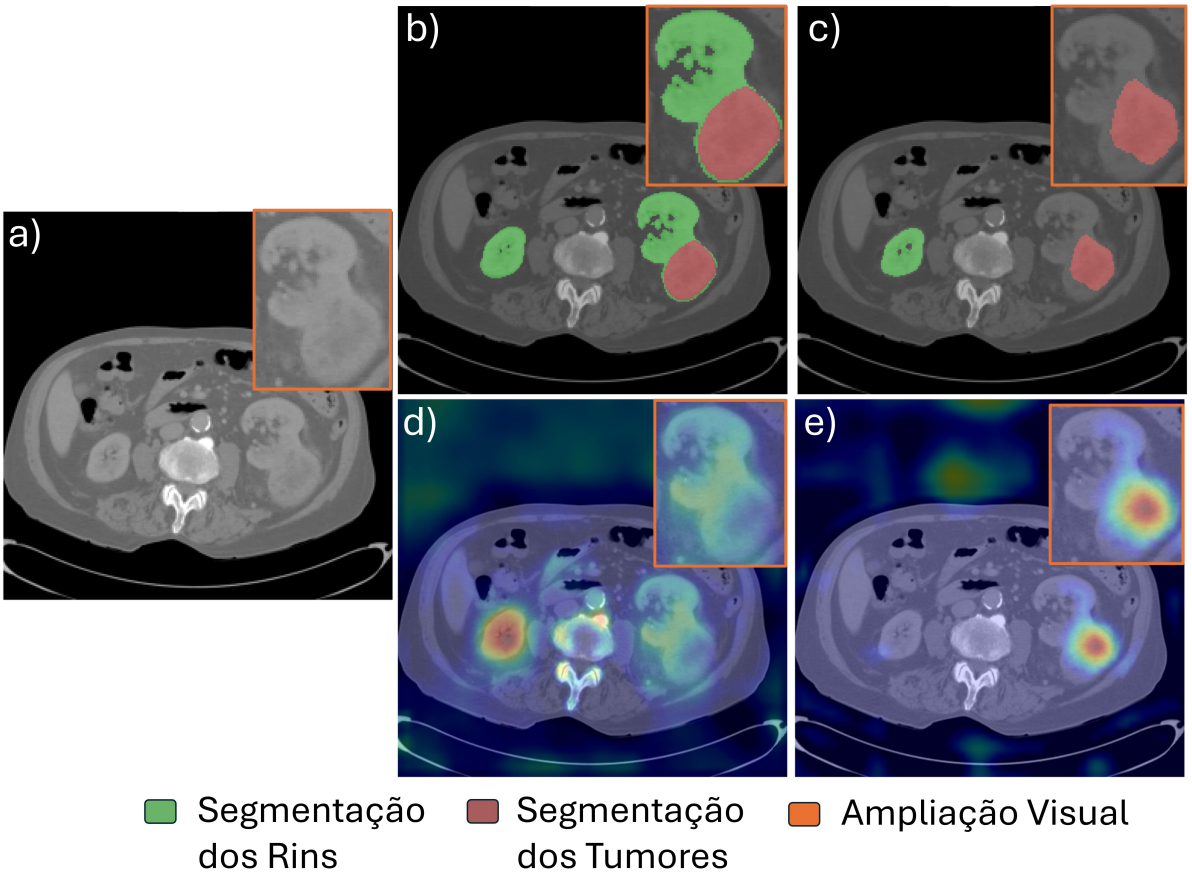
Fonte: Desenvolvido pelo Autor

segmentação de alguns casos, incluindo o Exame 00221, que contém tumores e cistos renais de dimensões reduzidas. Desta forma, observando a Figura 33 nota-se que o modelo não pôde identificar o cisto e o tumor localizados nos rins esquerdo e direito, respectivamente. Acredita-se que, devido ao tamanho reduzido e à semelhança com outras estruturas abdominais, incluindo os rins, não foi possível obter características discriminatórias dessas classes, contribuindo de maneira negativa ao modelo.

Destaca-se ainda, as dimensões dos cistos e tumores renais para este paciente permanecem reduzidas em todas as fatias do exame, tornando a inferência ainda mais desafiadora considerando que o modelo CPP-UNet possui sua arquitetura baseada em múltiplas resoluções tendo maior desempenho em casos onde há maior variabilidade de tamanho entre os exemplares das classes a serem segmentadas.

Esse fato pode ser observado de maneira mais detalhada por meio da análise do

Figura 32 – Segmentação incorreta do rim presente no Exame 00221 a partir do modelo PPM-Deeplabv3+. (a) Fatia original do exame 00221; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumores à imagem de tomografia computadorizada, respectivamente.



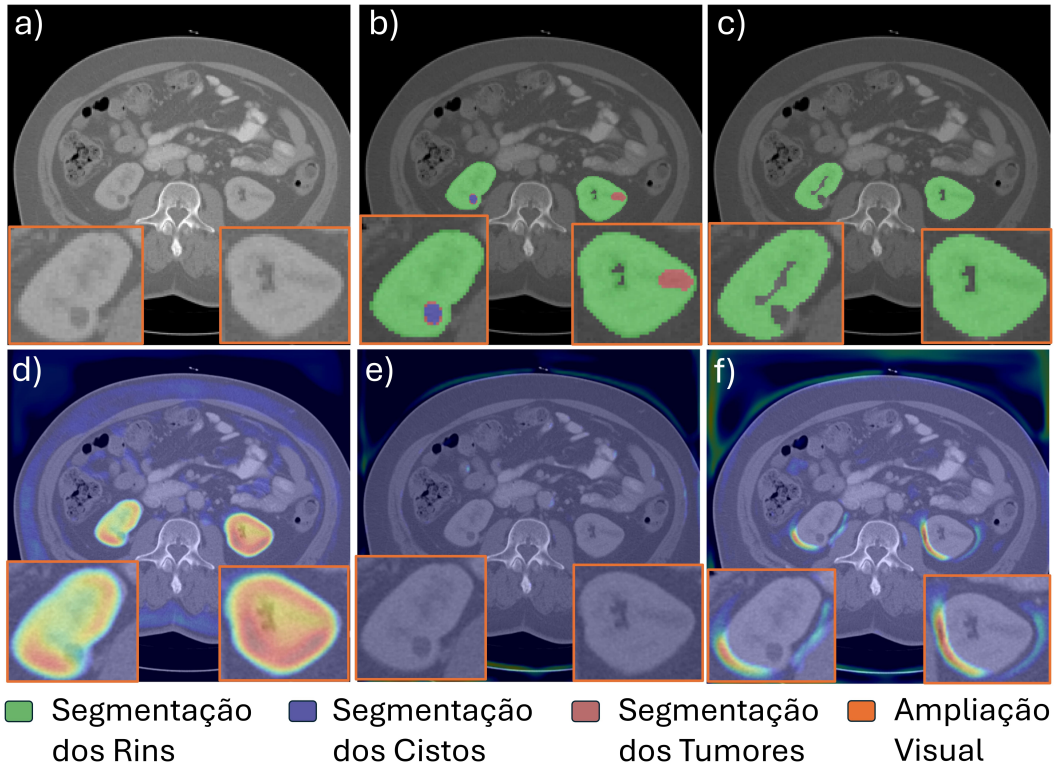
Fonte: Desenvolvido pelo Autor

heatmap gerado, ilustrado na Figura 33. As regiões foram corretamente ativadas para a classe dos rins, onde se apresenta uma intensidade maior, delimitando ambos os rins, esquerdo e direito. Esse comportamento foi possível pelo fato dos rins estarem bem definidos na fatia do exame e não apresentarem grandes deformações. Por outro lado, percebe-se que não houve ativação do modelo para as classes cisto e tumor renal, evidenciando a semelhança entre essas estruturas e os rins onde estão localizadas.

5.3 Experimentos com a base KiTS23

Esta seção detalha os resultados obtidos a partir da aplicação dos métodos de segmentação desenvolvidos nesta pesquisa na tarefa de segmentação de rins, cistos e tumores renais, utilizando o conjunto de dados KiTS23. Para avaliar o desempenho dos modelos, foram utilizadas as métricas de similaridade Dice e Jaccard, que permitem medir

Figura 33 – Segmentação incorreta do cisto e tumor renal presentes no Exame 00289 a partir do modelo CPP-UNet. (a) Fatia original do exame 00289; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) - (f) Sobreposições dos mapas de ativação dos rins, tumores e cistos à imagem de tomografia computadorizada, respectivamente.



Fonte: Desenvolvido pelo Autor

a sobreposição entre as segmentações geradas e as anotações de referência. Em seguida, são apresentados estudos de caso que oferecem uma análise qualitativa por meio da visualização das segmentações produzidas, evidenciando exemplos de segmentações bem-sucedidas, bem como situações em que ocorreram erros, visando identificar padrões de acerto e limitações do modelo.

5.3.1 Preparação da Base

Conforme descrito na Seção 4.1.2 o desafio KiTS disponibiliza 599 exames para o treinamento e avaliação dos modelos participantes da versão KiTS23. O desafio ainda especifica que 489 casos são destinados ao treinamento e 110 casos para teste e avaliação. No entanto, o conjunto de teste é reservado exclusivamente para a avaliação dos métodos submetidos ao desafio, sendo acessível apenas às equipes participantes. Assim, para os experimentos desta tese, os 489 casos disponibilizados publicamente foram subdivididos conforme a seguinte proporção: 70% (342 exames) para treinamento,

20% (97 exames) para teste e 10% (48 exames) para validação. A Tabela 5 apresenta em detalhes a distribuição adotada nos experimentos.

Tabela 5 – Divisão do conjunto de dados para treinamento, teste e validação na base KiTS23.

	Divisão Conjunto de Dados KiTS23		
	Proporção	Pacientes	Fatias
Treinamento	70%	342	66.655
Teste	20%	98	19.044
Validação	10%	49	9.522
Total	100%	489	95.221

5.3.2 Resultados

Esta subseção apresenta os resultados obtidos pelos modelos de segmentação desenvolvidos nesta tese ao serem aplicados ao conjunto de dados KiTS23. Seguindo a mesma abordagem da Seção 5.2, foram avaliados os modelos U-Net e Deeplabv3+ sem o módulo ASPP, que não incorporam mecanismos de pirâmide de *pooling* em suas arquiteturas. Além disso, foram testados os modelos PPM-UNet, Deeplabv3+ Padrão, PPM-Deeplabv3+ e CPP-UNet, os quais fazem uso dessa estratégia para aprimorar a extração de características relevantes na segmentação de estruturas renais e patologias associadas. Em termos de métricas de avaliação de desempenho, foram adotados os coeficientes de similaridade Dice (DSC) e Jaccard (JCC), conforme detalhado na Seção de Métricas (Seção 3.10). As avaliações foram realizadas considerando as classes de interesse, que incluem os rins, cistos, tumores e o fundo da imagem (*background* — BG). Vale ressaltar que todos os modelos mencionados anteriormente foram treinados com as funções de perda *Dice Categorical Focal Loss* e *Weighted Categorical Crossentropy Loss*. A Tabela 6 apresenta uma visão geral dos resultados obtidos na base KiTS23 para ambas as funções de perda.

Tabela 6 – Visão geral dos resultados dos modelos de segmentação desenvolvidos (PPM-Deeplabv3+ e CPP-UNet) em comparação com outros modelos aplicados na base KiTS23.

Método	**Dice Categorical Focal Loss**										
	Dice (DSC)					**Jaccard (JCC)**					
	BG	**Rim**	**Tumor**	**Cisto**	**Médio**	**BG**	**Rim**	**Tumor**	**Cisto**	**Médio**	**Loss**
Deeplabv3+ Sem ASPP	99,92%	93,10%	87,57%	93,64%	93,56%	99,85%	91,18%	86,95%	93,55%	92,88%	0,067529
U-Net Padrão	99,94%	94,37%	87,98%	93,52%	93,95%	99,87%	92,65%	87,18%	93,52%	93,30%	0,062889
PPM-UNet	99,85%	93,12%	84,19%	93,52%	92,67%	99,71%	91,25%	84,19%	93,52%	92,17%	0,076569
Deeplabv3+ Padrão	99,94%	93,23%	87,59%	94,35%	93,78%	99,87%	91,18%	86,71%	94,35%	93,03%	0,064528
PPM-Deeplabv3+	99,96%	95,17%	89,45%	96,25%	95,21%	99,92%	93,50%	88,84%	96,19%	94,61%	0,051018
CPP-UNet	99,96%	95,23%	**91,44%**	94,67%	95,32%	99,93%	93,66%	90,77%	94,67%	94,76%	0,047941
Método	**Weighted Categorical Crossentropy**										
	Dice (DSC)					**Jaccard (JCC)**					
	BG	**Rim**	**Tumor**	**Cisto**	**Médio**	**BG**	**Rim**	**Tumor**	**Cisto**	**Médio**	**Loss**
PPM-UNet	99,88%	89,42%	78,78%	93,28%	90,34%	99,77%	87,34%	78,00%	93,18%	89,57%	0,007113
Deeplabv3+ Padrão	99,92%	92,07%	85,66%	93,01%	92,67%	99,85%	89,78%	84,75%	92,88%	91,82%	0,007453
Deeplabv3+ Sem ASPP	99,95%	91,57%	85,83%	92,76%	92,53%	99,89%	89,26%	84,83%	92,44%	91,60%	0,004481
U-Net Padrão	99,94%	93,53%	86,85%	93,52%	93,46%	99,87%	91,71%	86,11%	93,52%	92,80%	0,024354
PPM-Deeplabv3+	99,89%	90,81%	83,59%	94,19%	92,12%	99,78%	88,28%	82,89%	94,19%	91,28%	0,084479
CPP-UNet	99,93%	94,36%	**88,17%**	95,99%	94,61%	99,85%	92,52%	87,38%	95,99%	93,94%	0,05618

Ao avaliar os resultados obtidos por todos os modelos com a função de perda *Dice Categorical Focal Loss*, percebe-se que as abordagens que incorporam o módulo de pirâmide de *pooling*, como PPM-UNet, Deeplabv3+ Padrão, PPM-Deeplabv3+ e CPP-UNet, demonstraram, de maneira geral, um desempenho comparável e, em alguns casos, superior em relação às abordagens U-Net Padrão e Deeplabv3+ sem o módulo ASPP. Indicando que a incorporação desse módulo às redes convolucionais desempenha um papel significativo na melhoria da segmentação de estruturas relacionadas ao câncer renal. Por outro lado, ao analisar os resultados obtidos com a função de perda *Weighted Categorical Crossentropy Loss*, observa-se que os valores médios do coeficiente de Dice foram semelhantes entre todos os modelos, com destaque para Deeplabv3+ Padrão, U-Net e CPP-UNet, que apresentaram os melhores índices.

É importante ressaltar que a abordagem CPP-UNet, desenvolvida neste trabalho, obteve o maior percentual médio de coeficiente Dice dentre todas as abordagens avaliadas, considerando ambas as funções de perda. O modelo atingiu valores médios de Dice e Jaccard de 95,32%, 94,61%, 94,76% e 93,94%, respectivamente, para todas as classes analisadas, evidenciando sua eficiência na segmentação dos rins, cistos e tumores renais.

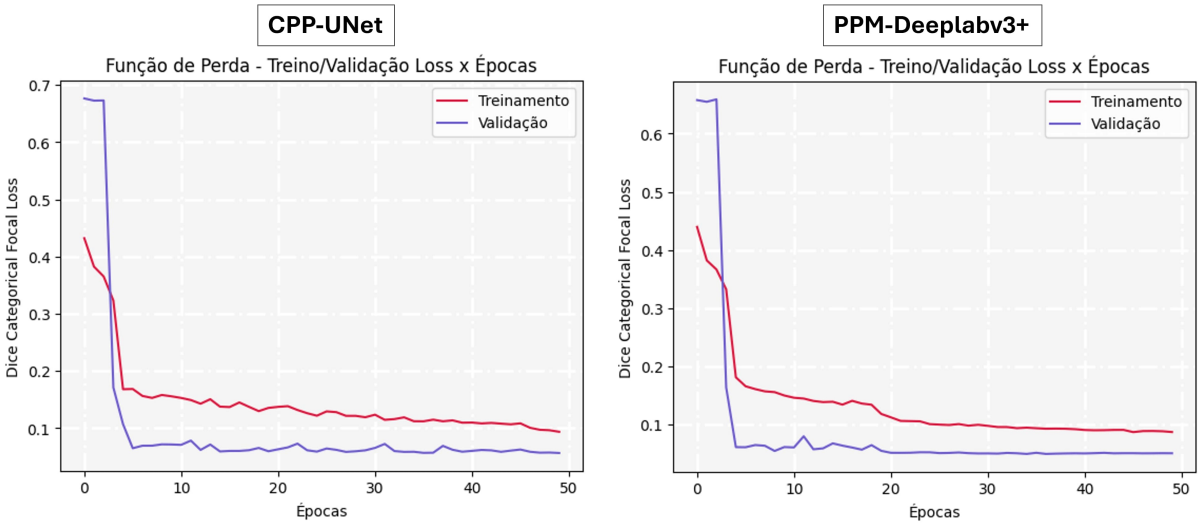
Em uma análise mais aprofundada da segmentação por classe, observa-se que os modelos propostos, PPM-Deeplabv3+ e CPP-UNet, obtiveram um desempenho superior em comparação com as demais abordagens na segmentação de rins, tumores e cistos renais ao utilizar a primeira função de perda avaliada. Detalhadamente, esses modelos alcançaram índices de Dice de 95,17% e 95,23% para a segmentação dos rins, 89,45% e 91,44% para a segmentação dos tumores e 94,67% e 96,25% para a segmentação dos cistos renais, demonstrando sua eficácia na identificação dessas estruturas.

Especificamente, o modelo CPP-UNet obteve o maior percentual de coeficiente

Dice na segmentação de tumores, superando o segundo melhor modelo aplicado, PPM-Deeplabv3+, por aproximadamente $2,0$ pontos percentuais. Além disso, esse desempenho foi $3,46$ pontos percentuais superior ao modelo U-Net padrão, que não utiliza o mecanismo de pirâmide de pooling. Esses resultados indicam que a aplicação desse mecanismo pode contribuir significativamente para a extração de informações mais relevantes no contexto do câncer renal. Importante destacar que para a segmentação de cistos renais as abordagens PPM-Deeplabv3+ foi superior a todos os modelos aplicados quando utilizando a função de perda *Dice Categorical Focal Loss*. Estes resultados superaram em cerca de $1,58$ pontos percentuais a segunda melhor abordagem, CPP-UNet aplicado à segunda função de perda. Quando comparado ao modelo padrão da Deeplabv3+, os resultados foram quase $2,0$ pontos percentuais superiores.

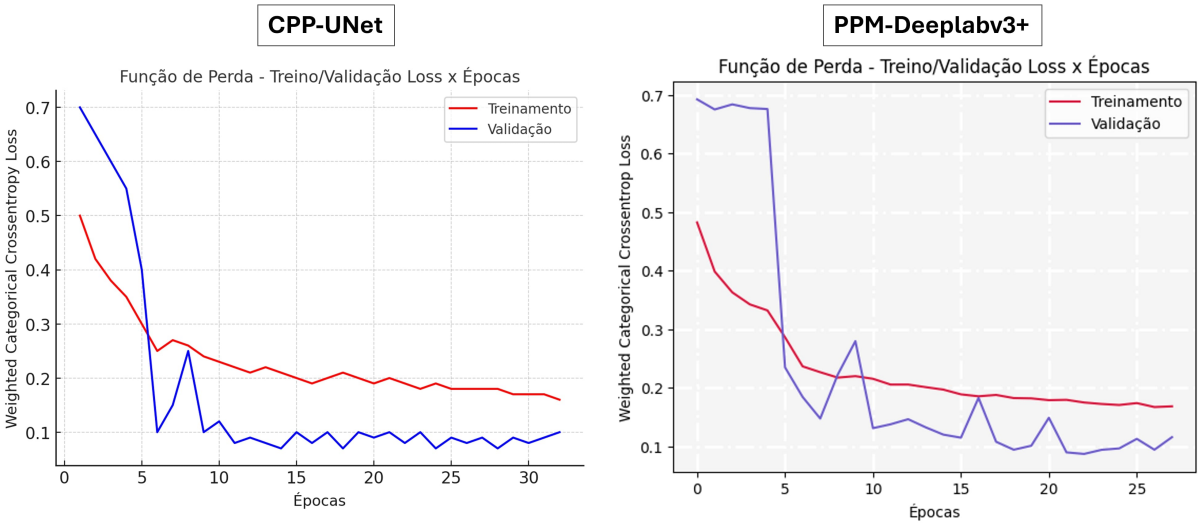
As curvas de perda auxiliam na análise do processo de otimização, fornecendo insights sobre o comportamento do modelo durante o treinamento. Elas permitem avaliar a convergência do aprendizado, identificar possíveis problemas, como overfitting ou underfitting, e ajustar hiperparâmetros para melhorar o desempenho do modelo. Nesse contexto, as Figuras 34 e 35 apresentam as curvas de perda associadas às funções *Dice Categorical Focal Loss* e *Weighted Categorical Crossentropy Loss*, utilizadas no treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base de dados KiTS21.

Figura 34 – Curvas da função de perda *Dice Categorical Focal Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS23.



Fonte: Desenvolvido pelo Autor

Figura 35 – Curvas da função de perda *Weighted Categorical Crossentropy Loss* referente ao treinamento dos modelos CPP-UNet e PPM-Deeplabv3+ na base KiTS23.



Fonte: Desenvolvido pelo Autor

A análise das curvas de perda permite avaliar a eficácia das funções de perda na penalização de erros durante o treinamento, além de auxiliar na escolha da abordagem mais adequada para garantir um aprendizado robusto e equilibrado entre as classes analisadas. Os resultados obtidos com os modelos de segmentação CPP-UNet e PPM-Deeplabv3+ utilizando funções de perda com ponderação indicam que a *Dice Categorical Focal Loss* apresentou o melhor desempenho em ambas as bases de imagens aplicadas nesta tese. Especificamente, quando essa função de perda priorizou as classes de tumores e cistos no modelo CPP-UNet, os resultados foram superiores às demais abordagens avaliadas nos conjuntos de dados analisados. Esse achado confirma a hipótese levantada na questão de pesquisa **Q4** (Seção 1.2), que investiga o impacto do uso de funções de perda com ponderação por classe como estratégia para mitigar o desbalanceamento presente nas bases do desafio KiTS.

Além disso, os resultados obtidos evidenciam que o desempenho das redes convolucionais tradicionais, como U-Net e Deeplab, pode ser significativamente aprimorado com a integração de módulos como ASPP e PPM. Essa combinação fortalece o extrator de características, tornando-o mais eficiente na segmentação de rins, cistos e tumores renais, especialmente ao lidar com as complexas variações morfológicas dessas estruturas em exames de tomografia computadorizada. Dessa forma, a adoção desses módulos não apenas melhora a precisão na segmentação das patologias renais, como também pode contribuir para um diagnóstico mais assertivo e para o aprimoramento das estratégias de tratamento do câncer renal.

5.3.3 Estudos de Caso

Nesta subseção, são apresentados exemplos de exames do conjunto de teste referente e base KiTS23 nos quais os modelos PPM-Deeplabv3+ e CPP-UNet foram aplicados para a segmentação de rins, cistos e tumores renais. O objetivo dessa análise é fornecer uma avaliação qualitativa dos resultados obtidos, destacando o desempenho dos modelos desenvolvidos nesta pesquisa. Além disso, conforme realizado na Seção 5.2.3, o método de explicabilidade Grad-CAM (SELVARAJU et al., 2017) foi utilizado para analisar o comportamento dos modelos PPM-Deeplabv3+ e CPP-UNet na segmentação das estruturas renais e suas patologias associadas, seguindo a mesma configuração descrita no estudo de caso anterior. Por fim, são discutidos aspectos relevantes sobre o desempenho dos modelos, destacando tanto os casos de segmentação bem-sucedida quanto as limitações identificadas dos modelos em efetuar a segmentação dos rins, cistos e tumores renais.

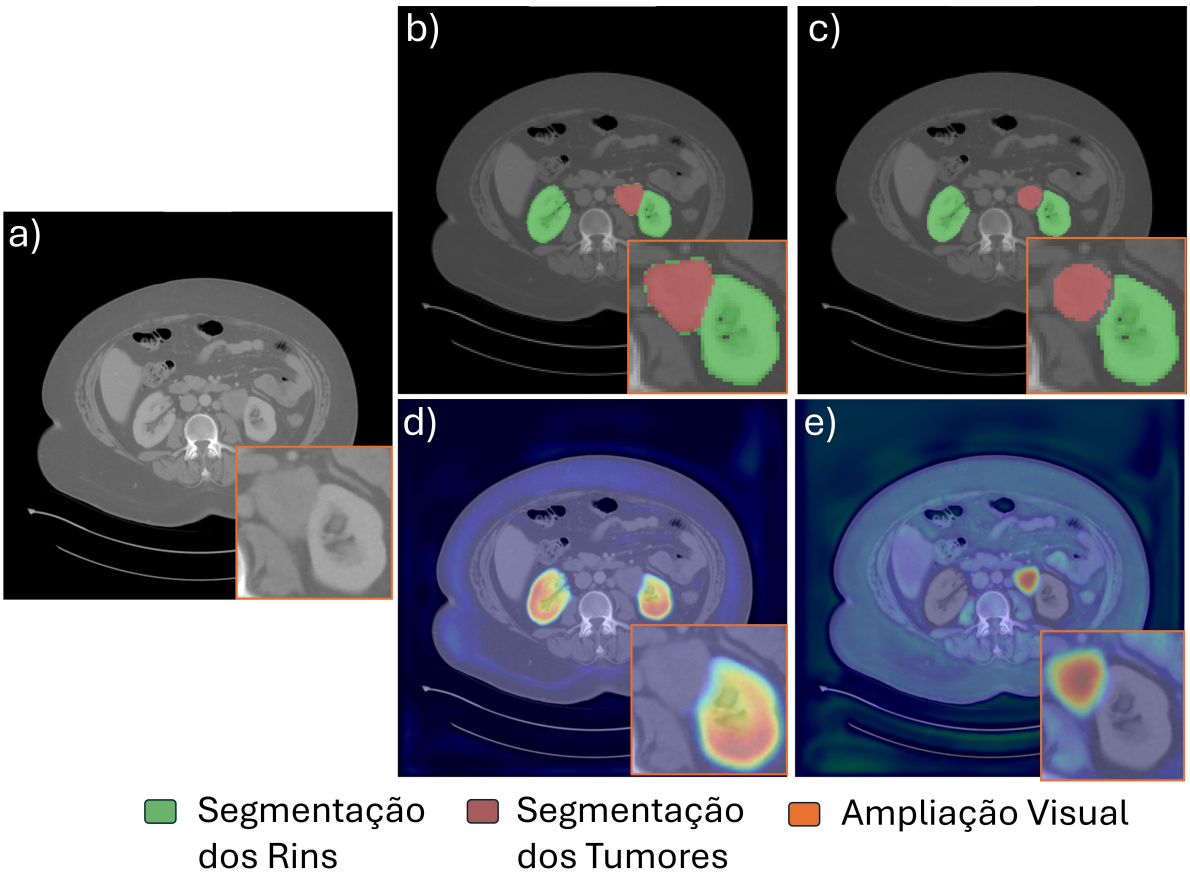
5.3.3.1 Casos de Sucesso

Exame — 00406 — O exame de tomografia computadorizada do paciente 00406 apresenta a presença de rins e tumores renais, conforme indicado pela segmentação realizada pelo especialista. A Figura 36 ilustra uma fatia da imagem do exame, destacando as regiões segmentadas, onde os rins são representados em verde e os tumores renais em vermelho. Adicionalmente, são apresentados os resultados da segmentação destas estruturas utilizando o modelo PPM-DeepLabv3+. Verifica-se que o modelo obteve êxito na segmentação dos rins e do tumor renal no exame 00406. Além disso, destaca-se sua capacidade de identificar as características do tumor renal associado ao rim esquerdo do paciente, mesmo diante da alta similaridade do tumor com outras estruturas abdominais, como pode ser observado na fatia da imagem analisada.

Complementando a análise do exame 00406, a técnica de explicabilidade Grad-CAM foi empregada ao modelo PPM-Deeplabv3+, mantendo a mesma configuração apresentada no estudo de caso anterior na Seção 5.2.3. Os resultados, ilustrados na Figura 36, evidenciam que as regiões correspondentes aos rins e ao tumor renal foram corretamente ativadas pelo modelo. Destaca-se que as áreas de maior intensidade de ativação estão concentradas tanto nos rins quanto no tumor renal associado ao rim direito do paciente. Essas observações indicam que o modelo conseguiu capturar informações discriminantes entre rins e tumores renais neste exame de tomografia computadorizada.

Exame — 00152 — Por outro lado, observa-se que a segmentação dos rins e tumores no Exame 00152 foi realizada corretamente pelo modelo CPP-UNet, conforme ilustrado na Figura 37. Nota-se que a segmentação do rim esquerdo do paciente produzida pelo modelo assemelha-se com a segmentação do médico especialista. Destaca-se ainda, que as regiões delimitadoras do rim direito e o tumor associado

Figura 36 – Segmentação correta dos rins e tumor renal presentes no Exame 00406 a partir do modelo PPM-DeepLabv3+. (a) Fatia original do exame 00406; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumor renal à imagem de tomografia computadorizada, respectivamente.



Fonte: Desenvolvido pelo Autor

possuem elevado grau de semelhança, conforme observado na fatia do Exame 00152. No entanto, mesmo que essa diferenciação não seja claramente perceptível, o modelo desenvolvido nesta tese obteve sucesso em delimitar, as áreas correspondentes aos rins e ao tumor renal, apresentando uma segmentação próxima à realizada pelo especialista.

Analisando os mapas de ativação gerados pelo modelo CPP-UNet para o Exame 00152, ilustrado na Figura 37, verifica-se que as regiões correspondentes aos rins e ao tumor renal foram corretamente destacadas pelo codificador do modelo. Além disso, observa-se que a maior concentração de ativações está localizada internamente às estruturas renais e ao tumor, o que indica que o modelo foi capaz de extrair características representativas dessas classes, mesmo diante das semelhanças visuais entre elas.

Dessa forma, observa-se que o modelo CPP-UNet, ao incorporar a combinação das pirâmides de pooling ASPP e PPM, explora de maneira eficiente o potencial

das convoluções dilatadas e tradicionais. Essa abordagem possibilita a extração de informações tanto no contexto local quanto global da imagem, tornando o modelo mais sensível a pequenas variações estruturais nos rins, especialmente em cistos e tumores. A eficácia dessa estratégia é evidenciada pelos resultados obtidos no Exame 00152, ilustrado na Figura 37, onde a diferenciação entre as regiões do rim direito e do tumor associado não é claramente perceptível na fatia original do exame, conforme apresentado na Figura 37(a). Esse achado confirma a hipótese levantada na questão de pesquisa **Q3** (Seção 1.2). Essa questão investiga o impacto do uso de métodos que consideram múltiplos aspectos contextuais da imagem (globais e locais) na extração de características sensíveis às sutis variações entre patologias renais e demais órgãos. Esse aspecto é crítico para uma segmentação mais assertiva, contribuindo para a precisão dos resultados. Além disso, a melhoria na segmentação pode auxiliar o médico especialista na identificação de regiões suspeitas com potencial presença de cistos e tumores renais, contribuindo para um diagnóstico mais preciso do câncer renal.

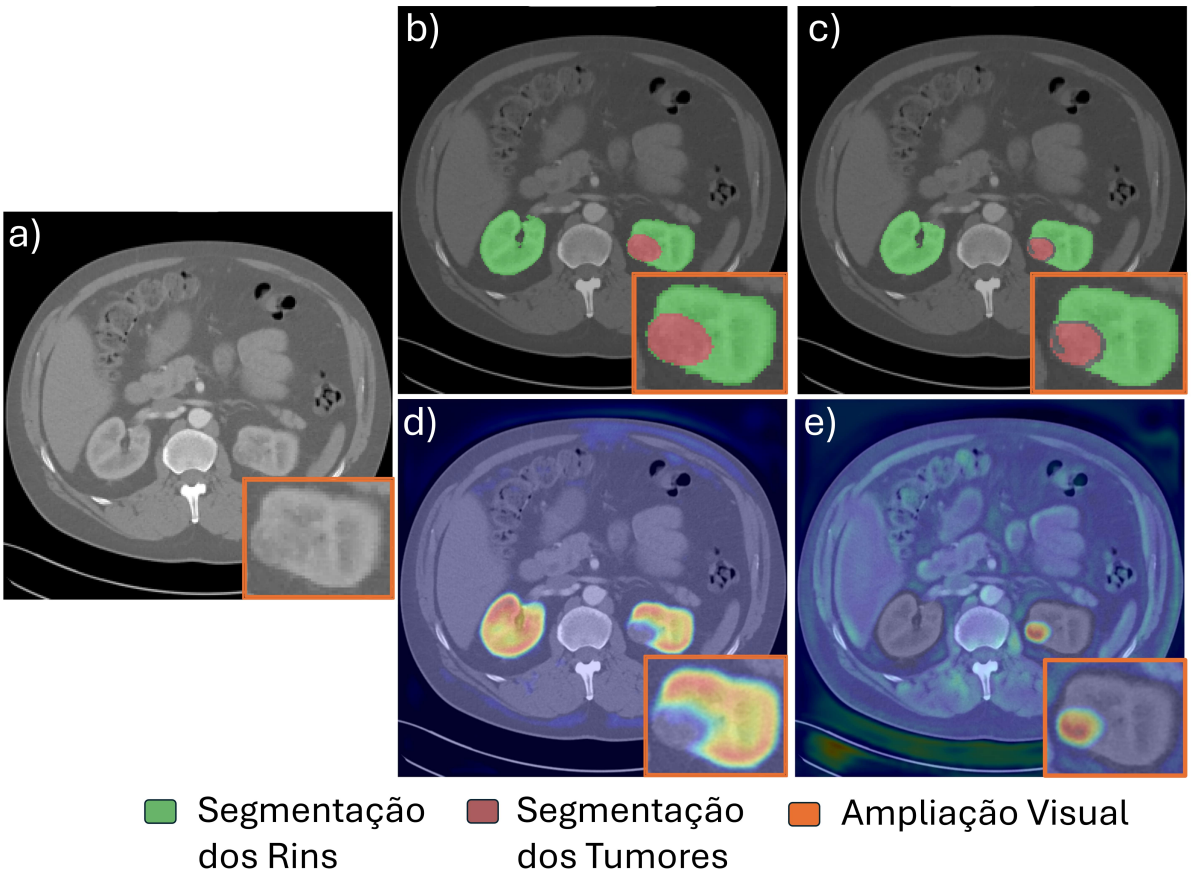
Exame - 00175 - Os cistos renais constituem uma condição patológica que, em determinados casos, pode estar associada a um risco de malignidade. Fatores como a presença de septações, calcificações, espessamento da parede ou conteúdo interno heterogêneo podem contribuir para essa suspeita (MUGLIA; WESTPHALEN, 2014). Assim, a segmentação dos cistos renais pode auxiliar na detecção precoce dessas lesões, desempenhando um papel fundamental na prevenção e no diagnóstico do câncer renal. Desta forma, a Figura 38 ilustra um exemplo de segmentação de rim e cisto renal efetuada pelo modelo CPP-UNet para o exame 00175. Avaliando os resultados obtidos, observamos que, em comparação com a segmentação realizada pelo especialista (*Ground Truth*), o modelo conseguiu identificar uma parte significativa dos contornos das estruturas correspondentes ao rim e ao cisto renal. Acredita-se que a delimitação completa do cisto não foi possível devido à sua conexão com estruturas internas do rim, como ilustrado na ampliação visual da Figura 38 (a).

As características relevantes consideradas pelo modelo CPP-UNet para o exame 00175 podem ser visualizadas na Figura 38. O mapa de ativação gerado para os rins e cistos demonstram que o modelo foi capaz de identificar corretamente a região onde está presente estas estruturas renais. Além disso, ao examinar a região do cisto renal, observa-se que as ativações de maior intensidade estão corretamente concentradas na área correspondente. No entanto, para uma segmentação completa do cisto, o modelo requer aprimoramento, visando melhorar seu desempenho nesses casos específicos.

5.3.3.2 Casos de Erro

Exame - 00475 - O exame 00475 apresenta a marcação dos rins esquerdo e direito, bem como do tumor renal associado ao rim esquerdo do paciente, conforme

Figura 37 – Segmentação correta dos rins e tumor renal aplicando o modelo CPP-UNet no Exame 00152. (a) Fatia original do exame 00152; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumor renal à imagem de tomografia computadorizada, respectivamente.

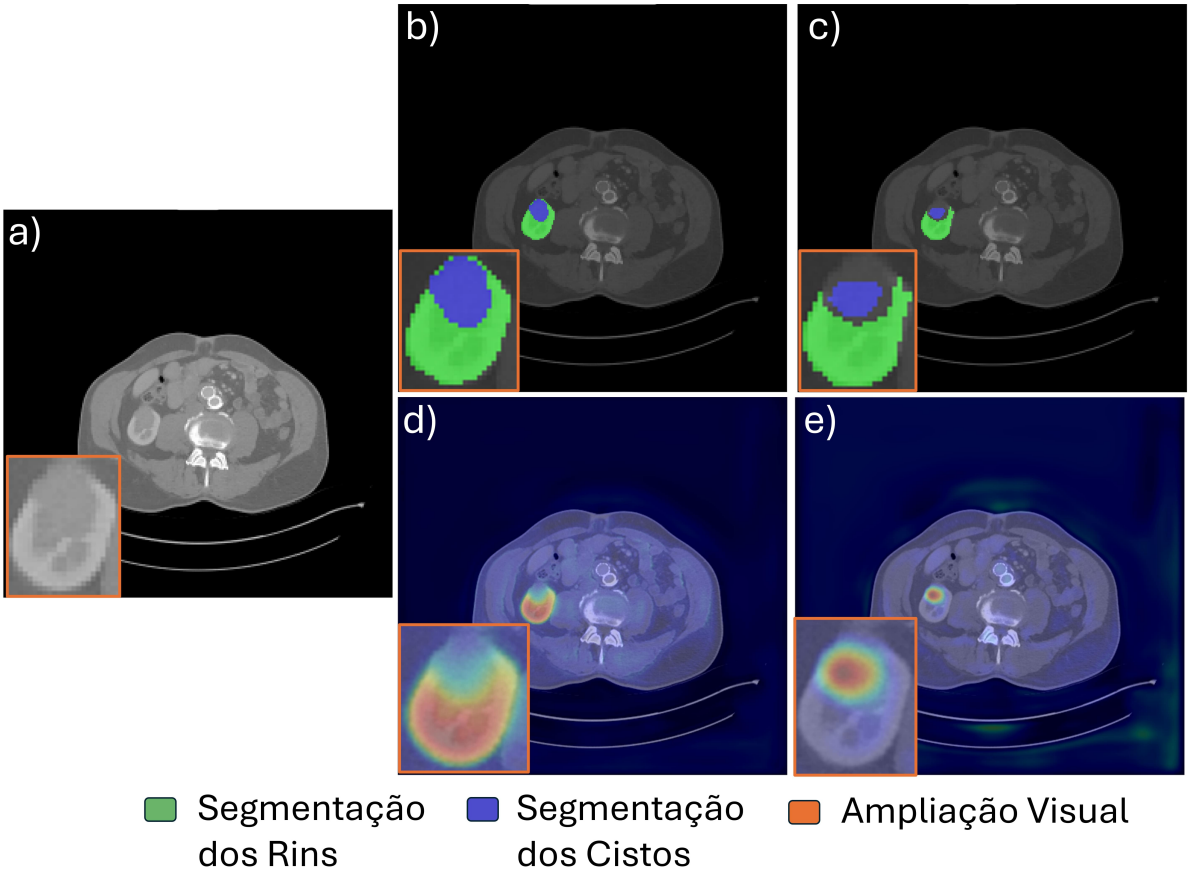


Fonte: Desenvolvido pelo Autor

ilustrado na Figura 39. Ao aplicar o modelo PPM-Deeplabv3+ a esse exame, observou-se que a segmentação dos rins foi realizada de forma aproximada em relação à marcação disponibilizada. No entanto, em relação à identificação do tumor renal, o modelo não conseguiu segmentá-lo adequadamente. Acredita-se que a semelhança em termos de forma e tamanho entre o rim esquerdo e o tumor renal tenha dificultado a diferenciação entre essas estruturas, comprometendo a segmentação. Esse aspecto torna-se ainda mais evidente na fatia do exame 00475, especialmente na ampliação visual da região correspondente ao rim esquerdo do paciente.

O mapa de ativação gerado pelo modelo para o exame 00475 evidencia essa limitação, uma vez que não foram identificadas ativações na região correspondente ao tumor renal, conforme demonstrado na Figura 39. Esse comportamento sugere que o modelo não reconheceu o tumor como uma estrutura distinta, o que levou à sua

Figura 38 – Exemplo de segmentação bem-sucedida do rim e do cisto renal utilizando o modelo CPP-UNet para o Exame com identificador 00175. (a) Fatia original do exame 00175; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação do rim esquerdo e cisto renal à imagem de tomografia computadorizada, respectivamente.

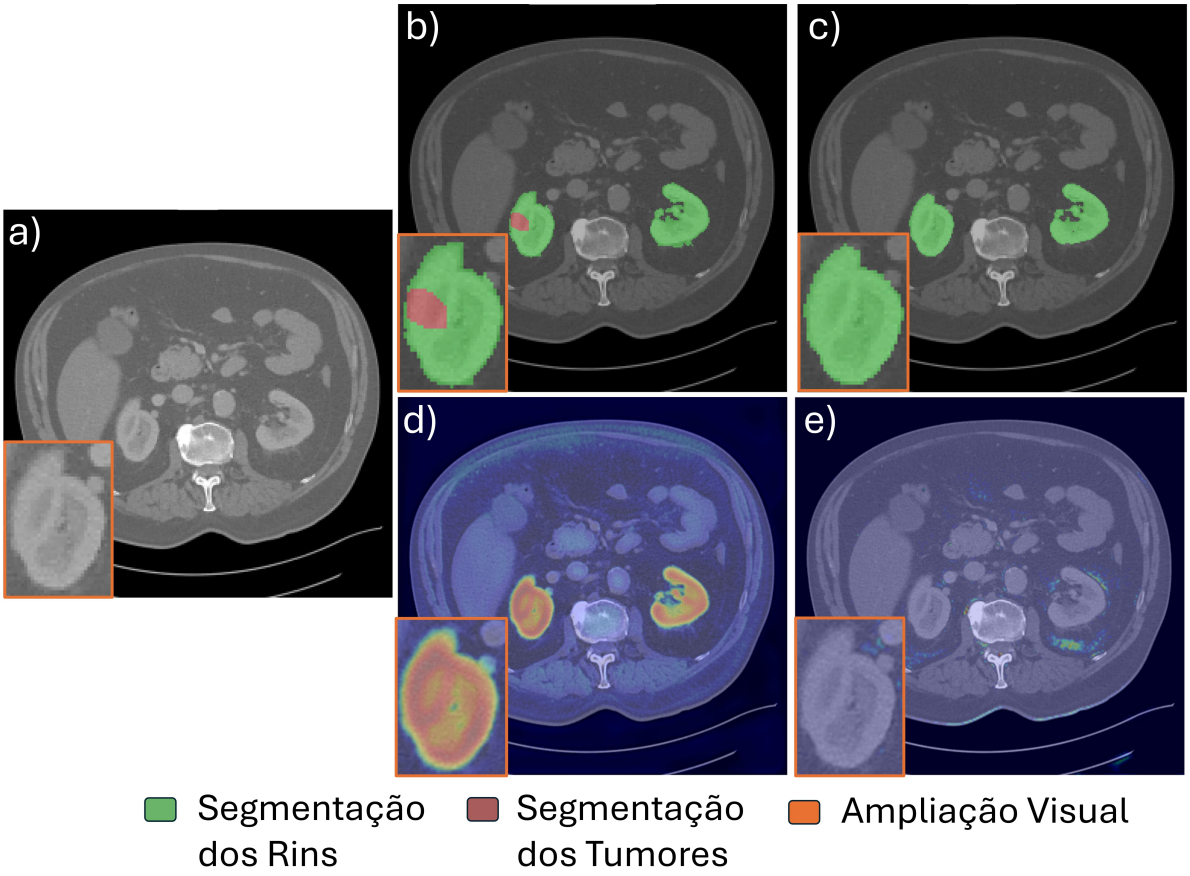


Fonte: Desenvolvido pelo Autor

incorreta inclusão na segmentação do rim esquerdo. Além disso, a ausência de respostas significativas na área do tumor indica que as características extraídas pelo modelo não foram suficientes para diferenciá-lo do tecido renal adjacente.

Exame - 00060 - Exemplo de segmentação dos rins e tumores renais efetuada pelo CPP-UNet para o Exame 00060 pode ser visualizado na Figura 40. Observa-se que, os rins, direito e esquerdo, contidos no exame foram corretamente segmentados, quando comparado a marcação do especialista. De forma semelhante, o tumor renal presente no rim esquerdo foi segmentado em quase sua totalidade pelo CPP-UNet. Por outro lado, ao analisar a segmentação do cisto renal associado ao rim direito do paciente, o modelo não foi capaz de realizar uma segmentação precisa desta patologia. Houve erros na identificação e delimitação dessa estrutura, resultando em uma segmentação incorreta.

Figura 39 – Exemplo de segmentação incorreta do tumor renal utilizando o modelo PPM-Deeplabv3+ para o Exame com identificador 00475. (a) Fatia original do exame 00475; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins e tumor renal à imagem de tomografia computadorizada, respectivamente.



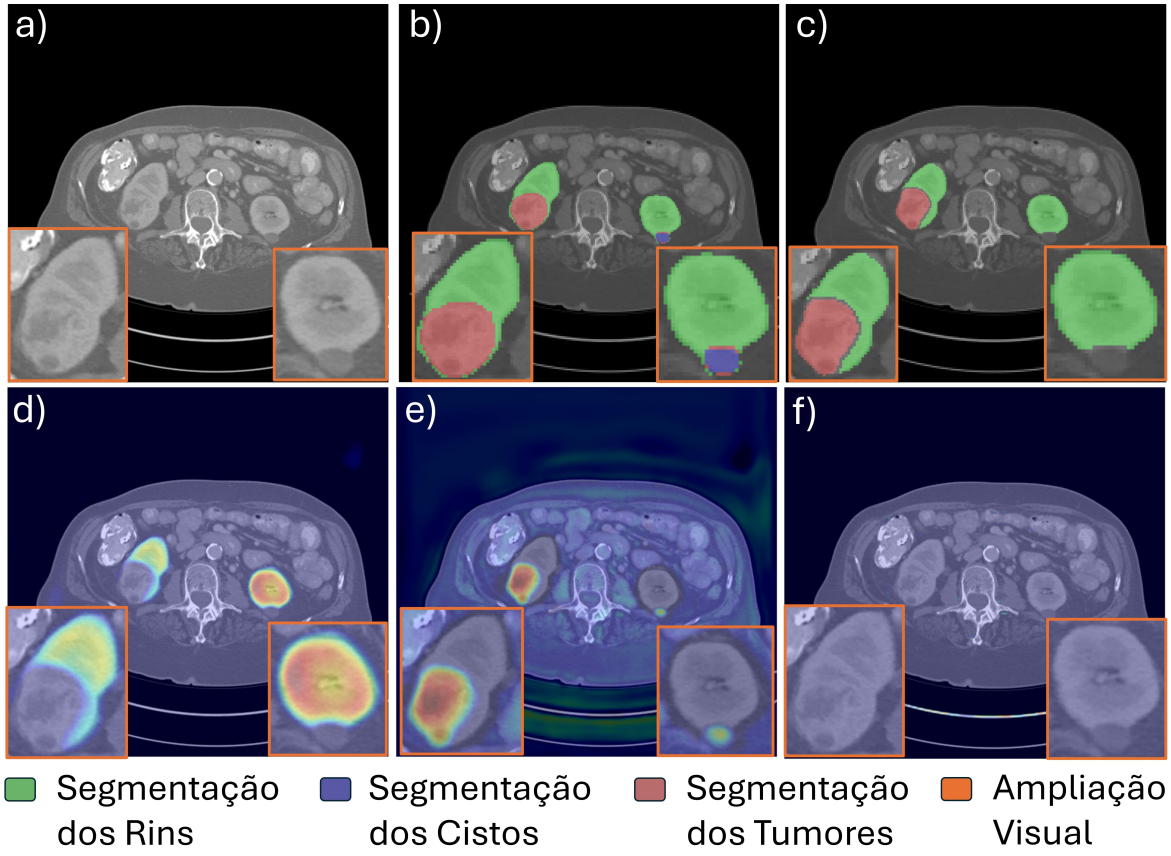
Fonte: Desenvolvido pelo Autor

Embora o modelo tenha delimitado corretamente os rins, não foi capaz de identificar a presença do cisto renal localizado na parte inferior do rim esquerdo. Esse comportamento torna-se mais evidente ao analisarmos a Figura 40, que apresenta o mapa de ativação gerado para cada uma das classes. Nota-se que as regiões de maior ativação estão concentradas nos rins e no tumor renal, enquanto a área correspondente ao cisto não apresenta ativações. Isso indica que o modelo não conseguiu extrair características relevantes para a segmentação do cisto presente no exame.

5.4 Comparação com Trabalhos Relacionados

Nesta seção, é realizada uma análise comparativa dos resultados obtidos pela metodologia desenvolvida nesta tese em relação às pesquisas relacionadas

Figura 40 – Segmentação incorreta do cisto renal aplicando o modelo CPP-UNet para o Exame 00060. (a) Fatia original do exame 00060; (b) Segmentação efetuada pelo especialista (*GroundTruth*); (c) Segmentação produzida pelo modelo;(d) e (e) Sobreposições dos mapas de ativação dos rins, cisto e tumor renal à imagem de tomografia computadorizada, respectivamente.



Fonte: Desenvolvido pelo Autor

apresentadas no Capítulo 2. Comparar abordagens correlatas é um desafio, uma vez que frequentemente há uma escassez de informações detalhadas sobre aspectos como configuração dos treinamentos, divisão dos conjuntos de dados e parametrização dos métodos. Dessa forma, mesmo quando a mesma base de imagens é utilizada, a comparação só pode ser feita de maneira aproximada.

Para tanto, as subseções a seguir apresentam uma comparação dos modelos PPM-Deeplabv3+ e CPP-UNet com as pesquisas que representam o estado da arte na segmentação de rins, cistos e tumores renais nas bases KiTS21 e KiTS23. Adicionalmente, foi aplicada a técnica de Classes de Avaliação Hierárquica (CAH), conforme descrito na Seção 4.1 para agrupar as classes em macroclasses em ”Massas Renais”, ”Rins e Massas” e ”Tumor” de modo a estabelecer critério de comparação entre as abordagens da literatura.

5.4.1 Abordagens aplicadas na base KiTS21

A Tabela 7, tem-se uma visão resumida dos resultados encontrados nas pesquisas relacionadas a base KiTS21, a quantidade de pacientes envolvidos e as métricas de desempenho empregadas em cada método. Desta forma, observa-se que os modelos propostos são promissores, uma vez que, em comparação com os obras que ocupam os melhores resultados, alcançaram resultados superiores, em torno de 91,75% e 91,38% de Dice para a segmentação de Massas Renais e 87,65% e 87,84% para segmentação de tumores renais. Esses resultados são particularmente relevantes, pois a segmentação de cistos e tumores renais desempenha um papel crucial no diagnóstico do câncer renal. Uma segmentação correta pode contribuir significativamente para a análise, o diagnóstico e o prognóstico da doença, auxiliando na identificação precoce e no planejamento terapêutico adequado.

Tabela 7 – Comparação dos resultados da metologia proposta com os trabalhos relacionados.

Base KiTS21						
Trabalho	**Método**	**Pacientes**	**Rins e Massas**	**Massas Renais**	**Tumor**	**Composição**
Zheng et al. (2024)	**MSALDS-UNet**	300	**96,10%**	**84,40%**	**87,40%**	**89,30%**
Sun et al. (2023)	2.5D MFFAU-Net	300	97,30%	88,70%	87,30%	91,10%
George (2022)	3D U-Net	300	97,48%	88,13%	87,10%	90,90%
Zhao, Chen e Wang (2022)	nnUNet	300	97,52%	88,51%	86,93%	90,99%
Golts et al. (2022)	3D U-Net	300	97,15%	87,90%	86,38%	90,48%
Wu e Liu (2022)	DenseU-Net com CA	300	96,93%	87,60%	85,09%	89,87%
Yang et al. (2022)	Residual 3D U-Net com ASPP	210	97,30%	87,40%	82,20%	88,97%
Zang et al. (2022)	3D U-Net	300	97,00%	85,10%	81,90%	88,00%
Sun et al. (2022)	DenseU-Net com AG	300	97,10%	88,30%	81,50%	88,97%
Li et al. (2022)	Residual e 3D U-Net	300	-	-	81,00%	-
Adam et al. (2022)	3D U-ResNet	300	95,10%	79,80%	78,80%	84,57%
Chen e Liu (2022)	ResSENormUnet e DenseTransUnet	300	-	-	77,79%	-
Heo (2022)	3D U-Net	300	93,69%	78.83%	75,03%	84,36%
Shen et al. (2022)	COTRNet	300	92,30%	55,30%	50,60%	66,07%
Modelos Propostos	**PPM-Deeplabv3+**	300	**91,25%**	**91,75%**	**87,65%**	**90,22%**
	CPP-UNet	300	**92,32%**	**91,38%**	**87,84%**	**90,51%**

Avaliando especificamente a tarefa de segmentação de massas renais (tumores e cistos), a metodologia proposta nesta tese demonstrou desempenho superior em comparação ao estado da arte no conjunto de dados KiTS21. Destaca-se que as taxas de Dice alcançadas foram de 91,38% e 91,75% ao utilizar os modelos PPM-Deeplabv3+ e CPP-UNet, respectivamente. Esses resultados representam um incremento de aproximadamente 7,35% em relação ao estudo conduzido por Zhao, Chen e Wang (2022), vencedor do desafio KiTS21 em 2021.

Além disso, na segmentação específica de tumores renais, o modelo CPP-UNet obteve o melhor desempenho, atingindo 87,84% de Dice, superando em 0,44% a abordagem considerada referência na literatura. Esses achados evidenciam o potencial da metodologia desenvolvida e reforçam sua relevância para a segmentação de estruturas renais em imagens de tomografia computadorizada.

Dando continuidade à comparação, observa-se que a maioria das abordagens

presentes na literatura (LI et al., 2022; GOLTS et al., 2022; GEORGE, 2022; HEO, 2022; ZANG et al., 2022; ADAM et al., 2022; YANG et al., 2022) utiliza a arquitetura U-Net como base para o desenvolvimento de suas metodologias. Essas abordagens incorporam adaptações específicas visando aumentar a robustez da rede e aprimorar sua precisão na segmentação de rins, cistos e tumores renais. Ao comparar os resultados obtidos pelos modelos propostos com os das abordagens mencionadas, verifica-se que os métodos desenvolvidos apresentam um desempenho promissor, especialmente na segmentação de massas renais e tumores. Esse desempenho reforça a eficácia da combinação de pirâmides de *pooling* com redes convolucionais, uma vez que essa abordagem permite a extração de características mais sensíveis às sutis variações entre cistos e tumores renais. Mediante a observação dos resultados obtem-se a resposta positiva para questão de pesquisa **Q2** (Seção 1.2) que trata do uso combinado de pirâmides de*pooling* aplincando diferentes tipos de convolução.

5.4.2 Abordagens aplicadas na base KiTS23

A Tabela 8 apresenta uma análise comparativa entre os modelos desenvolvidos nesta tese e as principais pesquisas voltadas à segmentação dos rins e suas respectivas patologias na base KiTS23. Nessa tabela, são destacadas informações relevantes, incluindo a técnica empregada, o número de pacientes utilizados nos experimentos e os resultados obtidos em termos do coeficiente de Dice. Além disso, são apresentados os melhores resultados obtido pelos modelos PPM-Deeplabv3+ e CPP-UNet.

Comparando os resultados obtidos pelos modelos propostos com os demais estudos apresentados na Tabela 8, observa-se que, de maneira geral, os métodos desenvolvidos demonstram um desempenho promissor. Os modelos alcançaram a maior média (Composição) considerando a segmentação de rins, massas renais e tumores, com métricas superiores a 91% no coeficiente de Dice. Esses resultados indicam que a abordagem proposta consegue segmentar de maneira equilibrada e eficiente todas as estruturas analisadas, incluindo rins, cistos e tumores renais.

Continuando análise comparativa, observa-se diretamente que ambos os modelos propostos obtiveram métricas superiores para segmentação de massas renais e tumor, com Dice de 92,85% e 89,45% para PPM-UNet e 93,06% 91,44% para CPP-UNet, respectivamente. Em detalhes, efetuando uma comparação direta com algumas pesquisas, Salahuddin et al. (2023) utilizam uma adaptação da rede U-Net, denominada nnU-Net, projetada para ajustar automaticamente sua arquitetura e estratégias de treinamento com base nas características dos dados de entrada. Este trabalho obteve um Dice de 86,50% para massas renais e 83,50% para tumor, considerando 599 pacientes. Os modelos desenvolvidos nesta pesquisa não aplicam mecanismo de ajuste automático de hiperparâmetros. Ainda assim, obtiveram resultados superiores quando observamos a

segmentação de rins e massas e tumores renais.

Tabela 8 – Comparação dos resultados da metodologia proposta com os trabalhos relacionados.

Base KiTS23						
Trabalho	**Método**	**Pacientes**	**Rins e Massas**	**Massas Renais**	**Tumor**	**Composição**
Chen e Zhang (2023)	**2.5D ResUnet**	489	**89,44%**	**85,85%**	**85,91%**	**87,07%**
Uhm et al. (2023)	3D U-Net e Pós-processamento	599	97,90%	82,60%	85,70%	88,73%
Salahuddin et al. (2023)	nnUNet com CLR	599	94,00%	86,50%	83,50%	88,00%
Myronenko et al. (2023)	SegResNet e SwinUNETR	599	95,60%	79,20%	75,80%	83,53%
Vispi (2023)	EfficientNet-B5 + UNet	599	97,71%	81,39%	73,81%	84,30%
Stoica, Breaban e Barbu (2023)	3D U-Net e Data Augmentation	599	94,70%	76,00%	71,30%	80,67%
Qian et al. (2023)	nnU-Net e Swin Transformer	489	-	-	68,70%	80,10%
Hu e Peng (2023)	Residual e 3D U-Net	599	93,30%	74,40%	67,90%	78,53%
Li, Peng e Zhang (2023)	nnU-Net + Residual Attention	599	93,60%	-	67,00%	-
Wang et al. (2023)	nnU-Net	599	86,60%	54,50%	49,00%	63,37%
Modelos Propostos	**PPM-Deeplabv3+**	489	**93,62%**	**92,85%**	**89,45%**	**91,97%**
	CPP-UNet	489	**93,78%**	**93,06%**	**91,44%**	**92,76%**

Comparando com outras pesquisas, Chen e Zhang (2023) utilizam uma abordagem híbrida de rede residual 2.5D integrada à rede convolucional U-Net. Redes que empregam blocos residuais se destacam pela capacidade de aprender representações de alto nível, graças à adição de conexões de salto entre camadas. No entanto, uma desvantagem significativa dessas redes é o aumento no número de parâmetros, o que pode acarretar um custo computacional maior e aumentar o risco de *overfitting*. Além disso, a similaridade entre as estruturas abdominais presentes nas imagens de tomografia computadorizada pode impactar negativamente as redes residuais, que enfrentam dificuldades na preservação de detalhes, comprometendo a identificação de características sutis, como os contornos de órgãos ou tumores. Este trabalho é considerado o padrão ouro para segmentação de tumores renais, alcançando um Dice de 85,91% ao considerar 599 pacientes da base KiTS23. Apesar de não utilizarem blocos residuais, os modelos desenvolvidos nesta pesquisa obtiveram resultados superiores na segmentação de tumores renais, atingindo 89,45% e 91,44% de Dice, respectivamente. É relevante destacar que o modelo CPP-UNet apresentou um desempenho 5,53% superior ao deste estudo, que já obteve os melhores resultados no estado da arte.

A maioria dos estudos utiliza variações da rede U-Net (Uhm et al. (2023), Salahuddin et al. (2023), Myronenko et al. (2023), Vispi (2023), Stoica, Breaban e Barbu (2023), Qian et al. (2023), Hu e Peng (2023), Li, Peng e Zhang (2023), Wang et al. (2023)), alcançando resultados promissores na segmentação de rins, massas renais e tumores. No entanto, uma limitação das abordagens baseadas na arquitetura U-Net é o aglomerado de parâmetros decorrente das conexões de salto e das camadas adicionais no caminho de expansão, o que pode aumentar o risco de overfitting e prejudicar a capacidade de generalização do modelo. Para superar essa limitação, os modelos propostos nesta pesquisa adotam uma abordagem híbrida, combinando convoluções dilatadas e tradicionais em sua estrutura, visando otimizar a extração de características sem comprometer a eficiência e robustez da segmentação. Adicionalmente, a adaptação

dos módulos PPM e ASPP, como codificadores das redes Deeplab e U-Net, permite a integração de características tanto no contexto local quanto global das imagens, tornando os métodos mais eficazes na segmentação de rins, cistos e tumores renais.

5.5 Limitações da Pesquisa e Metodologia

Embora os resultados obtenham indicações de que a aplicação dos módulos de pirâmide de pooling pode melhorar a precisão da segmentação de lesões renais quando associados às redes convolucionais, é importante ressaltar que ainda existem limitações, restrições e desafios a serem superados. Dessa forma, é fundamental destacar as limitações da metodologia adotada, especialmente para uma análise crítica dos resultados.

Apesar do sucesso para segmentação utilizando os modelos PPM-Deeplabv3+ e CPP-UNet, cada arquitetura possui suas limitações, que incluem o tamanho dos filtros de convolução, taxas de dilatação, bem como a quantidade de níveis da pirâmide de *pooling*. Tanto o PPM-Deeplabv3+ quanto o CPP-UNet apresentam, em suas arquiteturas, uma dependência significativa da escolha adequada da taxa de dilatação aplicada em cada nível do bloco ASPP, assim como da definição do tamanho do filtro de convolução utilizado no bloco PPM. Taxas de dilatação maiores, como 12 e 18, relacionam características situadas em regiões distantes na imagem, enquanto taxas menores, como 1 e 6 abrangem informações mais próximas. De maneira análoga, o tamanho dos filtros de convolução aplicados no bloco PPM influencia diretamente o tipo de característica extraída: filtros maiores capturam informações mais globais, permitindo uma análise estrutural abrangente da imagem, enquanto filtros menores destacam detalhes locais, fundamentais para a diferenciação de estruturas anatômicas semelhantes, como cistos e tumores renais.

Na segmentação de imagens médicas, as diferenças sutis entre cistos e tumores renais podem ser atenuadas ao longo do treinamento, dependendo da configuração dos filtros de convolução e da taxa de dilatação aplicada. Dessa forma, a escolha adequada desses parâmetros é fundamental para otimizar a extração e o aprendizado das características relevantes, garantindo a diferenciação entre as estruturas abdominais e as patologias renais. Embora as otimizações não tenham sido efetuadas, é importante destacar que o uso e adaptações dessas técnicas em suas definições padrão demonstram o potencial dos módulos de pirâmide de *pooling*.

Além disso, outro fator limitante é a quantidade de blocos PPM empregados em sequência nas arquiteturas desenvolvidas. A acumulação desses blocos resulta em um aumento no número de operações de convolução para cada nível da pirâmide, tornando o modelo mais complexo e potencialmente suscetível ao sobreajuste (*overfitting*). Portanto, a definição do número ideal de blocos PPM deve ser cuidadosamente analisada, buscando

um equilíbrio entre a precisão da segmentação e a manutenção do desempenho do modelo. Esses modelos são altamente sensíveis à configuração desses parâmetros, de modo que ajustes inadequados podem comprometer a fase de codificação, resultando na extração de características incorretas para a segmentação de rins, cistos e tumores renais.

Outra limitação diz respeito à parametrização da técnica de preprocessamento aplicada, o CLAHE. Esta técnica de correção utiliza o limite de contraste e tamanho da grade de divisão. O primeiro parâmetro controla o limite de corte do histograma, evitando a amplificação excessiva do contraste e a criação de artefatos na imagem. Por outro lado, o segundo parâmetro ajusta a aplicação da técnica efetuando a divisão da imagem em grades. Portanto, a escolha incorreta destes pode resultar na produção de imagens com ruídos excessivos, dificultando o treinamento para segmentação dos rins, cistos e tumores. Assim, estes parâmetros devem ser considerados para otimizar a captura e o aprendizado das características relevantes para segmentação de estruturas renais.

As funções de perda também representam um aspecto passível de aprimoramento na metodologia. A adoção de uma estratégia de ponderação entre as classes pode levar ao superajuste do modelo para uma classe específica, resultando em overfitting. Dessa forma, a otimização dos valores de ponderação para cada classe (fundo, rins, cistos e tumores) pode tornar possivel alcançar um equilíbrio adequado entre as categorias analisadas. Esse ajuste visa garantir um treinamento mais generalizado e discriminativo, promovendo um melhor desempenho do modelo em todas as classes.

O processo de otimização de hiperparâmetros, por meio de ferramentas como o *GridSearch*, pode contribuir para a identificação da configuração ideal para ambos os modelos. Além disso, a adaptação das arquiteturas para a integração dos módulos de pirâmides pode ser realizada em diferentes estágios, incluindo o codificador, o decodificador ou as *skip connections*. No entanto, essa flexibilidade pode aumentar significativamente a complexidade do modelo, tornando sua implementação inviável para determinadas aplicações.

5.6 Considerações Finais

Este capítulo apresentou em detalhes os experimentos para a segmentação de rins e patologias relacionadas. Estudos de casos também foram apresentados, observam-se casos de sucesso e falhas. Também foi feito um comparativo com os trabalhos relacionados, para analisar a relevância da pesquisa desenvolvida.

No próximo capítulo, são apresentadas as conclusões sobre a pesquisa desenvolvida. As principais contribuições também são apontadas, bem como sugestões para trabalhos futuros.

6 Conclusão e Trabalhos Futuros

O desenvolvimento de metodologias e arquiteturas para a segmentação semântica de órgãos e lesões tumorais em exames de tomografia computadorizada é uma tarefa complexa e desafiadora, especialmente quando se trata de exames de TC da região abdominal, que englobam uma variedade significativa de órgãos, como rins, baço, bexiga, entre outros. A segmentação dos rins e das patologias associadas é de extrema importância para permitir aos especialistas realizar o diagnóstico precoce do câncer renal. Neste trabalho, foi desenvolvida uma metodologia que busca a combinação das arquiteturas PPM e ASPP associadas ao codificador das redes Deeplabv3+ e CPP-UNet, a fim de desenvolver novos modelos de segmentação capazes de realizar análises em várias resoluções, aplicado especificamente no contexto da segmentação renal.

A metodologia desenvolvida consistiu em quatro etapas principais. Na primeira etapa, foram obtidos os conjuntos de dados das edições de 2021 e 2023 do desafio KiTS para a realização das análises. Em seguida, na segunda etapa, foram aplicadas técnicas de processamento de imagens, incluindo realce e redimensionamento, com o objetivo de evidenciar as estruturas anatômicas abdominais, especificamente rins, cistos e tumores renais. Na terceira etapa, os modelos de segmentação são empregados para a identificação dessas estruturas, utilizando, especificamente, os modelos PPM-Deeplabv3+ e CPP-UNet. Por fim, na quarta etapa, a precisão das segmentações obtidas é avaliada por meio de métricas quantitativas de desempenho.

Os resultados experimentais demonstraram um desempenho promissor na segmentação de rins e patologias renais. Para a segmentação de tumores e cistos renais no conjunto de dados KiTS21, a aplicação do modelo CPP-UNet resultou em coeficientes de Dice de 87,84% e 94,91%, e índices de Jaccard de 87,28% e 94,91%, respectivamente. Por outro lado, ao utilizar o modelo PPM-UNet na mesma tarefa, foram obtidos coeficientes de Dice de 87,65% e 95,84%, e índices de Jaccard de 86,94% e 95,74%, respectivamente. Esses resultados evidenciaram a eficácia dos modelos na segmentação das estruturas renais e suas patologias associadas. Além disso, ao serem aplicados ao conjunto de dados KiTS23, os modelos de segmentação demonstraram desempenho expressivo, especialmente na identificação e segmentação de estruturas associadas a cistos e tumores renais. Especificamente, na utilização do modelo PPM-Deeplabv3+, foram obtidos coeficientes de Dice de 91,44% e 94,67%, e índices de Jaccard de 90,77% e 94,67%, respectivamente. Semelhantemente, ao empregar o modelo CPP-UNet, os resultados alcançados foram de 91,44% e 94,67% para o coeficiente Dice e de 90,77% e 94,67% para o índice Jaccard, evidenciando a eficácia dos modelos na segmentação dessas estruturas patológicas.

Esses resultados demonstram a eficácia da metodologia desenvolvida na identificação e segmentação de rins, cistos e tumores renais em imagens de tomografia computadorizada. Além disso, destacam o potencial da utilização de módulos de pirâmide de *pooling*, como PPM e ASPP, em combinação com redes convolucionais U-Net e DeepLab, para a construção de uma arquitetura robusta, capaz de extrair características locais e globais em múltiplas resoluções, otimizando a segmentação no contexto do câncer renal. Desta forma, essa representação aprimorada é essencial para tarefas que demandam a compreensão de detalhes em diferentes escalas, como a segmentação de estruturas complexas em imagens médicas.

Além disso, a ampliação dos filtros aplicados nos módulos de pirâmide permite que a operação de convolução identifique e relacione características de estruturas espacialmente mais distantes na imagem, facilitando a análise de informações contextuais e detalhadas. Importante destacar que esta habilidade de capturar informações em múltiplas escalas torna a convolução *atrous* uma poderosa ferramenta em aplicações para segmentação semântica e análise de imagens médicas. Especificamente, a inserção dos blocos PPM visa aprimorar a capacidade das redes DeepLab e U-Net em obter informações contextuais em várias escalas fornecendo a capacidade de capturar características em contextos próximos à região de interesse na imagem, inicialmente, por meio do uso da convolução tradicional. Isso permite uma análise mais precisa e robusta de pequenas regiões da imagem, tornando as redes mais eficazes na tarefa de segmentação do câncer renal.

Essa abordagem contribui para a precisão do diagnóstico, permitindo uma identificação precoce das lesões renais e possibilitando um prognóstico mais assertivo. A segmentação automatizada dessas estruturas facilita a avaliação clínica e o planejamento terapêutico, auxiliando os especialistas na tomada de decisões mais eficazes. Como consequência, a detecção precoce e a melhor caracterização das lesões podem aumentar as chances de sucesso no tratamento, proporcionando melhores perspectivas de recuperação e qualidade de vida aos pacientes com câncer renal.

Como pesquisas futuras, recomenda-se investigar a adaptação e a integração dos módulos de pirâmide de pooling em arquiteturas que incorporam mecanismos de autoatenção (self-attention), como o TransUNET (CHEN et al., 2024), o U-Net Transformer (PETIT et al., 2021) e o Vision Mamba UNet (RUAN; LI; XIANG, 2024). Acredita-se que a capacidade dessas técnicas de autoatenção em modelar dependências e relações nas sequências de entrada possibilitará a extração de características mais contextualizadas, aprimorando a segmentação das estruturas renais, incluindo rins, tumores e cistos. Além disso, a investigação da aplicação da metodologia a outras estruturas anatômicas e patologias de interesse voltadas a tomografia computadorizada.

Referências

ADAM, J.; AGETHEN, N.; BOHNSACK, R.; FINZEL, R.; GÜNNEMANN, T.; PHILIPP, L.; PLUTAT, M.; RINK, M.; XUE, T.; THIELKE, F. et al. Extraction of kidney anatomy based on a 3d u-resnet with overlap-tile strategy. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 13–21.

ALMEIDA, L. B. C1. 2 multilayer perceptrons. **Handbook of Neural Computation C**, Elsevier Amsterdam, The Netherlands, v. 1, 1997.

American Cancer Society. **Survival Rates for Kidney Cancer**. 2024. Acesso em: 15 de Fevereiro 2025. Disponível em: <https://www.cancer.org/cancer/types/kidney-cancer/detection-diagnosis-staging/survival-rates.html>.

American Cancer Society. **What Causes Kidney Cancer? - Risk Factors for Kidney Cancer**. 2024. Acesso em: 15 de Fevereiro 2025. Disponível em: <https://www.cancer.org/cancer/types/kidney-cancer/causes-risks-prevention/risk-factors.html">.

American Cancer Society. **What Causes Prostate Cancer? - Genetic (DNA) Changes**. 2024. Acesso em: 15 de Fevereiro 2025. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/causes-risks-prevention/what-causes.html>.

American Cancer Society. **What Is Kidney Cancer?** 2024. Acesso em: 15 de Fevereiro 2025. Disponível em: <https://www.cancer.org/cancer/types/kidney-cancer/about/what-is-kidney-cancer.html>.

ARAP, M. A.; COELHO, R. F. **Tumorações Renais**. 2023. Disponível em: <https://www.medicinanet.com.br/conteudos/revisoes/3981/tumoracoes_renais.htm>. Acesso em: 22 de Fevereiro 2025.

BANSAL, R. C. Overview and literature survey of artificial neural networks applications to power systems (1992-2004). **Journal of Institution of Engineers (India)–Electrical Engineering**, v. 86, n. 1, p. 282–296, 2006.

BEBIS, G.; GEORGIOPOULOS, M. Feed-forward neural networks. **IEEE Potentials**, v. 13, n. 4, p. 27–31, 1994.

BILIC, P.; CHRIST, P.; LI, H. B.; VORONTSOV, E.; BEN-COHEN, A.; KAISSIS, G.; SZESKIN, A.; JACOBS, C.; MAMANI, G. E. H.; CHARTRAND, G. et al. The liver tumor segmentation benchmark (lits). **Medical Image Analysis**, Elsevier, v. 84, p. 102680, 2023.

BROWNLEE, J. **Better deep learning: train faster, reduce overfitting, and make better predictions**. 151 Calle de San Francisco: Machine Learning Mastery, 2018.

CHEN, C.; ZHANG, R. An ensemble of 2.5 d resunet based models for segmentation of kidney and masses. In: **International Challenge on Kidney and Kidney Tumor Segmentation**. Vancouver, Canada: Springer, 2023. p. 47–53.

CHEN, J.; MEI, J.; LI, X.; LU, Y.; YU, Q.; WEI, Q.; LUO, X.; XIE, Y.; ADELI, E.; WANG, Y. et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. **Medical Image Analysis**, Elsevier, v. 97, p. 103280, 2024.

CHEN, L.-C.; PAPANDREOU, G.; KOKKINOS, I.; MURPHY, K.; YUILLE, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 40, n. 4, p. 834–848, 2017.

CHEN, L.-C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: **Proceedings of the European conference on computer vision (ECCV)**. Munich, Germany: Springer, 2018. p. 801–818.

CHEN, M.; CHALLITA, U.; SAAD, W.; YIN, C.; DEBBAH, M. Artificial neural networks-based machine learning for wireless networks: A tutorial. **IEEE Communications Surveys & Tutorials**, IEEE, v. 21, n. 4, p. 3039–3071, 2019.

CHEN, Z.; LIU, H. 2.5 d cascaded semantic segmentation for kidney tumor cyst. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 28–34.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Honolulu, HI, USA: IEEE, 2017. p. 1251–1258.

CIERNIAK, R. **X-ray computed tomography in biomedical engineering**. London: Springer Science & Business Media, 2011.

CIRILLO, L.; INNOCENTI, S.; BECHERUCCI, F. Global epidemiology of kidney cancer. **Nephrology Dialysis Transplantation**, Oxford University Press, v. 39, n. 6, p. 920–928, 2024.

DANTAS, H. A. d. O. **Morfologia Sistema Urinário**. 2023. Disponível em: <https://ulbra-to.br/morfologia/2011/08/17/Sistema-Urinario>. Acesso em: 18 de Abril 2023.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: IEEE. **2009 IEEE conference on computer vision and pattern recognition**. Miami, FL, 2009. p. 248–255.

DING, Y.; LIN, L.; WANG, L.; ZHANG, M.; LI, D. Digging into the multi-scale structure for a more refined depth map and 3d reconstruction. **Neural Computing and Applications**, Springer, v. 32, p. 11217–11228, 2020.

EELBODE, T.; BERTELS, J.; BERMAN, M.; VANDERMEULEN, D.; MAES, F.; BISSCHOPS, R.; BLASCHKO, M. B. Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. **IEEE Transactions on Medical Imaging**, IEEE, v. 39, n. 11, p. 3679–3690, 2020.

Fairview Health Services. **Leading The Way To Better Healthcare**. 2024. Acesso em: 12 de Dezembro 2024. Disponível em: <https://www.fairview.org/>.

FERNANDES, B. J. T. Redes neurais com extração implícita de características para reconhecimento de padrões visuais. Universidade Federal de Pernambuco, 2013.

FILHO, O. M.; NETO, H. V. **Processamento digital de imagens**. Rio de Janeiro, Brasil: Brasport, 1999.

GEORGE, Y. A coarse-to-fine 3d u-net network for semantic segmentation of kidney ct scans. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 137–142.

GERVEN, M. van; BOHTE, S. M. Artificial neural networks as models of neural information processing. 2017.

GOFRIT, O. N.; YUTKIN, V.; ZORN, K. C.; DUVDEVANI, M.; LANDAU, E. H.; HIDAS, G.; PODE, D. The growth rate of "clinically significant" renal cancer. **Springerplus**, Springer, v. 4, p. 1–5, 2015.

GOLTS, A.; KHAPUN, D.; SHATS, D.; SHOSHAN, Y.; GILBOA-SOLOMON, F. An ensemble of 3d u-net based models for segmentation of kidney and masses in ct scans. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 103–115.

GONZALEZ, R. C.; WOODS, R. E. Processamento digital de imagem. **Pearson, ISBN-10: 8576054019**, v. 10, p. 11–27, 2010.

GROUP, D. F. W. et al. **NIfTI:—neuroimaging informatics technology initiative**. 2023. Disponível em: <https://nifti.nimh.nih.gov/>. Acesso em: 25 de março 2023.

GULLI, A.; PAL, S. **Deep learning with Keras**. Birmingham, UK: Packt Publishing Ltd, 2017.

GUYTON, A.; J.E., H. **Fundamentos de Fisiologia**. Rio de Janeiro, Brasil: Elsevier Brasil, 2017.

HAFEMANN, L. G. An analysis of deep neural networks for texture classification. Dissertação (Mestrado), Universidade Federal do Paraná, Curitiba, 2014.

HAYKIN, S. **Redes neurais: princípios e prática**. Porto Alegre, RS, Brasil: Bookman Editora, 2001.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. San Juan, PR, USA: IEEE, 2016. p. 770–778.

HELLER, N.; ISENSEE, F.; MAIER-HEIN, K. H.; HOU, X.; XIE, C.; LI, F.; NAN, Y.; MU, G.; LIN, Z.; HAN, M. et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. **Medical image analysis**, Elsevier, v. 67, p. 101821, 2021.

HELLER, N.; ISENSEE, F.; MAIER-HEIN, K. H.; HOU, X.; XIE, C.; LI, F.; NAN, Y.; MU, G.; LIN, Z.; HAN, M. et al. **The kits23 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes.** 2023. Disponível em: <https://kits-challenge.org/kits23/>. Acesso em: 25 de março 2023.

HELLER, N.; ISENSEE, F.; TROFIMOVA, D.; TEJPAUL, R.; ZHAO, Z.; CHEN, H.; WANG, L.; GOLTS, A.; KHAPUN, D.; SHATS, D.; SHOSHAN, Y.; GILBOA-SOLOMON, F.; GEORGE, Y.; YANG, X.; ZHANG, J.; ZHANG, J.; XIA, Y.; WU, M.; LIU, Z.; WALCZAK, E.; MCSWEENEY, S.; VASDEV, R.; HORNUNG, C.; SOLAIMAN, R.; SCHOEPHOERSTER, J.; ABERNATHY, B.; WU, D.; ABDULKADIR, S.; BYUN, B.; SPRIGGS, J.; STRUYK, G.; AUSTIN, A.; SIMPSON, B.; HAGSTROM, M.; VIRNIG, S.; FRENCH, J.; VENKATESH, N.; CHAN, S.; MOORE, K.; JACOBSEN, A.; AUSTIN, S.; AUSTIN, M.; REGMI, S.; PAPANIKOLOPOULOS, N.; WEIGHT, C. **The KiTS21 Challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase CT**. 2023.

HELLER, N.; SATHIANATHEN, N.; KALAPARA, A.; WALCZAK, E.; MOORE, K.; KALUZNIAK, H.; ROSENBERG, J.; BLAKE, P.; RENGEL, Z.; OESTREICH, M. et al. **The 2021 Kidney and Kidney Tumor Segmentation Challenge**. 2021. Disponível em: <https://kits21.kits-challenge.org/>.

HEO, J. Automatic segmentation in abdominal ct imaging for the kits21 challenge. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 98–102.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, 2012.

Hospital Sírio-Libanês. **Câncer de Rim: Saiba Mais**. 2024. Acesso em: 15 de Fevereiro 2025. Disponível em: <https://www.hospitalsiriolibanes.org.br/blog/urologia/cancer-de-rim-saiba-mais>.

HU, X.; PENG, Y. Gsca-net: A global spatial channel attention network for kidney, tumor and cyst segmentation. In: **International Challenge on Kidney and Kidney Tumor Segmentation**. Vancouver, BC, Canada: Springer, 2023. p. 67–76.

Instituto Nacional do Câncer. **INCA estima 704 mil casos de câncer por ano no Brasil até 2025**. 2022. Acesso em: 5 de Março 2025. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/noticias/2022/inca-estima-704-mil-casos-de-cancer-por-ano-no-brasil-ate-2025>.

Instituto Nacional do Câncer. **Incidência, mortalidade e morbidade hospitalar por câncer**. 2022. Acesso em: 5 de Março 2025. Disponível em: <https://www.inca.gov.br/numeros-de-cancer/morbidade-hospitalar>.

ISENSEE, F.; JAEGER, P. F.; FULL, S. M.; VOLLMUTH, P.; MAIER-HEIN, K. H. nnu-net: a self-adapting framework for u-net-based medical image segmentation. **Nature Methods**, Nature Publishing Group, v. 18, n. 2, p. 203–211, 2021.

Johns Hopkins Medicine. **Computed Tomography (CT) Scan**. 2023. Disponível em: <https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/computed-tomography-ct-scan>. Acesso em: 25 de março 2023.

KANG, K.; WANG, X. Fully convolutional neural networks for crowd segmentation. **arXiv preprint arXiv:1411.4464**, 2014.

KOVÁCS, Z. L. **Redes neurais artificiais**. São Paulo, Brasil: Editora Livraria da Fisica, 2002.

KULKARNI, N. M.; FUNG, A.; KAMBADAKONE, A. R.; YEH, B. M. Computed tomography techniques, protocols, advancements, and future directions in liver diseases. **Magnetic Resonance Imaging Clinics**, Elsevier, v. 29, n. 3, p. 305–320, 2021.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, Ieee, v. 86, n. 11, p. 2278–2324, 1998.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: IEEE. **Proceedings of 2010 IEEE international symposium on circuits and systems**. Paris, France: IEEE, 2010. p. 253–256.

LI, D.; CHEN, Z.; HASSAN, H.; XIE, W.; HUANG, B. A cascaded 3d segmentation model for renal enhanced ct images. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 123–128.

LI, J.; CHENG, J.-h.; SHI, J.-y.; HUANG, F. Brief introduction of back propagation (bp) neural network algorithm and its improvement. In: SPRINGER. **Advances in Computer Science and Information Engineering: Volume 2**. Zhengzhou, China, 2012. p. 553–558.

LI, Z.; PENG, Y.; ZHANG, Z. Two-stage segmentation framework with parallel decoders for the kidney and kidney tumor segmentation. In: **International Challenge on Kidney and Kidney Tumor Segmentation**. Vancouver, BC, Canada: Springer, 2023. p. 83–92.

LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLáR, P. Focal loss for dense object detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 42, n. 2, p. 318–327, 2020.

LIVINGSTONE, D. J. **Artificial neural networks: methods and applications**. United Kingdom, UK: Springer, 2008. v. 458.

M Health Fairview. **Healthcare founded in academics**. 2024. Acesso em: 12 de Dezembro 2024. Disponível em: <https://mhealthfairview.org/About-Us>.

MAAS, A. L.; HANNUN, A. Y.; NG, A. Y. et al. Rectifier nonlinearities improve neural network acoustic models. In: ATLANTA, GEORGIA, USA. **Proc. icml**. California, USA, 2013. v. 30, n. 1, p. 3.

MCCULLOCH, W. S. A logical calculus of the ideas imminent in nervous activity. **Bull. Math. Biophys.**, v. 5, p. 115–133, 1943.

MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: IEEE. **2016 fourth international conference on 3D vision (3DV)**. California, USA: IEEE, 2016. p. 565–571.

MORKOS, J.; RODRIGUEZ, K. A. P.; ZHOU, A.; KOLARICH, A. R.; FRANGAKIS, C.; RODRIGUEZ, R.; GEORGIADES, C. S. Percutaneous cryoablation for stage 1 renal cell carcinoma: outcomes from a 10-year prospective study and comparison with matched

cohorts from the national cancer database. **Radiology**, Radiological Society of North America, v. 296, n. 2, p. 452–459, 2020.

MOSTAFA, A. H.; ABDEL-GALIL, H.; BELAL, M. Ensemble model-based weighted categorical cross-entropy loss for facial expression recognition. In: IEEE. **2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)**. Cairo, Egypt: IEEE, 2021. p. 165–171.

MU, Y.; NGUYEN, T.; HAWICKHORST, B.; WRIGGERS, W.; SUN, J.; HE, J. The combined focal loss and dice loss function improves the segmentation of beta-sheets in medium-resolution cryo-electron-microscopy density maps. **Bioinformatics Advances**, Oxford University Press, v. 4, n. 1, p. vbae169, 2024.

MUGLIA, V. F.; WESTPHALEN, A. C. Classificação de bosniak para cistos renais complexos: histórico e análise crítica. **Radiologia Brasileira**, SciELO Brasil, v. 47, n. 6, p. 368–373, 2014.

MYRONENKO, A.; YANG, D.; HE, Y.; XU, D. Automated 3d segmentation of kidneys and tumors in miccai kits 2023 challenge. **arXiv preprint arXiv:2310.04110**, 2023. DOI:<https://doi.org/10.48550/arXiv.2310.04110>.

NefroClínicas. **Quais são as Funções dos Rins?** 2023. Disponível em: <https://nefroclinicas.com.br/quais-sao-as-funcoes-dos-rins/>. Acesso em: 22 de Fevereiro 2025.

NETTER, F. H.; FRANK, M. Atlas de anatomia humana. 2ª edição. **New Jersey: Artmed**, p. 401–16, 2000.

NIH - Nacional Institute of Biomedical Imaging and Bioengenineering. **Computed Tomography (CT)**. 2025. Disponível em: <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct>. Acesso em: 17 de Janeiro 2025.

PADALA, S. A.; BARSOUK, A.; THANDRA, K. C.; SAGINALA, K.; MOHAMMED, A.; VAKITI, A.; RAWLA, P.; BARSOUK, A. Epidemiology of renal cell carcinoma. **World journal of oncology**, Elmer Press, v. 11, n. 3, p. 79, 2020.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de imagens digitais: princípios, algoritmos e aplicações**. São Paulo, Brasil: Cengage Learning, 2008.

PELBERG, R.; PELBERG, R. Basic principles in computed tomography (ct). **Cardiac CT angiography manual**, Springer, p. 19–58, 2015.

PETIT, O.; THOME, N.; RAMBOUR, C.; THEMYR, L.; COLLINS, T.; SOLER, L. U-net transformer: Self and cross attention for medical image segmentation. In: SPRINGER. **Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, September 27, 2021, Proceedings 12**. Strasbourg, France: Springer, 2021. p. 267–276.

PETRIK, V.; APOK, V.; BRITTON, J. A.; BELL, B. A.; PAPADOPOULOS, M. C. Godfrey hounsfield and the dawn of computed tomography. **Neurosurgery**, Oxford University Press, v. 58, n. 4, p. 780–787, 2006.

PIZER, S. M.; AMBURN, E. P.; AUSTIN, J. D.; CROMARTIE, R.; GESELOWITZ, A.; GREER, T.; ROMENY, B. ter H.; ZIMMERMAN, J. B.; ZUIDERVELD, K. Adaptive histogram equalization and its variations. **Computer vision, graphics, and image processing**, Elsevier, v. 39, n. 3, p. 355–368, 1987.

QIAN, L.; LUO, L.; ZHONG, Y.; ZHONG, D. A hybrid network based on nnu-net and swin transformer for kidney tumor segmentation. In: **International Challenge on Kidney and Kidney Tumor Segmentation**. Vancouver, BC, Canada: Springer, 2023. p. 30–39.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. Munich, Germany, 2015. p. 234–241.

RUAN, J.; LI, J.; XIANG, S. Vm-unet: Vision mamba unet for medical image segmentation. **arXiv preprint arXiv:2402.02491**, 2024.

SALAHUDDIN, Z.; KUANG, S.; LAMBIN, P.; WOODRUFF, H. C. Leveraging uncertainty estimation for segmentation of kidney, kidney tumor and kidney cysts. In: **International Challenge on Kidney and Kidney Tumor Segmentation**. Vancouver, BC, Canada: Springer, 2023. p. 40–46.

SanarMED. **Função Renal: revisão anatômica e as funções renais**. 2023. Disponível em: <https://www.sanarmed.com/funcao-renal>. Acesso em: 22 de Fevereiro 2025.

SBN - Sociedade Brasileira de Nefrologia. **Compreendendo os rins**. 2023. Disponível em: <https://www.sbn.org.br/o-que-e-nefrologia/compreendendo-os-rins/>. Acesso em: 22 de Fevereiro 2025.

SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Proceedings of the IEEE international conference on computer vision**. Venice, Italy: IEEE, 2017. p. 618–626.

SHEN, Z.; YANG, H.; ZHANG, Z.; ZHENG, S. Automated kidney tumor segmentation with convolution and transformer network. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 1–12.

SILVERMAN, S. G.; KHORASANI, R. K.; O'CONNOR, S. D.; COCHON, L. R. Renal cancer at unenhanced ct: imaging features, detection rates, and outcomes. **Abdominal Radiology**, v. 43, n. 7, 2018.

SILVERTHORN, D. U. **Fisiologia humana: uma abordagem integrada**. Rio Grande do Sul, Brasil: Artmed editora, 2010.

SPROUFFSKE, K.; AKTIPIS, C. A.; RADICH, J. P.; CARROLL, M.; NEDELCU, A. M.; MALEY, C. C. An evolutionary explanation for the presence of cancer nonstem cells in neoplasms. **Evolutionary applications**, Wiley Online Library, v. 6, n. 1, p. 92–101, 2013.

STOICA, G.; BREABAN, M.; BARBU, V. Analyzing domain shift when using additional data for the miccai kits23 challenge. **arXiv preprint arXiv:2309.02001**, 2023. DOI:<https://doi.org/10.48550/arXiv.2309.02001>.

SUN, P.; MO, Z.; HU, F.; SONG, X.; MO, T.; YU, B.; ZHANG, Y.; CHEN, Z. Segmentation of kidney mass using agdenseu-net 2.5 d model. **Computers in Biology and Medicine**, Elsevier, Amsterdam, Holanda, v. 150, p. 106223, 2022.

SUN, P.; MO, Z.; HU, F.; SONG, X.; MO, T.; YU, B.; ZHANG, Y.; CHEN, Z. 2.5 d mffau-net: a convolutional neural network for kidney segmentation. **BMC Medical Informatics and Decision Making**, Springer, v. 23, n. 1, p. 92, 2023.

SUNANDINI, G.; SIVANPILLAI, R.; SOWMYA, V.; VARIYAR, V. S. Significance of atrous spatial pyramid pooling (aspp) in deeplabv3+ for water body segmentation. In: IEEE. **2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)**. Delhi-NCR, India, 2023. p. 744–749.

TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. **BMC medical imaging**, BioMed Central, v. 15, n. 1, p. 1–28, 2015.

TING, J.-A.; VIJAYAKUMAR, S.; SCHAAL, S. Encyclopedia of machine learning. In: **Encyclopedia of Machine Learning**. New York, USA: Springer USA, 2010. p. 613–624.

UHM, K.-H.; CHO, H.; XU, Z.; LIM, S.; JUNG, S.-W.; HONG, S.-H.; KO, S.-J. Exploring 3d u-net training configurations and post-processing strategies for the miccai 2023 kidney and tumor segmentation challenge. **arXiv preprint arXiv:2312.05528**, 2023. DOI:<https://doi.org/10.48550/arXiv.2312.05528>.

Univen. **Univen - Equipamentos de Diagnósticos por Imagem**. 2025. Disponível em: <https://www.univen.com.br/produto/tomografo-revolution-act-vet-ge-healthcare/>. Acesso em: 17 de Janeiro 2025.

VISPI, A. Recursive learning reinforced by redefining the train and validation volumes of an encoder-decoder segmentation model. In: **International Challenge on Kidney and Kidney Tumor Segmentation**. Vancouver, BC, Canada: Springer, 2023. p. 126–138.

VORA, K.; YAGNIK, S.; SCHOLAR, M. A survey on backpropagation algorithms for feedforward neural networks. **Int. J. Eng. Dev. Res**, v. 1, n. 3, p. 193–197, 2014.

WAKAMATSU, K.; ONO, S. Transunet with unified focal loss for class-imbalanced semantic segmentation. **Artificial Life and Robotics**, Springer Nature, v. 29, n. 1, p. 101–106, 2023.

WANG, Y.; DAI, Y.; ZHANG, J.; YIN, J. Cascaded nnu-net for kidney and kidney tumor segmentation. In: **International Challenge on Kidney and Kidney Tumor Segmentation**. Vancouver, BC, Canada: Springer, 2023. p. 114–119.

World Cancer Research Fund. **Kidney cancer statistics**. 2024. Acesso em: 5 de Março 2025. Disponível em: <https://www.wcrf.org/preventing-cancer/cancer-statistics/kidney-cancer-statistics/>.

World Health Organization. **Global cancer burden growing, amidst mounting need for services**. 2024. Acesso em: 5 de Março 2025. Disponível em: <https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services>.

WU, M.; LIU, Z. Less is more: Contrast attention assisted u-net for kidney, tumor and cyst segmentations. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 46–52.

XIAO, Y.; YANG, G.; SONG, S. **Lesion Segmentation in Surgical and Diagnostic Applications: MICCAI 2022 Challenges, CuRIOUS 2022, KiPA 2022 and MELA 2022, Held in Conjunction with MICCAI 2022, September 18–22, 2022, Proceedings**. Singapore: Springer Nature, 2023. v. 13648.

XU, Z. Martin bland (2015): An introduction to medical statistics. **Statistical Papers**, v. 58, n. 3, p. 953–954, set. 2017.

YANG, X.; ZHANG, J.; ZHANG, J.; XIA, Y. Transfer learning for kits21 challenge. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 158–163.

YEUNG, M.; SALA, E.; SCHÖNLIEB, C.-B.; RUNDO, L. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. **Computerized Medical Imaging and Graphics**, Elsevier, v. 95, p. 102026, 2022. Disponível em: <https://arxiv.org/abs/2102.04525>.

ZANG, M.; WYSOCZANSKI, A.; ANGELINI, E.; LAINE, A. F. 3d u-net based semantic segmentation of kidneys and renal masses on contrast-enhanced ct. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 143–150.

ZHANG, J.; XIE, Y.; XIA, Y.; SHEN, C. Dodnet: Learning to segment multi-organ and tumors from multiple partially labeled datasets. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. Nashville, TN, USA: [s.n.], 2021. p. 1195–1204.

ZHAO, H.; SHI, J.; QI, X.; WANG, X.; JIA, J. Pyramid scene parsing network. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. Honolulu, HI, USA: IEEE, 2017. p. 2881–2890.

ZHAO, Z.; CHEN, H.; WANG, L. A coarse-to-fine framework for the 2021 kidney and kidney tumor segmentation challenge. In: **Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings**. Strasbourg, France: Springer, 2022. p. 53–58.

ZHENG, S.; SUN, Q.; YE, X.; LI, W.; YU, L.; YANG, C. Multi-scale adversarial learning with difficult region supervision learning models for primary tumor segmentation. **Physics in Medicine & Biology**, IOP Publishing, v. 69, n. 8, p. 085009, 2024. DOI:<https://doi.org/10.1088/1361-6560/ad3321>.