

# **UNIVERSIDADE FEDERAL DO MARANHÃO**

## **Programa de Pós-Graduação em Ciência da Computação**

**Felipe Rogério Silva Teles**

***Segmentação Automática do Pâncreas em Tomografias  
Computadorizadas Abdominais Utilizando uma Abordagem  
Orientada a Atlas Probabilístico e Aprendizado Profundo***

**São Luís  
2026**

Felipe Rogério Silva Teles

**Segmentação Automática do Pâncreas em Tomografias Computadorizadas Abdominais Utilizando uma Abordagem Orientada a Atlas Probabilístico e Aprendizado Profundo**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. João Otávio Bandeira Diniz

São Luís - MA

2026

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

Teles, Felipe.

Segmentação Automática do Pâncreas em Tomografias Computadorizadas Abdominais Utilizando uma Abordagem Orientada a Atlas Probabilístico e Aprendizado Profundo / Felipe Teles. - 2026.

75 p.

Orientador(a): João Bandeira.

Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, Auditório do Núcleo de Computação Aplicada, 2026.

1. Segmentação Médica. 2. Pâncreas. 3. Aprendizado Profundo. 4. Processamento de Imagens. 5. Tomografia Computadorizada. I. Bandeira, João. II. Título.

Felipe Rogério Silva Teles

# **Segmentação Automática do Pâncreas em Tomografias Computadorizadas Abdominais Utilizando uma Abordagem Orientada a Atlas Probabilístico e Aprendizado Profundo**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Trabalho \_\_\_\_\_, São Luís - MA, 27 de Abril de 2026:

---

**Prof. Dr. João Otávio Bandeira Diniz**  
Orientador  
Universidade Federal do Maranhão

---

**Prof. Dr. Darlan Bruno Pontes  
Quintanilha**  
Examinador Interno  
Universidade Federal do Maranhão

---

**Prof. Dr. Carlos Fernando da Silva  
Ramos**  
Examinador Externo  
Instituto Superior de Engenharia do Porto

São Luís - MA  
2026

*Este trabalho é dedicado a todos aqueles que seguem seus sonhos.*

# Agradecimentos

Gostaria de agradecer em primeiro lugar a Deus, que me concebeu o dom de correr atrás de conhecimento e ajudar as pessoas ao meu redor. Sem ele nada disso seria possível.

Agradeço profundamente à minha família, em especial aos meus pais, à minha irmã e ao meu primo Vinícius, por sempre me oferecerem todo o suporte emocional de que precisei e por incentivarem meu desenvolvimento acadêmico. Aos meus amigos, um salve a Neilson, Daniel, Celso, Jhones, Lucas, Rodrigo, Rafael, Sergio, Luiz e Vinícius, por me acompanharem nesta enriquecedora jornada.

Estendo minha gratidão aos meus professores, pilares fundamentais na construção da minha base acadêmica, principalmente ao Dr. Anselmo Cardoso de Paiva, que me orientou em momentos decisivos dentro e fora da universidade.

Um agradecimento especial ao meu orientador, professor Dr. João Otávio Bandeira Diniz, pelos anos dedicados a pavimentar meu caminho na academia e a me direcionar ao sucesso.

*– E o que havia de errado com os outros alquimistas, que buscaram ouro e não conseguiram?*

*– Buscavam apenas ouro – respondeu seu companheiro. – Buscavam o tesouro de sua Lenda Pessoal, sem desejarem viver a própria Lenda.*

*(Trecho do livro "O Alquimista", de Paulo Coelho)*

# Resumo

A segmentação automática do pâncreas em imagens de Tomografia Computadorizada (TC) é uma tarefa desafiadora devido à elevada variabilidade anatômica do órgão, ao baixo contraste entre tecidos adjacentes e à grande quantidade de dados presentes nos exames volumétricos. Essa tarefa é importante para diversas aplicações clínicas, como diagnóstico auxiliado por computador, planejamento de tratamentos e análise quantitativa de estruturas anatômicas. No entanto, a identificação manual do pâncreas em exames de TC pode demandar tempo significativo dos especialistas, tornando o processo suscetível a variações entre observadores. Nesse contexto, métodos computacionais baseados em aprendizado profundo têm sido amplamente utilizados para auxiliar na análise automática de imagens médicas. Assim, este trabalho propõe um método automatizado para segmentação do pâncreas em imagens de TC abdominal. O método desenvolvido é composto por diferentes etapas, incluindo pré-processamento das imagens, filtragem automática de fatias sem a presença do órgão por meio de uma rede neural convolucional, extração da região de interesse utilizando um atlas probabilístico e aplicação de modelos de aprendizado profundo para realizar a segmentação. Para essa etapa, foram treinadas diferentes arquiteturas de redes neurais, incluindo DeepLabV3, U-Net++ e SegFormer. Posteriormente, os resultados gerados pelos modelos foram combinados por meio de uma estratégia de *ensemble* com o objetivo de aumentar a robustez da segmentação. O método foi avaliado utilizando a base de dados *Medical Segmentation Decathlon*, composta por exames de TC abdominal anotados manualmente por especialistas. Os resultados experimentais demonstraram que o ensemble proposto alcançou desempenho superior aos modelos individuais, obtendo coeficiente Dice de 78,55%, IoU de 68,17% e Recall de 84,41%. Dessa forma, os resultados obtidos indicam que a abordagem proposta apresenta potencial para auxiliar especialistas na análise de exames de TC, contribuindo para o desenvolvimento de sistemas de apoio ao diagnóstico baseados em aprendizado profundo.

**Palavras-chave:** Segmentação médica, Pâncreas, Tomografia Computadorizada, Aprendizado Profundo, Atlas Probabilístico.

# Abstract

Automatic segmentation of the pancreas in Computed Tomography (CT) images is a challenging task due to the organ's high anatomical variability, low contrast between adjacent tissues, and the large amount of data present in volumetric examinations. This task is important for various clinical applications, such as computer-aided diagnosis, treatment planning, and quantitative analysis of anatomical structures. However, manual identification of the pancreas in CT scans can demand significant time from specialists, making the process susceptible to variations between observers. In this context, computational methods based on deep learning have been widely used to assist in the automatic analysis of medical images. Thus, this work proposes an automated method for pancreas segmentation in abdominal CT images. The developed method consists of different stages, including image preprocessing, automatic filtering of slices without the organ using a convolutional neural network, extraction of the region of interest using a probabilistic atlas, and application of deep learning models to perform the segmentation. For this stage, different neural network architectures were trained, including DeepLabV3, U-Net++, and SegFormer. Subsequently, the results generated by the models were combined using an ensemble strategy to increase the robustness of the segmentation. The method was evaluated using the Medical Segmentation Decathlon database, composed of abdominal CT scans manually annotated by specialists. The experimental results demonstrated that the proposed ensemble achieved superior performance to the individual models, obtaining a Dice coefficient of 78.55%, an IoU of 68.17%, and a Recall of 84.41%. Thus, the results obtained indicate that the proposed approach has the potential to assist specialists in the analysis of CT scans, contributing to the development of diagnostic support systems based on deep learning.

**Keywords:** Medical segmentation, Pancreas, Computed Tomography, Deep Learning, Probabilistic Atlas.

# Lista de ilustrações

Figura 1 – Localização do pâncreas. . . . .	21
Figura 2 – Exemplo de TC. . . . .	23
Figura 3 – Neurônio biológico . . . . .	26
Figura 4 – Representação do Perceptron. . . . .	27
Figura 5 – Organização das camadas de uma CNN . . . . .	29
Figura 6 – Arquitetura <i>Transformer</i> . . . . .	30
Figura 7 – <i>Multi-Head Self-Attention</i> . . . . .	30
Figura 8 – Bloco residual . . . . .	32
Figura 9 – Estratégias de escalonamento da <i>EfficientNet</i> : baseline, largura, profundidade, resolução e escalonamento composto. . . . .	33
Figura 10 – Arquitetura U-net . . . . .	33
Figura 11 – Arquitetura <i>U-Net++</i> . . . . .	34
Figura 12 – Arquitetura da <i>DeepLabV3</i> . . . . .	35
Figura 13 – Arquitetura da ViT . . . . .	36
Figura 14 – Fonte: Xie et al. (2021) . . . . .	37
Figura 15 – Fonte: Zhao et al. (2024) . . . . .	37
Figura 16 – Fluxograma da metodologia proposta. . . . .	48
Figura 17 – Exemplos da base de dados. . . . .	49
Figura 18 – Representação probabilística do atlas. . . . .	51
Figura 19 – Extração da ROI. . . . .	51
Figura 20 – Mapa de calor do Cenário Favorável. . . . .	62
Figura 21 – Segmentação do Cenário Favorável. . . . .	62
Figura 22 – Mapa de calor do Cenário Desafiador. . . . .	63
Figura 23 – Segmentação do Cenário Desafiador. . . . .	64

# Lista de tabelas

Tabela 1	–	Resumo comparativo dos trabalhos relacionados em segmentação pancreática	47
Tabela 2	–	Resultados da validação cruzada para a <i>ResNet50</i> .	55
Tabela 3	–	Resultados da validação cruzada para a <i>EfficientNetB3</i> .	55
Tabela 4	–	Resultados da validação cruzada para a <i>ViT</i> .	55
Tabela 5	–	Resultados da validação cruzada para a <i>SwinTr</i> .	56
Tabela 6	–	Comparativo de desempenho médio entre os modelos avaliados na etapa de filtragem de fatias. O melhor resultado por métrica está destacado.	56
Tabela 7	–	Impacto do pós-processamento no desempenho da filtragem de fatias.	57
Tabela 8	–	Desempenho de segmentação em porcentagem com e sem detecção de ROI.	57
Tabela 9	–	Comparação de segmentação em porcentagem com diferentes <i>backbones</i> convolucionais.	58
Tabela 10	–	Comparação de segmentação em porcentagem com diferentes <i>backbones</i> baseados em <i>Transformer</i> .	59
Tabela 11	–	Desempenho dos modelos individuais e do <i>ensemble</i> no conjunto de teste.	59
Tabela 12	–	Melhores distribuições de pesos do <i>ensemble</i> e seus respectivos desempenhos.	60
Tabela 13	–	Comparação de pontuações de <i>Dice</i> para diferentes abordagens.	61

# Lista de abreviaturas e siglas

CAD	<i>Diagnóstico Auxiliado por Computador</i>
DL	<i>Aprendizado Profundo</i>
TC	<i>Tomografia Computadorizada</i>
ROI	<i>Região de Interesse</i>
CNN	<i>Redes Neurais Convolucionais</i>
ResNet	<i>Residual Neural Network</i>
NIH	<i>National Cancer Institute</i>
MSD	<i>Medical Segmentation Decathlon</i>
TCIA	<i>The Cancer Imaging Archive</i>
Grad-CAM	<i>Gradient-weighted Class Activation Mapping</i>
RSTN	<i>Recurrent Saliency Transformation Network</i>
LSTM	<i>Long Short-Term Memory</i>
ECC	<i>External Contour Cropping</i>
DCNN	<i>Redes Neurais Convolucionais Profundas</i>
MCMC	<i>Markov Chain Monte Carlo</i>
FCN	<i>Fully Convolutional Network</i>
SrSNET	<i>Segmentation-reSegmentation Net</i>
MLP	<i>Perceptron de Múltiplas Camadas</i>
VP	<i>Verdadeiro Positivo</i>
FP	<i>Falso Positivo</i>
VN	<i>Verdadeiro Negativo</i>
FN	<i>Falso Negativo</i>
IoU	<i>Intersection over Union</i>

ViT	<i>Vision Transformers</i>
ASPP	<i>Atrous Spatial Pyramid Pooling</i>
MiT	<i>Mix Transformer</i>
SwinTr	<i>Swin Transformer</i>
HU	<i>Hounsfield Unit</i>
LLM	<i>Large Language Models</i>
FCN	<i>Fully Convolutional Network</i>
GT	<i>Ground Truth</i>
BCE	<i>Binary Cross-Entropy</i>
GPUs	<i>Graphics Processing Unit</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>1.1</b>	<b>Objetivos</b>	<b>18</b>
1.1.1	Objetivos Específicos	18
1.1.2	Contribuições	19
<b>1.2</b>	<b>Organização do Trabalho</b>	<b>19</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>21</b>
<b>2.1</b>	<b>Anatomia do pâncreas</b>	<b>21</b>
2.1.1	Patologias do pâncreas	22
<b>2.2</b>	<b>Tomografia Computadorizada</b>	<b>22</b>
<b>2.3</b>	<b>Processamento de Imagens Digitais</b>	<b>23</b>
2.3.1	Janelamento	24
2.3.2	Segmentação	24
2.3.3	Atlas probabilístico	25
<b>2.4</b>	<b>Aprendizado de Máquina</b>	<b>25</b>
2.4.1	Aprendizado Profundo	25
2.4.2	Redes Neurais Artificiais	26
2.4.3	Perceptron	27
2.4.4	Função de Ativação	27
2.4.5	Perceptron de Múltiplas Camadas	28
2.4.6	Backpropagation	28
2.4.7	Redes Neurais Convolucionais	28
2.4.8	<i>Transformers</i>	29
2.4.9	<i>Grad-CAM</i>	30
2.4.10	Arquiteturas de Aprendizado Profundo	31
2.4.10.1	<i>ResNet</i>	31
2.4.10.2	<i>EfficientNet</i>	32
2.4.10.3	<i>U-Net</i>	33
2.4.10.4	<i>U-Net++</i>	34
2.4.10.5	<i>DeepLabV3</i>	35
2.4.11	<i>Vision Transformers</i>	35
2.4.11.1	<i>SegFormer</i>	36
2.4.11.2	<i>Swin Transformer</i>	37
2.4.12	Otimização de Redes Neurais	38
2.4.12.1	Transferência de Aprendizado	38

2.4.12.2	Parada antecipada . . . . .	38
2.4.12.3	Validação Cruzada . . . . .	39
2.4.12.4	Busca Em Grade . . . . .	40
2.4.13	<i>Ensemble</i> de Redes . . . . .	40
2.4.13.1	<i>Stacking</i> com Média Ponderada . . . . .	41
<b>2.5</b>	<b>Métricas de Avaliação</b> . . . . .	<b>41</b>
<b>2.6</b>	<b>Considerações Finais</b> . . . . .	<b>43</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS</b> . . . . .	<b>44</b>
<b>3.1</b>	<b>Segmentação do pâncreas utilizando modelos tridimensionais</b> . . . . .	<b>44</b>
<b>3.2</b>	<b>Segmentação do pâncreas utilizando abordagem bidimensional</b> . . . . .	<b>45</b>
<b>3.3</b>	<b>Segmentação do pâncreas utilizando abordagem híbrida</b> . . . . .	<b>45</b>
<b>3.4</b>	<b>Considerações finais</b> . . . . .	<b>46</b>
<b>4</b>	<b>MATERIAL E MÉTODO PROPOSTO</b> . . . . .	<b>48</b>
<b>4.1</b>	<b>Base de dados e pré-processamento</b> . . . . .	<b>49</b>
<b>4.2</b>	<b>Remoção de Fatias sem Presença do Pâncreas</b> . . . . .	<b>49</b>
<b>4.3</b>	<b>Deteção de ROI</b> . . . . .	<b>50</b>
<b>4.4</b>	<b>Segmentação Final</b> . . . . .	<b>52</b>
<b>4.5</b>	<b>Validação do Método</b> . . . . .	<b>52</b>
<b>4.6</b>	<b>Considerações Finais</b> . . . . .	<b>53</b>
<b>5</b>	<b>RESULTADOS</b> . . . . .	<b>54</b>
<b>5.1</b>	<b>Ambiente de Treinamento</b> . . . . .	<b>54</b>
<b>5.2</b>	<b>Experimentos</b> . . . . .	<b>54</b>
5.2.1	Experimento 1: Seleção de Rede para a Filtragem de Fatias . . . . .	54
5.2.2	Experimento 2: Validação da Filtragem de Fatias . . . . .	56
5.2.3	Experimento 3: Segmentação Com e Sem Deteção de ROI . . . . .	57
5.2.4	Experimento 4: Avaliação de diferentes arquiteturas de <i>backbone</i> . . . . .	58
5.2.5	Experimento 5: Treinamento das Redes de Segmentação . . . . .	59
5.2.6	Experimento 6: Aplicação do <i>Ensemble</i> . . . . .	60
<b>5.3</b>	<b>Comparação com trabalhos relacionados</b> . . . . .	<b>60</b>
5.3.1	Estudo de Caso . . . . .	61
5.3.1.1	Estudo de Caso 1 . . . . .	61
5.3.1.2	Estudo de Caso 2 . . . . .	63
<b>5.4</b>	<b>Considerações Finais</b> . . . . .	<b>64</b>
<b>6</b>	<b>CONCLUSÃO</b> . . . . .	<b>65</b>
<b>6.1</b>	<b>Trabalhos Futuros</b> . . . . .	<b>66</b>
<b>6.2</b>	<b>Agradecimentos</b> . . . . .	<b>67</b>
<b>6.3</b>	<b>Produções Científicas</b> . . . . .	<b>67</b>

**REFERÊNCIAS** ..... 69

# 1 Introdução

De acordo com a Agência Internacional de Pesquisa sobre o Câncer, o câncer de pâncreas figura entre os tipos mais letais, ocupando a sexta posição em mortalidade global (BRAY et al., 2024). Em 2022, a doença foi responsável por mais de 450 mil óbitos em todo o mundo, evidenciando sua elevada agressividade e o desafio associado ao diagnóstico precoce (BRAY et al., 2024). Nesse cenário, o desenvolvimento de soluções computacionais capazes de auxiliar especialistas no processo diagnóstico torna-se fundamental, contribuindo para o aumento das taxas de sucesso terapêutico e para a redução da mortalidade.

No âmbito do tratamento e do planejamento de radioterapia, bem como no monitoramento da evolução da doença, a identificação precisa do pâncreas em imagens médicas constitui uma etapa inicial essencial (DANIEL et al., 2024). Considerando sua localização na região abdominal, a análise do órgão é realizada, em geral, por meio de exames de Tomografia Computadorizada (TC) do abdômen, não sendo necessária a utilização de imagens de corpo inteiro. A TC destaca-se como uma das principais ferramentas no diagnóstico de patologias pancreáticas, por fornecer imagens de alta resolução que possibilitam a localização precisa da região de interesse e a identificação de possíveis lesões (SAMEI; PELC, 2020).

Introduzida entre as décadas de 1960 e 1970 (BUZUG, 2011), a técnica baseia-se na emissão de radiação eletromagnética na faixa dos raios X, cuja interação com os tecidos biológicos permite a reconstrução de imagens internas do corpo humano. A qualidade dessas imagens depende diretamente de parâmetros físicos, como a energia dos elétrons e a tensão aplicada entre o cátodo e o ânodo, fatores que influenciam o nível de contraste e a definição das estruturas anatômicas (YANG et al., 2018).

Com o avanço das técnicas de processamento de imagens e aprendizado de máquina, os sistemas de diagnóstico auxiliado por computador (CAD, do inglês *Computer-Aided Diagnosis*) têm se consolidado como uma realidade cada vez mais presente na prática médica (DINIZ et al., 2021b; FERNANDES et al., 2023; DANIEL et al., 2024). Tais sistemas contribuem para o aumento da precisão diagnóstica, redução de custos operacionais e diminuição do tempo de espera dos pacientes, impactando positivamente a qualidade do atendimento em saúde.

O uso de técnicas de aprendizado profundo (DL, do inglês *Deep Learning*) tem sido amplamente adotado para diversas tarefas na área médica (JUNIOR; CRUZ; DINIZ, 2024; DINIZ et al., 2021a). Entre essas abordagens, as Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) destacam-se por sua eficácia, oferecendo soluções para desafios complexos, como a segmentação automática do pâncreas em TC.

Entretanto, uma nova classe de modelos de DL, os *Transformers* (VASWANI, 2017), vem ganhando crescente relevância também em atividades médicas, por introduzirem o conceito de atenção aos modelos, permitindo a captura de dependências de longo alcance nos dados e ampliando o potencial para análises mais precisas.

Todavia, a tarefa de segmentar o pâncreas ainda é desafiadora devido a três fatores principais. O principal deles é a elevada variação anatômica do pâncreas entre os pacientes, que apresenta diferenças significativas em forma, tamanho e posição (YAO; SONG; LIU, 2020). Outro fator relevante é a qualidade da imagem, que depende da radiação utilizada durante o procedimento, fazendo com que os primeiros exames tendam a apresentar uma resolução inferior, dificultando a identificação de patologias no órgão de estudo. Por fim, destaca-se o elevado custo computacional associado ao processamento de dados volumétricos, uma vez que exames de TC geram imagens tridimensionais que demandam maior capacidade de processamento (LEE et al., 2021).

## 1.1 Objetivos

Este estudo propõe um método automatizado para segmentação do pâncreas em imagens de TC, combinando técnicas avançadas de pré-processamento de imagem com algoritmos de DL.

A proposta visa oferecer aos médicos uma ferramenta diagnóstica mais confiável e precisa, reduzindo o tempo de espera dos pacientes e otimizando os custos operacionais das clínicas e hospitais.

### 1.1.1 Objetivos Específicos

Para alcançar o objetivo geral deste trabalho, alguns objetivos específicos devem ser atingidos:

- Fazer uma análise sistemática da literatura a fim de encontrar possíveis melhorias para a tarefa de segmentação do pâncreas.
- Avaliar uma abordagem baseada em DL para identificar e filtrar automaticamente as fatias que não contêm o pâncreas.
- Avaliar o uso de um atlas probabilístico para auxiliar na delimitação da região de interesse (ROI) nas imagens de treinamento.
- Avaliar o uso de modelos de DL treinados para segmentar o pâncreas.
- Avaliar uma técnica de combinação dos modelos treinados através do *ensemble*.

- Analisar o desempenho por meio de métricas de classificação (acurácia, especificidade, precisão e *F1-score*) e segmentação médica (*Dice*, *Intersection Over Union* (IoU) e *Recall*).

### 1.1.2 Contribuições

Destacam-se como principais contribuições:

- Realização de uma análise sistemática da literatura sobre segmentação do pâncreas em imagens de TC, aliada à avaliação do desempenho de redes de DL, evidenciando desafios, limitações e oportunidades de melhoria para a tarefa;
- Proposição de um método automático para segmentação da região pancreática em imagens de TC, alcançando precisão e sensibilidade promissoras, com segmentações de alta qualidade em mais da metade das fatias de cada exame;
- Desenvolvimento de técnicas avançadas de processamento de imagem e seleção de fatias relevantes, reduzindo o número de amostras necessárias para o treinamento, diminuindo o custo computacional e eliminando características desnecessárias, o que resultou em melhorias na qualidade final da segmentação.

## 1.2 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 apresenta a fundamentação teórica das técnicas utilizadas, abordando conceitos essenciais sobre processamento de imagens médicas, DL e arquiteturas de redes neurais empregadas na segmentação.
- O Capítulo 3 descreve trabalhos relacionados à segmentação do pâncreas, incluindo estudos que utilizam abordagens semelhantes ou relevantes para este trabalho.
- O Capítulo 4 apresenta as etapas que compõem a metodologia proposta, incluindo a descrição do conjunto de dados utilizado, os procedimentos de pré-processamento das imagens, as arquiteturas de redes neurais adotadas, bem como os detalhes do processo de treinamento e avaliação dos modelos.
- O Capítulo 5 apresenta os resultados obtidos e as discussões relacionadas aos experimentos realizados para a segmentação automática da região pancreática em imagens de TC.

- O Capítulo 6 apresenta as considerações finais do trabalho, destacando as principais contribuições, limitações do estudo e possíveis direções para trabalhos futuros, além de mencionar os artigos científicos desenvolvidos a partir desta pesquisa.

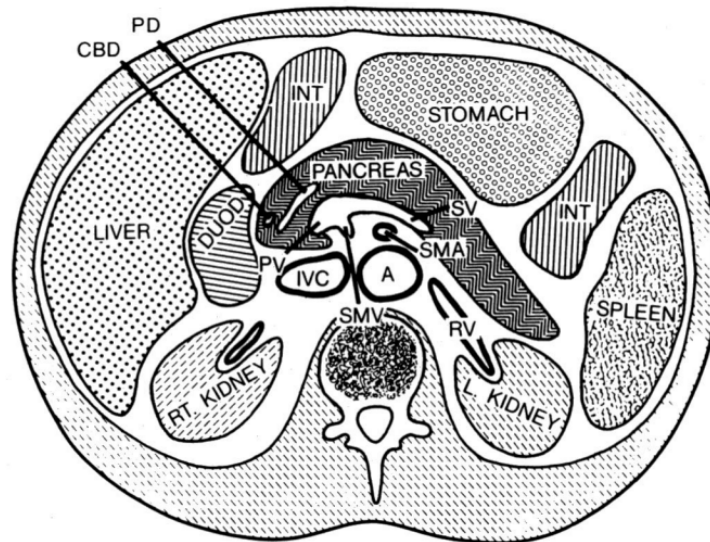
## 2 Fundamentação Teórica

Este capítulo apresenta os fundamentos teóricos necessários para a compreensão do problema abordado e das técnicas utilizadas no desenvolvimento do método proposto. São discutidos conceitos relacionados à anatomia do pâncreas, TC, processamento de imagens digitais, aprendizado profundo e métricas de avaliação aplicadas à segmentação de imagens médicas.

### 2.1 Anatomia do pâncreas

O pâncreas é um órgão complexo que está localizado na porção superior do abdômen, estendendo-se entre a concavidade do duodeno e o hilo esplênico, como visto na Figura 1. Sua anatomia é caracterizada por elevada variabilidade entre indivíduos, tanto em forma quanto em tamanho e posicionamento (MAHADEVAN, 2019; LONGNECKER; THOMPSON, 2023).

Figura 1 – Localização do pâncreas.



Fonte: Bockman (1993)

Do ponto de vista funcional, o pâncreas exerce funções exócrinas e endócrinas. A porção exócrina é responsável pela secreção de enzimas digestivas, como lipases, amilases e proteases, enquanto a porção endócrina, constituída pelas ilhotas de Langerhans, regula o metabolismo glicêmico por meio da secreção de hormônios como insulina e glucagon (MAHADEVAN, 2019).

Essa variabilidade anatômica e funcional representa um dos principais desafios para a segmentação automática do órgão em imagens médicas, uma vez que dificulta a generalização de modelos computacionais.

### 2.1.1 Patologias do pâncreas

Alterações no pâncreas podem gerar respostas inflamatórias, principalmente em condições como pancreatite aguda e crônica. Durante as fases iniciais de inflamação, observa-se hiperemia e edema do tecido, que se manifestam como regiões avermelhadas e aumentadas. À medida que o processo inflamatório evolui, pode ocorrer necrose gordurosa, fibrose e perda progressiva da função pancreática, comprometendo tanto a digestão quanto o equilíbrio metabólico. Em casos crônicos, a fibrose substitui áreas funcionais do órgão, resultando em diminuição da produção de enzimas e, em estágios avançados, possível desenvolvimento de diabetes secundário (FRANTZ; SOUZA-MELLO; LACERDA, 2012).

Embora as alterações pancreáticas se concentrem majoritariamente na região abdominal superior, manifestações menos comuns podem repercutir em outras áreas do corpo, como alterações metabólicas sistêmicas e complicações vasculares. A progressão das lesões inflamatórias tende a modificar gradualmente a aparência do tecido, que passa de uma coloração avermelhada em estágios iniciais para tons mais escuros e irregulares conforme se intensificam a inflamação e a fibrose. Quando o tecido pancreático funcional é substituído por tecido fibroso, há perda significativa de sua capacidade fisiológica, comprometendo de forma relevante a qualidade de vida do indivíduo (DOLENŠEK et al., 2017; FUSCO; EL-GOHARY; GITTES, 2017).

## 2.2 Tomografia Computadorizada

A Tomografia Computadorizada (TC) é uma técnica de imagem baseada na atenuação de raios X, capaz de gerar reconstruções tridimensionais a partir de múltiplas projeções adquiridas ao redor do objeto analisado (SAMEI; PELC, 2020).

No contexto médico, os valores de atenuação são representados em unidades Hounsfield (HU), o que permite a diferenciação entre diferentes tipos de tecido. Essa característica torna a TC uma ferramenta essencial para o diagnóstico e acompanhamento de patologias (SAMEI; PELC, 2020; BUZUG, 2011).

Entretanto, a qualidade das imagens obtidas depende de diversos fatores, incluindo parâmetros de aquisição e níveis de radiação, o que pode impactar diretamente a visibilidade das estruturas anatômicas e, conseqüentemente, o desempenho de métodos automatizados de segmentação.

O princípio físico envolve a emissão de fótons de alta energia, geralmente produzidos

a partir de um alvo metálico, como o tungstênio. Ao atravessarem o corpo, esses fótons sofrem processos de absorção e espalhamento que dependem das propriedades físicas dos tecidos, como densidade e número atômico. Essas variações são utilizadas para reconstruir imagens baseadas em coeficientes de atenuação (CANTATORE; MÜLLER, 2011). A Figura 2 apresenta um exemplo de uma máquina de TC moderna.

Figura 2 – Exemplo de TC.



Fonte: Cantatore e Müller (2011)

No contexto médico, a TC constitui uma ferramenta de imagem de elevada versatilidade, capaz de caracterizar não apenas a morfologia anatômica, mas também propriedades quantitativas relacionadas à densidade dos tecidos, em virtude da conversão das medidas de atenuação dos raios X em unidades numéricas padronizadas (as unidades Hounsfield). Isto acaba por permitir a avaliação não invasiva de estruturas internas com alta resolução espacial, tornando a TC amplamente empregada para fins diagnósticos, planejamento terapêutico e acompanhamento de diversas patologias (MAZONAKIS; DAMILAKIS, 2016).

### 2.3 Processamento de Imagens Digitais

O processamento digital de imagens remete o conjunto de métodos computacionais voltados à aquisição, transformação e interpretação de imagens. As operações são divididas em três níveis: baixo, médio e alto. As de nível baixo são mais primitivas, muito utilizadas para melhorar a qualidade da imagem, removendo ruídos ou aumentando o contraste. As de nível médio realizam tarefas como segmentação de objetos na imagem, descrição desses objetos e sua classificação. As de nível alto extraem significado de um conjunto de objetos

extraídos pelas operações de nível médio, realizando a análise da imagem com funções cognitivas associadas à visão (GONZALEZ; WOODS, 2000).

No contexto deste trabalho, o processamento de imagens desempenha papel fundamental na melhoria da qualidade dos dados de entrada e na redução de informações irrelevantes, contribuindo diretamente para o desempenho dos modelos de aprendizado profundo.

### 2.3.1 Janelamento

O janelamento é uma técnica de processamento de imagens médicas que consiste em mapear uma faixa de valores da escala de *Hounsfield* para os níveis de cinza visíveis da imagem. Essa faixa é definida por dois parâmetros, sendo o centro da janela, que determina o valor *Hounsfield Unit* (HU) central da faixa de interesse, e a largura da janela, que define a amplitude do intervalo. *Pixels* com valores abaixo do limite inferior são exibidos como preto, enquanto os acima do limite superior são exibidos como branco, e os intermediários são mapeados linearmente para os níveis de cinza (GONZALEZ; WOODS, 2000).

Essa técnica é fundamental para a visualização e análise de estruturas específicas em imagens de tomografia computadorizada, uma vez que a escala de *Hounsfield* abrange uma faixa muito ampla de valores, inviabilizando a visualização simultânea de todos os tecidos com contraste adequado (PAN et al., 2024). Ao ajustar a janela para uma faixa compatível com o tecido de interesse, é possível realçar estruturas relevantes e suprimir regiões irrelevantes. Neste trabalho, o janelamento é aplicado no pré-processamento das fatias de TC para realçar as estruturas abdominais relevantes à segmentação do pâncreas.

### 2.3.2 Segmentação

A segmentação de imagens consiste no processo de particionamento de uma imagem em regiões ou objetos significativos, sendo uma das etapas mais críticas do processamento digital de imagens (GONZALEZ; WOODS, 2000). O objetivo é separar as estruturas de interesse do restante da imagem, de forma que os elementos resultantes sejam mais simples e significativos para análise posterior.

Algumas das abordagens clássicas de segmentação baseiam-se em propriedades como descontinuidade, ou seja, detecção de bordas e contornos como limiarização, crescimento de regiões e agrupamento (CONCI; AZEVEDO, 1998).

Com o advento do aprendizado profundo, métodos baseados em redes neurais passaram a dominar as tarefas de segmentação semântica, superando as abordagens tradicionais, especialmente em imagens médicas com alta variabilidade anatômica, como é o caso do pâncreas (NETO et al., 2024).

### 2.3.3 Atlas probabilístico

O atlas probabilístico é uma técnica estatística utilizada para modelar a distribuição espacial de estruturas anatômicas em uma população. Ele fornece uma representação média da localização de um órgão, permitindo a identificação de regiões de maior probabilidade de ocorrência (DINIZ et al., 2021a; CASAMITJANA et al., 2025). A fórmula matemática que define o atlas está presente na Equação 2.1.

$$A = \frac{1}{N} \sum_{i=1}^N \frac{I_i}{255} \quad (2.1)$$

onde  $N$  é o número total de *pixels* na área, e  $I_i$  denota a intensidade do  $i$ -ésimo pixel, normalizado para o intervalo  $[0, 1]$ .

Sua utilização permite restringir a área de busca durante a segmentação, reduzindo a complexidade computacional e aumentando a eficiência dos modelos (DINIZ et al., 2021a).

No contexto deste trabalho, o atlas probabilístico é utilizado para delimitar a ROI, contribuindo para a redução de informações irrelevantes nas imagens de entrada.

## 2.4 Aprendizado de Máquina

O aprendizado de máquina baseia-se na construção de modelos capazes de aprender padrões a partir de dados, sendo amplamente aplicado em tarefas de classificação, regressão e segmentação. Nele, uma máquina é treinada a executar uma determinada tarefa por meio de um conjunto de exemplos (LAKSHMANAN; GÖRNER; GILLARD, 2021).

Nesse processo, a máquina aproxima-se da função ideal de solução do problema ao acumular aprendizado dos exemplos anteriores (RAUBER, 2005). Isso é feito com pouca intervenção humana. O aprendizado de máquina é aplicado em vários problemas como reconhecimento de fala, compreensão de texto, diagnósticos médicos e aproximações e previsões nas mais diversas áreas (GOODFELLOW et al., 2016).

### 2.4.1 Aprendizado Profundo

O aprendizado profundo é uma subárea do aprendizado de máquina que utiliza redes neurais com múltiplas camadas para modelar relações complexas nos dados. Sua principal vantagem é a capacidade de aprender representações hierárquicas diretamente a partir das entradas (PIRES; SANTOS; PEREIRA, 2025).

É uma técnica robusta, capaz de processar dados que possuem muitas dimensões, fator limitante para as técnicas de aprendizado raso. Geralmente, uma arquitetura profunda

define uma função a partir de uma composição de funções, com a entrada de cada uma sendo a saída da anterior (CAO et al., 2022; GOODFELLOW et al., 2016).

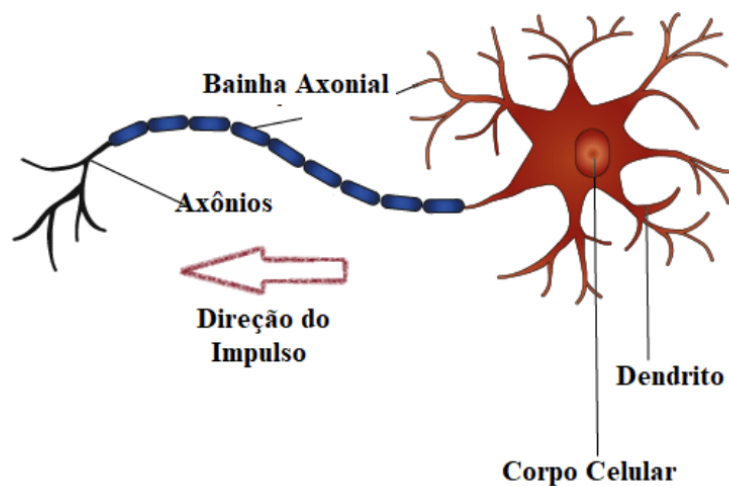
As subseções seguintes detalham os componentes do aprendizado profundo utilizados neste trabalho.

## 2.4.2 Redes Neurais Artificiais

As redes neurais artificiais tornaram-se o alicerce da análise moderna de imagens médicas, fornecendo mecanismos avançados para detecção, segmentação e classificação de estruturas anatômicas (PAN et al., 2024). Entre as arquiteturas mais influentes encontram-se as CNNs (LECUN et al., 1998) e, mais recentemente, os modelos baseados na arquitetura *Transformer* (VASWANI, 2017), que redefiniram o paradigma de processamento de informação em visão computacional.

A denominação redes neurais artificiais tem origem no sistema nervoso biológico, uma vez que esses modelos computacionais foram inspirados no funcionamento dos neurônios (RAUBER, 2005), como pode ser analisado melhor na Figura 3.

Figura 3 – Neurônio biológico



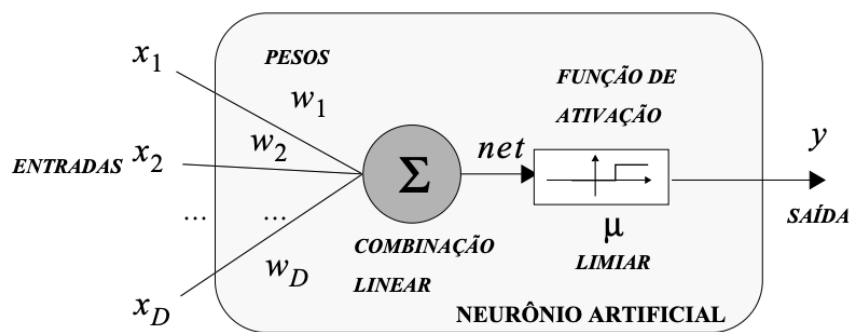
Fonte: Furtado (2019)

Em termos conceituais, um neurônio artificial busca reproduzir, de forma simplificada, o comportamento do neurônio biológico: os sinais de entrada representam os estímulos recebidos, os pesos simulam a força das conexões sinápticas, a soma ponderada corresponde à integração dos sinais, e a função de ativação modela o disparo do neurônio. A saída desse processo representa a resposta do neurônio artificial, e o conjunto de múltiplos neurônios interconectados, organizados em camadas, é responsável pela formação do resultado final da rede neural como um todo (FLECK et al., 2016).

### 2.4.3 Perceptron

O primeiro modelo formal de neurônio artificial foi proposto por [McCulloch e Pitts \(1943\)](#). Por meio de um modelo matemático, ele descrevia o neurônio como uma unidade lógica capaz de realizar operações booleanas simples. Esse modelo pioneiro estabeleceu as bases conceituais para a representação computacional do comportamento neuronal, embora não incorporasse mecanismos de aprendizado ([RAUBER, 2005](#)). Posteriormente, foi introduzido o *perceptron* ([ROSENBLATT, 1958](#)), cuja representação se faz presente na Figura 4.

Figura 4 – Representação do Perceptron.



Fonte: [Rauber \(2005\)](#)

O *perceptron* é um classificador linear capaz de aprender a partir de exemplos por meio do ajuste iterativo de seus pesos, com base no erro entre a saída prevista e a saída desejada. Ele é composto por entradas  $x_i$ , cada uma associada a um peso  $w_i$ , cuja soma ponderada é calculada e submetida a uma função de ativação, geralmente uma função degrau, que determina a saída do neurônio. O processo de aprendizado ocorre por meio da regra delta, na qual os pesos são ajustados proporcionalmente ao erro cometido na classificação, permitindo que o modelo aprenda progressivamente a separar classes linearmente separáveis ([HAYKIN, 2001](#)).

### 2.4.4 Função de Ativação

A função de ativação é um componente essencial das redes neurais artificiais, responsável por introduzir não-linearidade ao modelo, permitindo que ele aprenda padrões complexos nos dados ([HAYKIN, 2001](#)). Sem essa não-linearidade, uma rede neural com múltiplas camadas seria matematicamente equivalente a um único neurônio linear, limitando drasticamente sua capacidade de representação.

### 2.4.5 Perceptron de Múltiplas Camadas

A Rede Neural *Perceptron* de Múltiplas Camadas (MLP, do inglês *Multilayer Perceptron*) é uma arquitetura de rede neural artificial formada por diversas camadas de neurônios interligados. Nessa estrutura, a primeira camada recebe os dados de entrada, enquanto as camadas subsequentes utilizam como entrada a saída produzida pela camada imediatamente anterior, em seguida, cada neurônio realiza um processamento sobre os dados recebidos e encaminha o resultado para a próxima camada da rede. A última camada é responsável por produzir a saída final do modelo, representando o nível mais alto de abstração das informações processadas. As camadas situadas entre a camada de entrada e a camada de saída são denominadas camadas ocultas (RUMELHART et al., 1986).

O funcionamento de uma camada em uma MLP pode ser descrito pela Equação 2.2. Nessa equação,  $y^{(l)}$  representa o vetor de saída da  $l$ -ésima camada,  $\sigma$  corresponde à função de ativação aplicada aos neurônios,  $W^{(l)}$  indica a matriz de pesos que conecta as camadas  $l - 1$  e  $l$ ,  $y^{(l-1)}$  representa a saída da camada anterior e  $b^{(l)}$  corresponde ao termo de viés (*bias*).

$$y^{(l)} = \sigma \left( W^{(l)} y^{(l-1)} + b^{(l)} \right) \quad (2.2)$$

### 2.4.6 Backpropagation

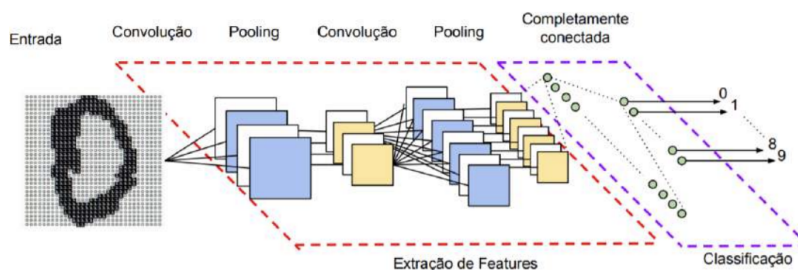
O *backpropagation*, ou retropropagação do erro, é o algoritmo fundamental para o treinamento de redes neurais com múltiplas camadas, sendo responsável por calcular o gradiente da função de perda em relação a cada peso da rede (RUMELHART; HINTON; WILLIAMS, 1986). O algoritmo opera em duas etapas distintas: a propagação direta (*forward pass*), na qual os dados de entrada percorrem a rede até a geração da saída, e a propagação reversa (*backward pass*), na qual o erro é retropropagado da camada de saída até a camada de entrada, utilizando a regra da cadeia do cálculo diferencial para atualizar os pesos (HINTON; SALAKHUTDINOV, 2006).

A eficiência do *backpropagation* reside na sua capacidade de distribuir o peso pelo erro entre todos os neurônios da rede de forma computacionalmente viável, tornando possível o treinamento de arquiteturas profundas com milhões de parâmetros (HINTON; SALAKHUTDINOV, 2006).

### 2.4.7 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNN), por sua vez, são baseadas no córtex visual dos seres vivos (LECUN et al., 1998). Sua arquitetura tem como característica três tipos de camadas: camadas de convolução, camadas de pooling e camadas completamente conexas, como ilustrado na Figura 5.

Figura 5 – Organização das camadas de uma CNN



Fonte: Vargas, Paes e Vasconcelos (2016)

A camada de convolução, geralmente posicionada após a entrada dos dados, concentra a maior parte do custo computacional e é composta por mapas de características cujos neurônios se conectam a regiões locais da camada anterior, denominadas campos receptivos, compartilhando pesos para reduzir a complexidade do modelo (ZEILER; FERGUS, 2014).

A camada de pooling reduz a dimensionalidade dos mapas de ativação, diminuindo o sobreajuste e o custo computacional, sendo a max pooling uma das abordagens mais utilizadas. As camadas finais são, em geral, completamente conectadas e responsáveis pela classificação, embora arquiteturas mais recentes substituam essas camadas por operações de pooling visando maior eficiência (ZEILER; FERGUS, 2014).

#### 2.4.8 Transformers

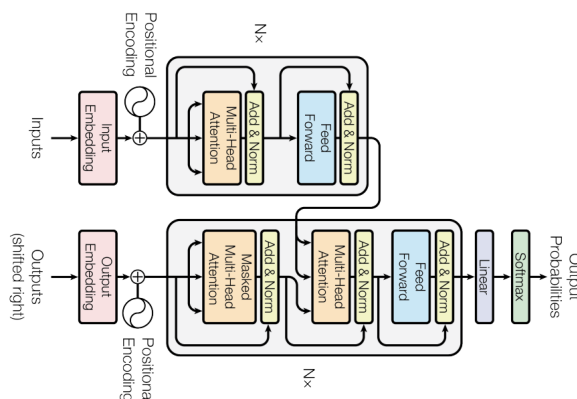
A *Transformer* é uma rede neural profunda proposta por Vaswani (2017), composta principalmente por mecanismos de *self-attention* e camadas totalmente conectadas. Sua arquitetura é estruturada como uma pilha de camadas de *self-attention*, na qual o codificador é formado por N camadas idênticas. Cada uma dessas camadas contém duas subcamadas: uma de *self-attention multihead* e outra do tipo totalmente conectada, semelhante ao perceptron descrito na Seção 2.4.3.

De forma análoga, o decodificador também é constituído por N camadas, porém inclui uma subcamada adicional de autoatenção responsável por processar as representações geradas pelo codificador. A arquitetura completa da *Transformer* é ilustrada na Figura 6.

Cada mecanismo de atenção é responsável por modelar dependências entre os elementos de uma sequência ao atribuir pesos diferenciados às suas representações, de modo a quantificar a relevância relativa de cada elemento no contexto global.

Esse processo é formalizado por meio do mapeamento das entradas em três espaços vetoriais distintos, denominados *queries*, *keys* e *values*, a partir dos quais são calculadas pontuações de similaridade que determinam a contribuição de cada posição da sequência

Figura 6 – Arquitetura *Transformer*

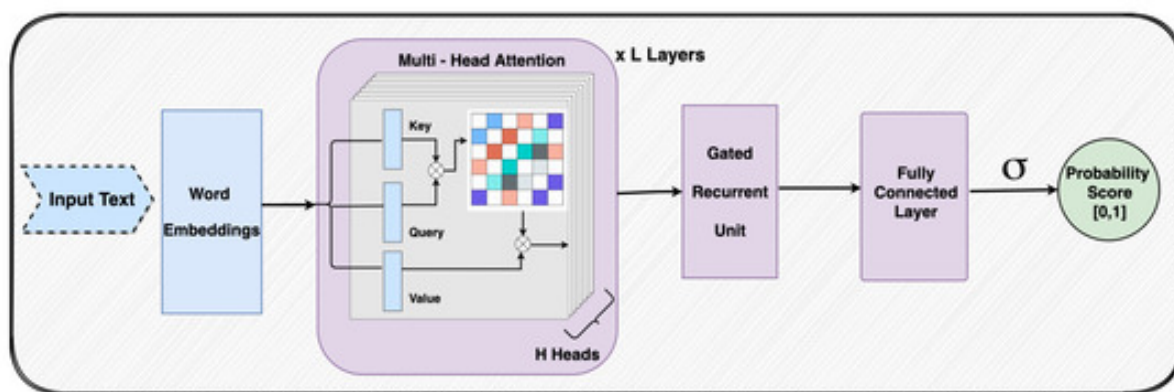


Fonte: Vaswani (2017)

para a representação final.

A Figura 7 apresenta a extensão desse mecanismo, as chamadas múltiplas cabeças de atenção (*Multi-Head Self-Attention*).

Figura 7 – *Multi-Head Self-Attention*



Fonte: Akula e Garibay (2021)

Com essa técnica, o modelo passa a operar simultaneamente em diferentes subespaços de representação, possibilitando a aprendizagem de padrões complementares de dependência semântica e estrutural, o que resulta em representações mais expressivas e robustas. (VASWANI, 2017; DOSOVITSKIY, 2020)

#### 2.4.9 Grad-CAM

O *Gradient-weighted Class Activation Mapping* (Grad-CAM) é uma técnica de interpretabilidade aplicada a redes neurais, cujo objetivo é produzir mapas de ativação

que destacam as regiões da imagem de entrada mais relevantes para a decisão do modelo (SELVARAJU et al., 2017). Diferentemente de abordagens anteriores, o Grad-CAM não exige modificações na arquitetura da rede, sendo aplicável a qualquer rede treinada.

O método utiliza os gradientes do *score* de uma classe de interesse em relação aos mapas de ativação da última camada convolucional. Esses gradientes são agregados globalmente para obter pesos que indicam a importância de cada mapa de ativação. A combinação ponderada desses mapas gera um mapa de calor que é redimensionado para as dimensões da imagem original. Regiões com valores mais elevados no mapa correspondem às áreas que mais influenciaram a predição do modelo para a classe em questão. Neste trabalho, o Grad-CAM é utilizado para verificar se o modelo de segmentação está de fato identificando regiões anatomicamente relevantes ao pâncreas.

## 2.4.10 Arquiteturas de Aprendizado Profundo

Os avanços em aprendizado profundo revolucionaram a classificação de imagens médicas, tornando possível extrair automaticamente características discriminativas de exames complexos e de difícil interpretação visual. Entretanto, a segmentação do pâncreas segue sendo uma tarefa complexa devido a quantidade alta de características semelhantes e sem padrão que a imagem de TC apresenta. Sabendo disso, este trabalho utilizou algumas arquiteturas de aprendizado profundo, técnica apresentada na Seção 2.4.

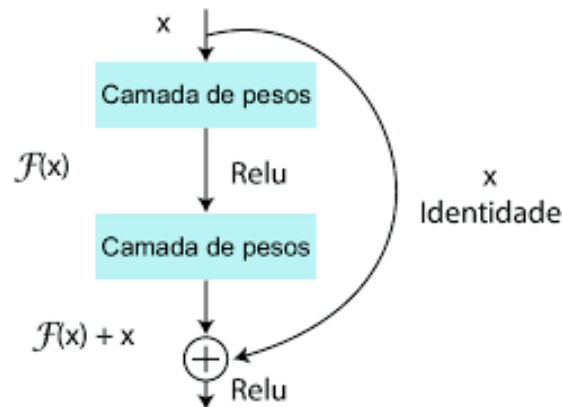
### 2.4.10.1 ResNet

A *ResNet*, proposta por (HE et al., 2016), é uma arquitetura de redes neurais profundas desenvolvida para mitigar o problema da degradação do desempenho observado em redes muito profundas, no qual o aumento do número de camadas não resulta necessariamente em melhor acurácia. A principal inovação da *ResNet* é o uso de conexões residuais, que permitem que a informação seja propagada diretamente entre camadas não adjacentes, facilitando o fluxo de gradientes durante o treinamento. O bloco residual, ilustrado na Figura 8, é composto por uma sequência de camadas convolucionais seguida de uma conexão de atalho, que soma a entrada original do bloco à sua saída transformada (HE et al., 2016).

Além da arquitetura base, a *ResNet* possui diferentes variações que se distinguem principalmente pelo número de camadas utilizadas na rede. Essas variações são identificadas por números que indicam a profundidade total do modelo, como a *ResNet50*.

Em geral, modelos com maior número de camadas apresentam maior capacidade de representação, possibilitando a extração de características mais complexas das imagens. No entanto, redes mais profundas também demandam maior custo computacional e maior quantidade de dados para treinamento.

Figura 8 – Bloco residual



Fonte: Souza et al. (2020)

Neste trabalho, foram avaliadas as arquiteturas *ResNet50* e *ResNet101* como *backbones*, permitindo analisar o impacto do aumento da profundidade da rede no desempenho da tarefa de segmentação.

#### 2.4.10.2 *EfficientNet*

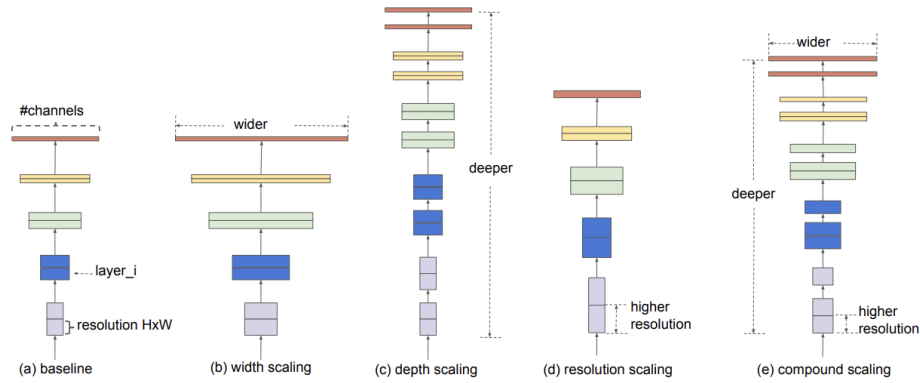
A *EfficientNet* é uma família de redes neurais convolucionais proposta por Tan, Le et al. (2019), projetada para alcançar alto desempenho em tarefas de visão computacional com maior eficiência computacional. Diferentemente de abordagens tradicionais que aumentam arbitrariamente a profundidade ou a largura das redes, a *EfficientNet* introduz um método sistemático de escalonamento denominado *compound scaling*. Esse método permite ampliar simultaneamente três dimensões da rede, sendo elas profundidade, largura e resolução das imagens de entrada, utilizando um conjunto fixo de coeficientes que equilibram essas escalas.

Para ilustrar esse conceito, diferentes estratégias de escalonamento podem ser aplicadas individualmente ou de forma conjunta na arquitetura. Essas abordagens incluem o aumento da largura da rede, da profundidade e da resolução das imagens de entrada, além da combinação dessas dimensões por meio do escalonamento composto, conforme apresentado na Figura 9.

A arquitetura base da *EfficientNet* é denominada *EfficientNet-B0*, obtida por meio de um processo de busca automática de arquitetura. A partir desse modelo inicial, são geradas versões maiores da rede, como *EfficientNet-B1* até *EfficientNet-B7*, que aumentam progressivamente sua capacidade representacional mantendo um uso eficiente de parâmetros e recursos computacionais. Essa estratégia permite obter modelos com melhor desempenho em comparação a arquiteturas tradicionais, mantendo menor complexidade.

Devido à sua eficiência e elevada capacidade de extração de características, as

Figura 9 – Estratégias de escalonamento da *EfficientNet*: baseline, largura, profundidade, resolução e escalonamento composto.



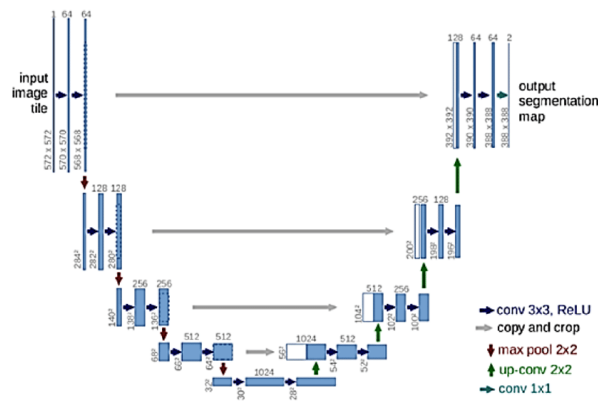
Fonte: Tan, Le et al. (2019)

arquiteturas da família *EfficientNet* têm sido amplamente utilizadas em diferentes tarefas de visão computacional, incluindo classificação de imagens e segmentação semântica. Neste trabalho, variantes dessa arquitetura foram utilizadas como *backbone* para extração de características nas redes avaliadas.

### 2.4.10.3 U-Net

A *U-Net*, proposta por Ronneberger, Fischer e Brox (2015), é uma arquitetura de rede neural convolucional amplamente utilizada em tarefas de segmentação de imagens, especialmente no contexto de imagens médicas. Sua estrutura é caracterizada por um formato em “U”, composto por um caminho de contração (*encoder*) e um caminho de expansão (*decoder*). Sua arquitetura está presente na Figura 10

Figura 10 – Arquitetura U-net



Fonte: Ronneberger, Fischer e Brox (2015)

No *encoder*, a rede extrai progressivamente características de alto nível por meio de camadas convolucionais e operações de *downsampling*, reduzindo a resolução espacial

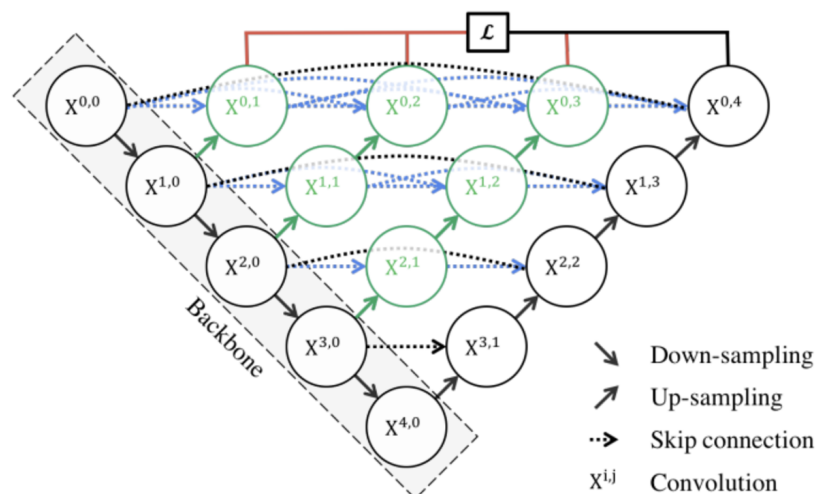
enquanto aumenta a profundidade dos mapas de características. Já no *decoder*, ocorre o processo inverso, no qual operações de *upsampling* restauram a resolução espacial da imagem, permitindo a reconstrução detalhada da segmentação.

Um dos principais diferenciais da U-Net é o uso de conexões de atalho entre camadas correspondentes do *encoder* e do *decoder*. Essas conexões concatenam mapas de características de alta resolução provenientes do caminho de contração com os mapas do caminho de expansão, preservando informações espaciais importantes que poderiam ser perdidas durante o *downsampling*.

#### 2.4.10.4 U-Net++

Os modelos *U-Net* tradicionais podem perder informações espaciais críticas devido à subamostragem excessiva durante o treinamento. Para mitigar isso, a *U-Net++*, proposta por Zhou et al. (2018), introduz conexões densas que permitem uma transição mais gradual entre os recursos do *encoder* e do *decoder*. Essas conexões funcionam como atalhos que permitem combinar características de diferentes níveis de processamento, melhorando a precisão da segmentação, particularmente para estruturas anatômicas complexas (ZHOU et al., 2018). Sua arquitetura está presente na Figura 11.

Figura 11 – Arquitetura *U-Net++*



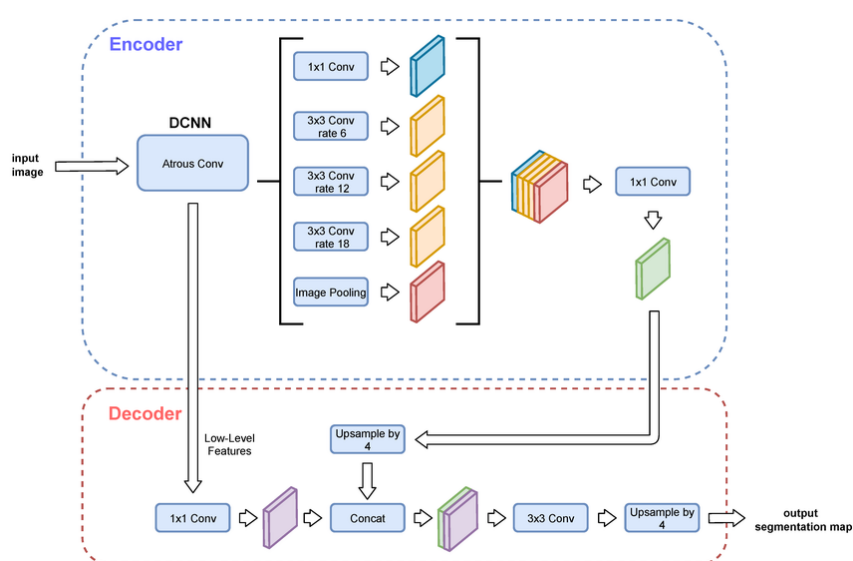
Fonte: Zhou et al. (2018)

Um fator essencial no *U-Net++* é a seleção de um *backbone* apropriado, que serve como extrator de características. Esta escolha influencia na capacidade do modelo em aprender representações hierárquicas. As escolhas comuns incluem *ResNet* (Seção 2.4.10.1) e *EfficientNet* (Seção 2.4.10.2), ambos os quais oferecem um equilíbrio entre profundidade e eficiência (ABDALLA; EL-GINDY et al., 2021).

### 2.4.10.5 DeepLabV3

A *DeepLabV3*, proposta por [Chen et al. \(2017\)](#), é uma arquitetura de segmentação semântica que se destaca pelo uso eficiente de convoluções dilatadas para capturar informações contextuais em múltiplas escalas sem a necessidade de reduzir excessivamente a resolução espacial dos mapas de características. Esta abordagem permite aumentar o campo receptivo da rede sem aumentar significativamente o número de parâmetros ou o custo computacional, preservando detalhes importantes para a segmentação precisa de objetos. Sua arquitetura está presente na Figura 12.

Figura 12 – Arquitetura da *DeeplabV3*



Fonte: [Pedrayes et al. \(2021\)](#)

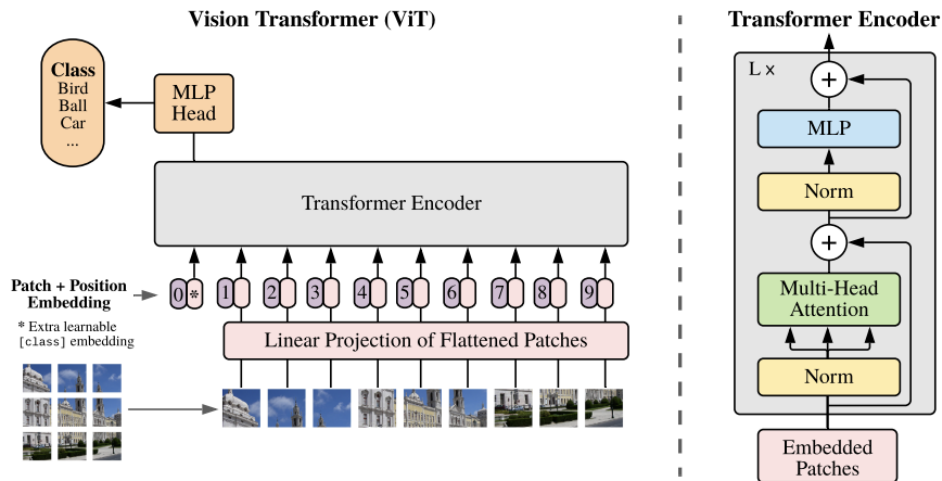
Um dos principais componentes da *DeepLabV3* é o módulo *Atrous Spatial Pyramid Pooling* (ASPP), que aplica convoluções com diferentes taxas de dilatação em paralelo, combinando informações locais e globais. Esse mecanismo possibilita à rede lidar melhor com objetos de tamanhos variados dentro de uma mesma imagem, tornando a arquitetura robusta para diferentes cenários e amplamente utilizada como base para sistemas avançados de segmentação semântica.

### 2.4.11 Vision Transformers

Com base no *Transformer*, [Dosovitskiy \(2020\)](#) propuseram o *Vision Transformer* (ViT), que aplica a *Transformer* no contexto de imagens. Este trabalho apresenta que uma imagem pode ser entendida como uma sequência de *pixels*, tal qual um texto é uma sequência de palavras. O principal desafio é realizar a autoatenção entre todos os *pixels*, o que é pouco viável computacionalmente.

Então, os autores da ViT dividiram as imagens de entrada em recortes, determinando esses recortes como sendo os elementos da sequência. Tornou-se possível, então, realizar a tarefa de classificação utilizando a ViT. A arquitetura da ViT manteve-se quase a mesma do codificador da *Transformer* original. Essa arquitetura, bem como sua camada de codificação, podem ser observadas na Figura 13.

Figura 13 – Arquitetura da Vit



Fonte: [Dosovitskiy \(2020\)](#)

A complexidade computacional da ViT depende principalmente do tamanho da imagem e dos recortes. Cada recorte tem associado a ele o vetor de codificação posicional, que indica a posição na imagem em que o recorte está. Os valores desses vetores são aprendidos. Ademais, é adicionado também um outro vetor para ser aprendido, referente à classificação, para cada recorte. Os recortes passam por uma projeção linear para tornarem-se vetores unidimensionais e a autoatenção ocorre normalmente. A função utilizada nos perceptrons de cada camada é a GELU ([DOSOVITSKIY, 2020](#); [HENDRYCKS, 2016](#)).

#### 2.4.11.1 SegFormer

O *SegFormer* é um modelo de segmentação semântica baseado em *Transformers*, proposto por [Xie et al. \(2021\)](#), que busca combinar simplicidade arquitetural e alta eficiência computacional. A rede, presente na Figura 14 é composta por um codificador hierárquico baseado em *Transformers*, denominado *Mix Transformer* (MiT), responsável por extrair representações multiescala da imagem de entrada, e por um decodificador leve baseado em MLP, que agrega as informações provenientes de diferentes níveis do codificador para produzir o mapa de segmentação final ([XIE et al., 2021](#)).

Diferentemente de outras arquiteturas baseadas em ViT, o *SegFormer* não utiliza codificação posicional explícita, o que aumenta sua robustez a variações de resolução entre

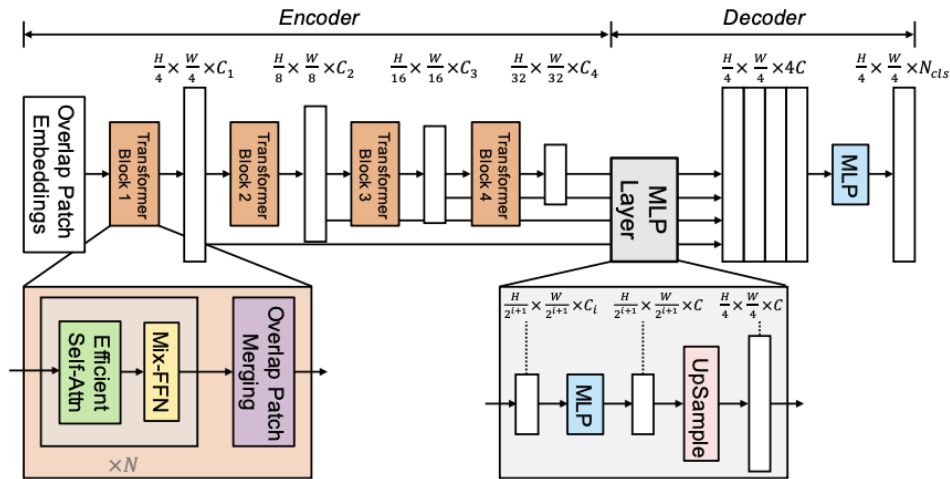


Figura 14 – Fonte: Xie et al. (2021)

treinamento e teste. Além disso, a arquitetura foi projetada para ser escalável, que além de equilibrar custo computacional e desempenho, permite também o uso de diferentes arquiteturas como codificador (XIE et al., 2021).

#### 2.4.11.2 Swin Transformer

O *Swin Transformer* (SwinTr), proposto por Liu et al. (2021), é uma arquitetura baseada em *Transformers* hierárquicos projetada para aplicações de visão computacional, sua arquitetura está presente na Figura 15. Diferentemente dos *Transformers* tradicionais, que operam com autoatenção global e apresentam alto custo computacional, a *SwinTr* utiliza mecanismos de autoatenção restrita a janelas locais, que reduzem a complexidade e tornam o modelo mais escalável para imagens de alta resolução.

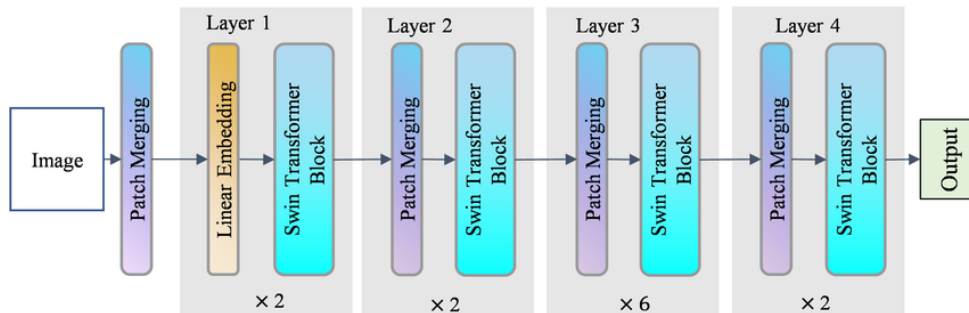


Figura 15 – Fonte: Zhao et al. (2024)

Um diferencial importante da *SwinTr* é o uso de janelas deslocadas entre camadas consecutivas, permitindo a troca de informações entre diferentes regiões da imagem e a modelagem de dependências de longo alcance de forma eficiente. Além disso, sua estrutura hierárquica, com redução progressiva da resolução espacial e aumento da dimensionalidade

das características, o torna compatível com arquiteturas do tipo *encoder–decoder*, sendo amplamente utilizado como *backbone* em tarefas como classificação, detecção e segmentação de imagens (LIU et al., 2021).

## 2.4.12 Otimização de Redes Neurais

A otimização de redes neurais profundas envolve o uso de estratégias que visam melhorar o desempenho do modelo durante o treinamento, buscando maior capacidade de generalização e melhor adaptação aos dados disponíveis. Neste trabalho, são empregadas técnicas de transferência de aprendizado, parada antecipada, validação cruzada e busca sistemática de hiperparâmetros.

### 2.4.12.1 Transferência de Aprendizado

O treinamento de uma rede profunda requer uma grande quantidade de dados, o que representa um desafio significativo para sua obtenção. Quanto menos variação de características a base de dados tiver, maior o risco da rede aprender apenas no cenário específico, ocasionando ineficácia geral do treinamento (HAWTHORNE; ISAACS, 2018).

Esse desafio é encontrado principalmente no contexto de imagens médicas, como o deste trabalho, pois conjuntos de imagens médicas em larga escala e anotados, além de escassos, muitas das vezes possuem muita variedade de informações e poucos grupos de imagens com as mesmas características. Uma forma de amenizar essa dificuldade é a transferência de aprendizado (SHIN et al., 2016). Esta técnica é caracterizada pelo treinamento de uma rede neural profunda em um grande conjunto de dados e, em seguida, realiza-se um ajuste fino do modelo com os dados relativos a situação do estudo (HAWTHORNE; ISAACS, 2018).

A transferência de aprendizado tem como objetivo conseguir maior generalização nas camadas menos específicas, por meio do treinamento em uma enorme base de imagens gerais que possua milhares de imagens e muitas classes, como a *ImageNet* (DENG et al., 2009). Assim, essa técnica melhora a eficácia dos modelos de aprendizado profundo (SHIN et al., 2016; HAWTHORNE; ISAACS, 2018; ZHANG et al., 2024). Neste trabalho, utiliza-se a transferência de aprendizado como inicialização das redes profundas utilizadas para segmentação explicada na Seção 4.4.

### 2.4.12.2 Parada antecipada

A parada antecipada, ou *Early Stopping*, é uma técnica utilizada durante o treinamento de redes neurais profundas com o objetivo de evitar o sobreajuste e melhorar a capacidade de generalização do modelo (PRECHELT, 1998). Essa abordagem consiste

em monitorar o desempenho do modelo em um conjunto de validação ao longo das épocas de treinamento e interromper o processo quando não há melhoria significativa na métrica de interesse após um determinado número de iterações, denominado paciência.

Ao interromper o treinamento antes que o modelo comece a se ajustar excessivamente aos dados de treinamento, esta técnica contribui para obter modelos mais robustos e com melhor desempenho em dados não vistos. Além disso, essa técnica também reduz o custo computacional do treinamento, uma vez que evita a execução desnecessária de épocas adicionais quando o modelo já atingiu seu melhor desempenho (PRECHELT, 2002).

Neste trabalho, esta técnica foi aplicada durante o treinamento das redes profundas utilizadas na tarefa de segmentação, monitorando a métrica de validação ao longo das épocas. O treinamento era interrompido automaticamente quando não eram observadas melhorias na métrica monitorada após um número predefinido de épocas consecutivas.

#### 2.4.12.3 Validação Cruzada

A validação cruzada é uma técnica amplamente utilizada para avaliar a capacidade de generalização de modelos de aprendizado de máquina. Nesse método, o conjunto de dados disponível é dividido em múltiplas partições, chamadas de *folds*. Em cada iteração do processo, uma dessas partições é utilizada como conjunto de validação enquanto as demais são empregadas para o treinamento do modelo. Esse procedimento é repetido até que todas as partições tenham sido utilizadas como conjunto de validação ao menos uma vez (ROBERTS et al., 2017).

A principal vantagem da validação cruzada é fornecer uma estimativa mais robusta do desempenho do modelo, reduzindo a dependência de uma única divisão entre dados de treinamento e validação (ARLOT; CELISSE, 2010). Essa abordagem é particularmente útil em cenários com quantidade limitada de dados, como em aplicações médicas, pois permite aproveitar melhor o conjunto de dados disponível durante o processo de treinamento e avaliação.

Uma das variantes mais utilizadas é o *K-Fold*, no qual o conjunto de dados é dividido em  $K$  partições de tamanho aproximadamente igual. O modelo é treinado  $K$  vezes, cada vez utilizando  $K - 1$  partições para treinamento e a partição restante para validação. O desempenho final é obtido pela média das métricas calculadas em cada uma das  $K$  iterações, acompanhada do respectivo desvio-padrão, que quantifica a variabilidade do modelo entre os *folds* (ARLOT; CELISSE, 2010).

A escolha do valor de  $K$  envolve um compromisso entre viés e variância: valores menores de  $K$  resultam em conjuntos de treinamento menores e estimativas com maior viés, enquanto valores maiores aproximam-se do desempenho real do modelo, porém com maior custo computacional. Na prática,  $K = 5$  ou  $K = 10$  são os valores mais adotados

na literatura (ROBERTS et al., 2017).

#### 2.4.12.4 Busca Em Grade

A busca em grade, ou *Grid Search*, é uma técnica utilizada para seleção de hiperparâmetros em modelos de aprendizado de máquina. Esse método consiste em definir previamente um conjunto de valores possíveis para cada hiperparâmetro e avaliar sistematicamente todas as combinações possíveis entre esses valores. Para cada combinação testada, o modelo é treinado e avaliado utilizando uma estratégia de validação, permitindo identificar a configuração que apresenta melhor desempenho (BERGSTRA; BENGIO, 2012).

Embora o *Grid Search* possa demandar maior custo computacional quando há muitos hiperparâmetros envolvidos, ele é amplamente utilizado por sua simplicidade e por permitir uma exploração estruturada do espaço de busca.

Neste trabalho, essa técnica foi utilizada para identificar configurações adequadas de hiperparâmetros relacionadas ao treinamento das redes neurais, contribuindo para a obtenção de melhores resultados no processo de segmentação.

#### 2.4.13 Ensemble de Redes

O *ensemble* é um paradigma de aprendizado de máquina no qual múltiplos modelos, os classificadores base, são combinados após seu treinamento para produzir uma única predição. Diferentemente dos métodos tradicionais, que buscam aprender apenas uma hipótese a partir dos dados de treinamento, os métodos de *ensemble* constroem um conjunto de hipóteses e combinam seus resultados, geralmente por meio de votação, média ou outro mecanismo de combinação. Essa abordagem tende a melhorar a capacidade de generalização do modelo final, reduzindo erros e aumentando a precisão em comparação com o uso de um único modelo (POLIKAR, 2012).

O uso de diferentes modelos pode capturar padrões distintos nos dados, e a combinação dessas perspectivas pode resultar em um sistema mais robusto. Em muitos casos, modelos individuais relativamente fracos, quando combinados adequadamente, podem formar um modelo forte capaz de realizar previsões mais confiáveis (ZHOU, 2021).

Na literatura, os métodos de *ensemble* são categorizados por diferentes tipos de abordagens. Neste trabalho, adota-se o *Stacking* com Média Ponderada, no qual as predições dos modelos base são utilizadas como entrada para um meta-modelo que aprende a combinar essas saídas de forma otimizada (WOLPERT, 1992).

### 2.4.13.1 *Stacking* com Média Ponderada

O *Stacking* com média ponderada é uma variante do *Stacking* tradicional na qual, em vez de treinar um meta-modelo, a combinação das predições dos modelos base é realizada por meio de uma soma ponderada direta (SHAHOSSEINI; HU; PHAM, 2022):

$$\hat{y} = \sum_{i=1}^N w_i \cdot \hat{y}_i, \quad \text{sujeito a } \sum_{i=1}^N w_i = 1 \quad (2.3)$$

onde  $N$  é o número de modelos,  $\hat{y}_i$  é a predição do  $i$ -ésimo modelo e  $w_i$  é o peso atribuído a ele.

## 2.5 Métricas de Avaliação

Nesta seção serão explicadas as métricas utilizadas para avaliar a eficácia da segmentação obtida pelo método deste trabalho. Em tarefas de segmentação, as métricas avaliam a qualidade geral da classificação de cada *pixel* e a corretude da localização dos objetos. Para isso, as máscaras geradas pela rede neural são comparadas com as máscaras obtidas a partir da anotação realizada pelos especialistas (GTs, do inglês *ground truth*) (MÜLLER; SOTO-REY; KRAMER, 2022).

Dessa forma, é possível avaliar a eficácia do método na segmentação do pâncreas. A comparação das predições com as GTs gera a matriz de confusão. Nessa matriz são apresentadas as quantidades de verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN). VP são *pixels* que foram corretamente segmentados como tendo o pâncreas presente. FP são os *pixels* que não contêm o órgão, mas que foram segmentados como se contivesse. VN são *pixels* que não apresentam o órgão e que não foram segmentados como se o apresentassem. Finalmente, FN são os *pixels* que contêm o pâncreas mas não foram segmentados pelo método (MÜLLER; SOTO-REY; KRAMER, 2022).

A precisão, definida na Equação 2.4, utilizada na tarefa de classificação de fatias, verifica dentre as predições positivas realizadas pelo modelo, quantas estão corretas. Essa métrica é penalizada quando o número de FP é maior que o de VP. No contexto da classificação, a precisão indica a proporção de fatias corretamente classificadas como contendo o pâncreas em relação ao total de fatias preditas como positivas.

A sensibilidade, ou *Recall*, apresentada na Equação 2.5, mede a capacidade do modelo em identificar corretamente as amostras positivas. Na classificação de fatias, o *Recall* indica o quanto o modelo consegue detectar corretamente as imagens que contêm o pâncreas. No contexto da segmentação, essa métrica expressa a capacidade da rede em detectar os pixels pertencentes ao órgão, sendo de extrema importância no contexto

médico, pois valores baixos indicam a presença de regiões do pâncreas não segmentadas.

A acurácia, definida na Equação 2.6, é aplicada exclusivamente à tarefa de classificação e representa a proporção total de predições corretas realizadas pelo modelo. Embora forneça uma visão global do desempenho do classificador, a acurácia pode ser influenciada por desbalanceamentos entre as classes.

A especificidade, apresentada na Equação 2.7, é utilizada apenas na classificação de fatias, mede a capacidade do modelo em identificar corretamente as amostras negativas. Essa métrica corresponde à razão entre os VN e a soma de verdadeiros negativos e FP, indicando o quanto o modelo evita classificar incorretamente fatias sem o pâncreas como positivas.

O *F1-Score*, ou coeficiente de similaridade *Dice*, definido na Equação 2.8, é empregado principalmente na tarefa de segmentação. Essa métrica corresponde à média harmônica entre a precisão (Equação 2.4) e a sensibilidade (Equação 2.5), combinando a capacidade do modelo em segmentar corretamente os *pixels* pertencentes ao pâncreas (VP) com a habilidade de evitar a segmentação excessiva de regiões incorretas (FP). Dessa forma, o *Dice* fornece uma avaliação balanceada entre detecção e formato, sendo uma das métricas mais relevantes para a análise de resultados em segmentação médica.

Além do coeficiente *Dice*, outra métrica utilizada em tarefas de segmentação é o *Intersection over Union* (IoU), também conhecido como índice de *Jaccard*, presente na Equação 2.9. Essa métrica mede a sobreposição entre a região predita pelo modelo e a região real anotada pelos especialistas. O IoU é calculado como a razão entre a interseção das duas regiões e a união entre elas.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.4)$$

$$\text{Sensibilidade (Recall)} = \frac{VP}{VP + FN} \quad (2.5)$$

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.6)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (2.7)$$

$$\text{F1-Score (Dice)} = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (2.8)$$

$$\text{IoU} = \frac{VP}{VP + FP + FN} \quad (2.9)$$

## 2.6 Considerações Finais

Este capítulo apresentou a fundamentação teórica do trabalho desenvolvido, abordando os conceitos necessários para compreender o problema estudado, o método desenvolvido e a base de imagens utilizada. Também foram discutidos temas relacionados ao processamento de imagens, aprendizado profundo e métricas de avaliação empregadas neste estudo. No próximo capítulo será apresentado o método proposto.

## 3 Trabalhos Relacionados

A tarefa de segmentação do pâncreas utilizando aprendizado profundo segue três abordagens principais. A primeira consiste em uma abordagem 3D, processando dados volumétricos onde os *voxels* são as unidades fundamentais. A segunda opera em fatias bidimensionais individuais, utilizando *pixels* como elemento base. A terceira segue um modelo híbrido, processando fatias adjacentes em canais distintos. Além disso, destacam-se três bases de dados, sendo elas a *Medical Segmentation Decathlon* (MSD), a *National Cancer Institute Pancreas-CT Dataset* (NIH), e a *The Cancer Imaging Archive* (TCIA).

Esta seção apresenta uma revisão sistemática dos principais trabalhos relacionados a essas metodologias, destacando seus avanços, limitações e contribuições para o campo.

### 3.1 Segmentação do pâncreas utilizando modelos tridimensionais

Zhao et al. (2019) desenvolveram uma estratégia em duas etapas utilizando redes CNN 3D. Inicialmente, uma *U-Net* tridimensional era aplicada a volumes subsamplados para identificação da região candidata. Em seguida, a segmentação era refinada em recortes de alta resolução por meio de outra *U-Net* 3D. No conjunto NIH, o método atingiu um *Dice* médio de 85,99%. Entretanto, a abordagem demanda elevada capacidade computacional e depende de uma correta delimitação da região candidata, sob o risco de exclusão parcial da glândula.

Tian et al. (2021) propuseram um método baseado em *Markov Chain Monte Carlo* (MCMC) e CNNs para determinar a localização da região pancreática antes da segmentação. Eles empregaram uma *U-Net* 3D e testaram seu método no conjunto de dados NIH. Na fase de pré-processamento, eles usaram superpixels para aprimorar a extração de recursos. Sua abordagem alcançou uma pontuação *Dice* de 78,13% no conjunto de dados de teste.

Zhu et al. (2023) introduziram um *framework* de controle de qualidade de imagem combinado com um algoritmo de segmentação para pâncreas. Eles usaram uma *U-Net* 3D para segmentar o pâncreas, avaliando sua abordagem nos conjuntos de dados MSD e NIH. O melhor resultado foi um *Dice* de 75,43% no conjunto de dados NIH.

Por fim, Tondji et al. (2025) propuseram um modelo de segmentação do pâncreas baseado em *U-Net* 3D com mecanismo de atenção, projetado para aprimorar a capacidade da rede em destacar regiões anatômicas relevantes em cenários de baixo contraste e alta complexidade estrutural. A abordagem foi avaliada em dois conjuntos de dados de referência amplamente utilizados, o NIH e o MSD. No conjunto MSD, o método alcançou um *Dice* médio de 78,1%.

## 3.2 Segmentação do pâncreas utilizando abordagem bidimensional

Em 2017, Yu et al. (2017) desenvolveram a *Recurrent Saliency Transformation Network* (RSTN), um método iterativo de segmentação *coarse-to-fine*. A estratégia baseia-se na geração de uma segmentação grosseira inicial, seguida por refinamentos sucessivos por meio de um módulo de transformação de saliência que repondera intensidades com base em máscaras anteriores. Essa abordagem foi realizada na base de dados NIH e alcançou *Dice* médio de aproximadamente 82,73%.

Cai et al. (2019) propuseram um modelo bidimensional recorrente, no qual cada fatia era processada por uma rede convolucional 2D, seguida de uma camada *Long Short-Term Memory* (LSTM) responsável por impor coerência entre as fatias consecutivas. Essa abordagem, aplicada na base de dados NIH, atingiu *Dice* médio de 76,1%. Apesar de integrar dependências sequenciais, o método ainda não explora de forma plena a informação tridimensional do volume.

A rede *Deep U-Net*, que foi introduzida por Nishio, Noguchi e Fujimoto (2020), investiga uma versão mais profunda da *U-Net*, incorporando diferentes estratégias de *data augmentation*, como *mixup*, para melhorar a robustez do modelo. A abordagem foi aplicada fatia a fatia, com posterior empilhamento para reconstrução do volume. Os resultados mostraram valores de *Dice* entre 70,30% e 78,90%. Ainda assim, as melhorias observadas foram modestas e dependentes tanto da qualidade do conjunto de dados quanto das técnicas de aumento utilizadas.

Kurnaz et al. (2024) propuseram a *Pascal U-Net* para segmentação do pâncreas. Ao contrário da *U-Net* tradicional, a *Pascal U-Net* estabelece conexões por meio de camadas convolucionais em vez de links diretos entre o codificador e o decodificador. Seu modelo foi avaliado no conjunto de dados TCIA, atingindo uma pontuação *Dice* média de 71,35%.

Juwita, Hassan e Datta (2025) propuseram a MOMUNet (*MobileNetV2 + U-Net*), uma rede leve para segmentação do pâncreas em TC, combinada com uma estratégia de *External Contour Cropping* (ECC) para mitigar o desbalanceamento entre fundo e pâncreas. Avaliada nas bases de dados NIH e MSD, a MOMUNet com ECC obteve 75,29% como pontuação média de *Dice* na base de dados MSD.

## 3.3 Segmentação do pâncreas utilizando abordagem híbrida

Zheng et al. (2020) desenvolveram uma rede 2.5D com módulo de correlação de fatias e pré-treinamento *self-supervised* para melhorar a comunicação entre fatias adjacentes na segmentação do pâncreas em TC. O método foi validado na base de dados NIH utilizando validação cruzada de quatro dobras, alcançando *Dice* médio de aproximadamente 86,21%, com alta sensibilidade e especificidade, demonstrando robustez na segmentação mesmo

diante da variabilidade anatômica.

Li et al. (2021) propuseram um método híbrido que integra abordagens baseadas em atlas e em rede neural. Seu estudo usou a base de dados NIH, com binarização aplicada entre  $-100HU$  e  $240HU$ . Além disso, eles avaliaram o método na base de dados MSD. A entrada consistiu em imagens 2.5D, processadas por uma *Fully Convolutional Network* (FCN). A pontuação média *Dice* alcançada foi de 71,36%.

Ferrara et al. (2024) propuseram o uso de Redes Neurais Convolucionais Profundas (DCNNs) como uma etapa de pré-processamento para segmentar a região pancreática antes de identificar tumores malignos. Eles treinaram seu modelo usando o conjunto de dados MSD, adotando uma abordagem 2.5D com uma arquitetura *Segmentation-reSegmentation Net* (SrSNet). A pontuação *Dice* média foi 66,8%.

He et al. (2025) introduziram um modelo baseado em U-Net aprimorado com múltiplos estágios de atenção, projetado para destacar características discriminativas ao longo das diferentes camadas da rede. A estratégia, implementada em uma configuração 2.5D por meio do uso de fatias adjacentes, alcançou *Dice* de 87,16% na base de dados NIH. O método apresentou bom equilíbrio entre acurácia e eficiência, embora sua dependência do contexto inter-fatia limite a captura completa das relações volumétricas.

### 3.4 Considerações finais

Nesta seção, são apresentadas as principais considerações a partir da análise dos trabalhos relacionados e dos resultados obtidos por eles. Busca-se sintetizar os achados mais relevantes, destacando os avanços e limitações das abordagens estudadas.

A análise dos trabalhos relacionados evidencia que métodos tridimensionais, como os propostos por Tian et al. (2021) e Zhao et al. (2019), tendem a apresentar maior acurácia, em virtude do uso completo da informação volumétrica. No entanto, seu elevado custo computacional limita sua aplicação prática em cenários com restrições operacionais.

Por outro lado, métodos bidimensionais oferecem maior eficiência computacional, mas apresentam desempenho inferior devido à ausência de contexto espacial entre fatias. Abordagens híbridas, como as de Li et al. (2021), Zheng et al. (2020) e He et al. (2025), surgem como alternativa intermediária, buscando equilibrar desempenho e eficiência, com resultados competitivos.

Adicionalmente, observa-se que a maioria dos trabalhos concentra esforços na melhoria das arquiteturas de redes neurais, como observado em Zhu et al. (2023) e Ferrara et al. (2024), enquanto poucos exploram estratégias voltadas à redução da complexidade do problema antes da etapa de segmentação, como filtragem de fatias e delimitação de regiões de interesse.

A Tabela 1 apresenta um resumo comparativo dos principais trabalhos analisados, destacando suas abordagens, bases de dados utilizadas e desempenho reportado.

Tabela 1 – Resumo comparativo dos trabalhos relacionados em segmentação pancreática

Abordagem	Trabalho	Base de dados	Dice
<b>3D</b>	Zhao et al. (2019)	NIH	85,99%
	Tian et al. (2021)	NIH	78,13%
	Zhu et al. (2023)	NIH, MSD	75,43%
	Tondji et al. (2025)	MSD	78,1%
<b>2D</b>	Yu et al. (2017)	NIH	82,73%
	Cai et al. (2019)	NIH	76,1%
	Nishio, Noguchi e Fujimoto (2020)	NIH	78,9%
	Kurnaz et al. (2024)	TCIA	71,35%
	Juwita, Hassan e Datta (2025)	NIH, MSD	75,29%
<b>Híbrida</b>	Zheng et al. (2020)	NIH	86,21%
	Li et al. (2021)	NIH, MSD	71,36%
	Ferrara et al. (2024)	MSD	66,8%
	He et al. (2025)	NIH	87,16%

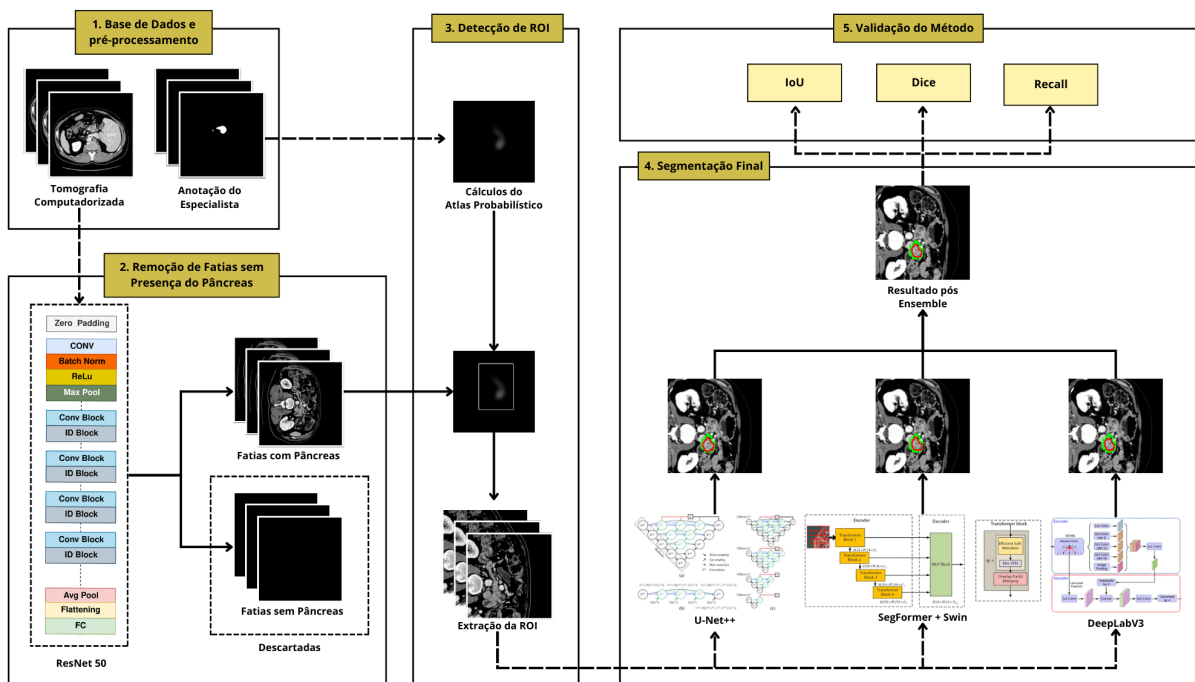
Nesse contexto, observa-se que os trabalhos relacionados concentram-se majoritariamente na melhoria do coeficiente *Dice* por meio de arquiteturas 2D, 3D ou híbridas, com destaque para abordagens 3D e híbridas que alcançam maiores valores de segmentação. No entanto, há uma lacuna quanto à integração estruturada de estratégias complementares, como o uso de ensemble de modelos e etapas explícitas de redução de falsos positivos, bem como na preocupação com a eficiência computacional.

Assim, o método proposto neste trabalho diferencia-se ao combinar técnicas de pré-processamento, seleção de fatias relevantes e utilização de atlas probabilístico, com o objetivo de restringir o espaço de busca, reduzir a incidência de falsos positivos e diminuir o custo computacional, sem comprometer a qualidade da segmentação e adicionando mais confiabilidade ao estudo.

## 4 Material e Método proposto

Este capítulo apresenta o método proposto para a segmentação automática do pâncreas em imagens de TC. A abordagem foi estruturada em etapas sequenciais, com o objetivo de reduzir a complexidade do problema e melhorar a eficiência computacional sem comprometer a qualidade da segmentação. O método é ilustrado na Figura 16 e consiste em quatro etapas.

Figura 16 – Fluxograma da metodologia proposta.



Fonte: Elaborado pelo autor.

Primeiro, realiza-se um pré-processamento da base de dados transformando o volume em um conjunto de fatias por paciente. Em seguida, utiliza-se uma rede neural de classificação para realizar o filtro de fatias que contêm o órgão em estudo, diminuindo assim a quantidade de falsos positivos. Depois, ocorre o processamento de imagens para a extração da ROI das imagens, para diminuir a área a ser processada pelas redes e pelo ensemble. Após o corte da imagem, utiliza-se arquiteturas CNN e *Transformers* para a tarefa de segmentar o pâncreas, para, em seguida, realizar o *ensemble* e melhorar o resultado da segmentação. Por fim, realiza-se a etapa de avaliação dos resultados por meio de métricas consolidadas na literatura.

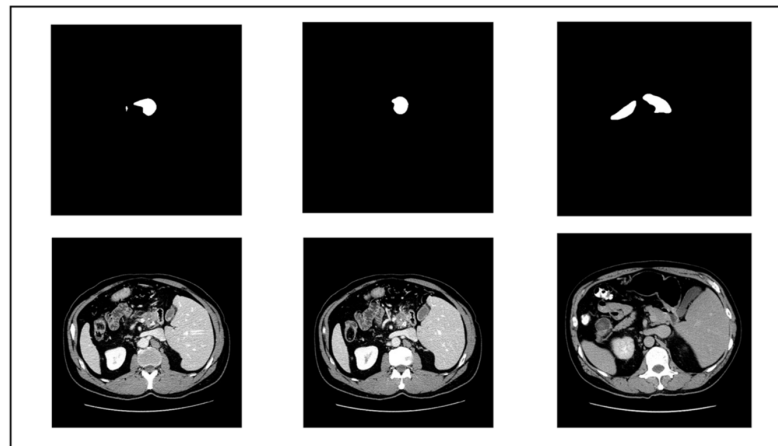
## 4.1 Base de dados e pré-processamento

A base de dados utilizada por este estudo é a MSD ([ANTONELLI et al., 2022](#)), que contém 420 volumes de TC abdominais coletados pelo *Memorial Sloan Kettering Cancer Center*.

Cada volume tem aproximadamente 100 fatias (imagens 2D) com resolução de  $512 \times 512$ . O parênquima pancreático foi anotado manualmente em cada fatia por um radiologista usando o aplicativo *Scout*.

Exemplos de fatias da base de dados podem ser vistos na Figura 17, onde as anotações do pâncreas variam entre uma única região ou dois objetos distintos, dependendo da fatia.

Figura 17 – Exemplos da base de dados.



Fonte: Elaborado pelo autor.

Antes de enviar as imagens para a CNN de classificação, os volumes de TC passaram por um processo de pré-processamento em duas etapas. Na primeira etapa, os valores de intensidade dos volumes foram limitados, utilizando a técnica de janelamento (Seção 2.3.1). O intervalo selecionado foi de  $[-150, 250]$  HU, pois melhora a visualização das estruturas abdominais ao excluir regiões de ar e estruturas altamente densas, como ossos ([FERNANDES et al., 2023](#)).

## 4.2 Remoção de Fatias sem Presença do Pâncreas

Nesta etapa, as fatias por paciente são reduzidas, retirando do conjunto aquelas que não incluem o pâncreas. Utiliza-se uma CNN treinada para classificação binária, onde a classe positiva corresponde a uma fatia com o pâncreas e a classe negativa a uma sem o pâncreas. Esta tarefa é crucial, pois a maioria das fatias não contém o órgão, o que pode

beneficiar os resultados para uma classe e comprometer a confiabilidade da segmentação final.

Para esta tarefa, foram avaliadas quatro redes de classificação, onde foi selecionado um modelo de cada rede. As redes escolhidas foram a *ResNet50* (Seção 2.4.10.1), a *EfficientNet-B5* (Seção 2.4.10.2), a ViT (Seção 2.4.11) e a *SwinTR* (Seção 2.4.11.2), todos ajustados usando pesos pré-treinados do *ImageNet* (DENG et al., 2009). Para avaliar o desempenho do treinamento, empregou-se a validação cruzada (Seção 2.4.12.3), com o objetivo de identificar o modelo estatisticamente mais adequado para a realização do filtro.

Após a etapa de validação, o modelo selecionado demonstrou capacidade de filtrar fatias que não continham o pâncreas. No entanto, devido à natureza probabilística do classificador, ainda podem ocorrer erros do tipo falso negativo, nos quais fatias contendo o órgão são incorretamente descartadas.

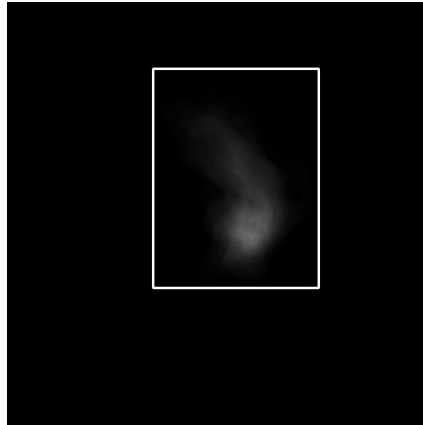
Para mitigar esse problema, foi aplicado um pós-processamento baseado na continuidade anatômica do pâncreas ao longo do volume. Especificamente, para cada sequência de fatias, identifica-se a primeira e a última fatia classificadas como positivas, e todas as fatias intermediárias nesse intervalo passam a ser consideradas como positivas. Essa estratégia permite recuperar regiões que foram indevidamente descartadas, reduzindo a incidência de falsos negativos e preservando a coerência anatômica da segmentação.

### 4.3 Detecção de ROI

Após a seleção das fatias contendo o pâncreas, foi realizada a delimitação da ROI com o objetivo de reduzir o espaço de busca dos modelos de segmentação.

Para essa finalidade, foi utilizado um Atlas Probabilístico (Seção 2.3.3), construído a partir da média das máscaras de treinamento. Esse atlas fornece uma representação estatística da localização do pâncreas no conjunto de dados. Na Figura 18 é possível ver uma representação probabilística do atlas.

Figura 18 – Representação probabilística do atlas.

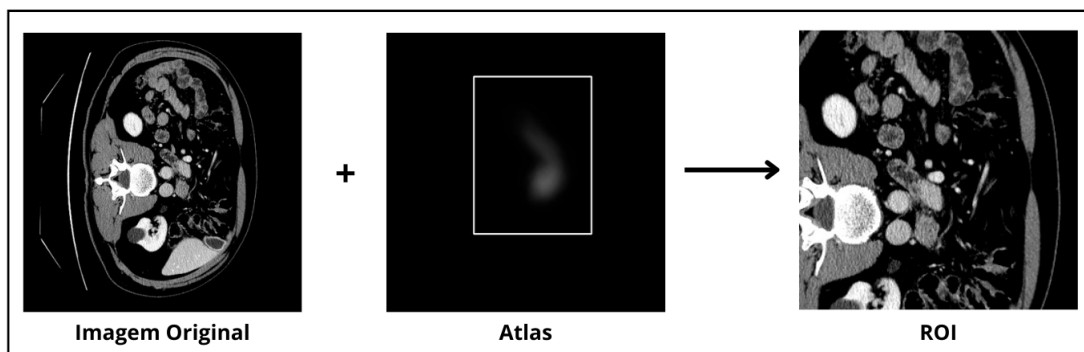


Fonte: Elaborado pelo autor.

Com base no atlas probabilístico, foi definida uma *bounding box* responsável por delimitar a região com maior probabilidade de ocorrência do pâncreas. Para isso, inicialmente identifica-se, ao longo do eixo horizontal, o primeiro pixel à esquerda e o último pixel à direita com probabilidade relevante de pertencimento ao órgão. A partir desses limites, é adicionada uma margem de segurança, com o objetivo de compensar variações anatômicas e possíveis imprecisões do atlas. Em seguida, a região delimitada é ajustada para dimensões compatíveis com a entrada da rede neural, garantindo que altura e largura sejam múltiplos de 16.

Dessa forma, as imagens são recortadas preservando apenas a região de interesse (ROI), mantendo consistência dimensional entre as amostras e reduzindo áreas irrelevantes para o processo de segmentação. A Figura 19 apresenta a aplicação do atlas para obtenção da ROI.

Figura 19 – Extração da ROI.



Fonte: Elaborado pelo autor.

Esse procedimento contribui para reduzir a quantidade de informações irrelevantes presentes nas imagens originais, direcionando o modelo de segmentação para uma região

mais específica onde o pâncreas provavelmente está localizado. Dessa forma, a utilização do Atlas Probabilístico auxilia na diminuição do espaço de busca do modelo, podendo melhorar a eficiência computacional e a qualidade da segmentação ao focar apenas nas regiões mais relevantes da imagem.

## 4.4 Segmentação Final

Após filtrar as fatias sem o pâncreas e detectar a ROI usando o atlas probabilístico, o próximo passo é a segmentação do pâncreas. Para isso, foram treinadas três redes neurais, sendo elas a *U-Net++* (Seção 2.4.10.4), a *DeepLabV3* (Seção 2.4.10.5) e a *SegFormer* (Seção 2.4.11.1). Após o treinamento individual, foi aplicada uma estratégia de *ensemble* baseada em *stacking* com média ponderada (Seção 2.4.13), combinando as previsões dos modelos para gerar a segmentação final.

Todas as redes foram treinadas de forma independente sobre as regiões de interesse previamente delimitadas, permitindo uma comparação justa entre os modelos e fornecendo as previsões necessárias para a etapa posterior de combinação *ensemble* (ABDELRAHMAN; VIRIRI, 2023).

Para que o *ensemble* apresente um desempenho superior ao dos modelos individuais, é fundamental definir adequadamente o peso atribuído a cada rede no processo de combinação (NADERALVOJOD; HERNANDEZ-BOUSSARD, 2024). Dessa forma, foi aplicada uma estratégia de busca baseada em *Grid Search* para avaliar diferentes combinações de pesos no ensemble, priorizando aquelas que resultassem no melhor desempenho de segmentação. Para garantir a interpretabilidade e estabilidade da combinação, foi imposta a restrição de que a soma dos pesos fosse igual a 1. Assim, os pesos foram gerados aleatoriamente dentro de um intervalo de variação de 0,5, sob a condição de normalização, isto é, assegurando que a soma entre eles fosse sempre igual a 1. Cada combinação válida foi então avaliada, permitindo identificar a configuração que melhor equilibra a contribuição individual de cada modelo.

## 4.5 Validação do Método

Para validar o método, são adotadas as métricas padrão de imagens médicas, descritas na Seção 2.5. Na etapa de filtragem de fatias, Acurácia, Especificidade, Precisão, Sensibilidade e *F1-Score* avaliam o desempenho do modelo na detecção de fatias com o pâncreas, minimizando a quantidade de fatias da classe negativa.

Para segmentação, o *Dice* mede a sobreposição entre a máscara prevista e a marcação do especialista, enquanto o *Recall* avalia se a maior parte do pâncreas foi segmentada

corretamente, reduzindo o risco de regiões críticas ausentes. Além disso, a métrica IoU é utilizada para avaliar a similaridade entre a segmentação predita e a anotação de referência, mensurando a proporção entre a interseção e a união das regiões segmentadas.

## 4.6 Considerações Finais

Este capítulo apresentou a metodologia proposta para a segmentação do pâncreas em imagens de TC. O método inclui etapas de pré-processamento das imagens, classificação de fatias para identificar a presença do órgão, segmentação utilizando redes neurais profundas e avaliação dos resultados por meio de métricas de desempenho. No capítulo seguinte, são apresentados os experimentos realizados e a análise dos resultados obtidos com a aplicação do método proposto.

## 5 Resultados

Este capítulo apresenta os resultados obtidos com a aplicação do método proposto para a segmentação do pâncreas em imagens de TC. Inicialmente, são discutidos os resultados da etapa de classificação de fatias, seguida da avaliação da segmentação e, por fim, os resultados são comparados com os trabalhos relacionados discutidos no Capítulo 3 e, em seguida, estudos de caso são apresentados para uma análise qualitativa.

### 5.1 Ambiente de Treinamento

Os experimentos foram conduzidos usando o ambiente de computação em nuvem Google Colab Pro, que fornece acesso a uma GPU Tesla T4 com 12 GB de RAM. A implementação foi escrita em Python, utilizando como bibliotecas principais *TensorFlow*, *PyTorch*, *NumPy*, *OpenCV* e *Matplotlib*.

### 5.2 Experimentos

Esta seção descreve os diferentes cenários experimentais projetados para validar cada etapa do método proposto.

Para a validação do método, utiliza-se a base de dados descrita na Seção 4.1, onde 420 volumes foram divididos em 80% para treinamento e 20% para teste, garantindo que fatias do mesmo paciente não fossem compartilhadas entre os conjuntos, evitando assim vazamento de dados. Todas as técnicas foram aplicadas apenas ao conjunto de treinamento, enquanto o conjunto de teste foi usado exclusivamente para avaliação do método proposto.

#### 5.2.1 Experimento 1: Seleção de Rede para a Filtragem de Fatias

Nesta etapa, quatro arquiteturas, descritas na Seção 4.2, foram avaliadas para a tarefa de classificar se as fatias continham o pâncreas ou não.

Todos os modelos foram inicializados com os pesos pré-treinados na ImageNet (DENG et al., 2009) para, em seguida, ocorrer o treinamento em 30 épocas usando o otimizador *Adam* com uma taxa de aprendizado de 0,0001, um tamanho de *batch* de 16 e a função de perda *Binary Cross-Entropy* (BCE). Para encontrar os hiperparâmetros ótimos, aplica-se um método de otimização baseado em *Grid Search*, que explora sistematicamente um conjunto predefinido de valores no otimizador (*Adam* e *AdamW*), a taxa de aprendizado (0,01, 0,001 e 0,0001) e o tamanho do lote (8, 12 e 16), afim de encontrar o melhor *F1-Score* entre eles.

Para avaliar o desempenho completo dos modelos, foi realizado um método de validação cruzada, *K-Fold* (Seção 2.4.12.3). Os resultados individuais de cada *Fold* por modelo são apresentados nas Tabelas 2 a 5, e um comparativo consolidado entre todas as arquiteturas é exibido na Tabela 6.

Tabela 2 – Resultados da validação cruzada para a *ResNet50*.

Fold	Sensibilidade(%)	Precisão(%)	<i>F1-Score</i> (%)	Especificidade(%)	Acurácia(%)
1	91,40	94,53	91,65	97,36	95,20
2	92,39	93,73	91,22	96,64	95,25
3	92,22	95,58	92,10	97,17	95,28
4	92,16	94,94	91,66	96,79	95,11
5	91,24	92,73	89,79	96,13	94,64
<b>Média</b>	<b>91,88 ± 0,53</b>	<b>94,30 ± 1,10</b>	<b>91,29 ± 0,89</b>	<b>96,82 ± 0,48</b>	<b>95,09 ± 0,26</b>

A Tabela 2 apresenta os resultados da *ResNet50* (Seção 2.4.10.1). O modelo demonstrou desempenho consistente ao longo dos *folds*, com *F1-Score* médio de 91,29% ± 0,89 e elevada especificidade média de 96,82% ± 0,48, indicando boa capacidade de rejeitar fatias sem o pâncreas.

Tabela 3 – Resultados da validação cruzada para a *EfficientNetB3*.

Fold	Sensibilidade(%)	Precisão(%)	<i>F1-Score</i> (%)	Especificidade(%)	Acurácia(%)
1	97,30	93,06	95,04	95,72	96,29
2	96,95	94,28	95,40	96,70	96,78
3	95,35	94,56	94,93	96,58	96,11
4	95,27	95,43	95,23	97,12	96,45
5	96,53	93,11	94,49	96,30	96,37
<b>Média</b>	<b>96,28 ± 0,93</b>	<b>94,09 ± 1,01</b>	<b>95,02 ± 0,34</b>	<b>96,48 ± 0,52</b>	<b>96,40 ± 0,25</b>

A Tabela 3 exibe os resultados obtidos pela *EfficientNetB3* (Seção 2.4.10.2). Esta arquitetura alcançou o melhor desempenho geral entre os modelos avaliados, com *F1-Score* médio de 95,02% ± 0,34, sendo o maior valor e, ao mesmo tempo, o menor desvio-padrão entre todos os modelos, evidenciando não apenas superioridade em desempenho, mas também alta estabilidade e confiabilidade ao longo dos *folds*.

Tabela 4 – Resultados da validação cruzada para a ViT.

Fold	Sensibilidade(%)	Precisão(%)	<i>F1-Score</i> (%)	Especificidade(%)	Acurácia(%)
1	85,14	47,77	56,96	36,58	54,19
2	84,36	72,63	73,62	78,90	80,70
3	92,91	66,88	76,18	65,03	75,69
4	81,22	75,21	73,54	78,10	79,23
5	85,48	70,49	76,15	81,92	83,01
<b>Média</b>	<b>85,82 ± 4,30</b>	<b>66,60 ± 10,96</b>	<b>71,29 ± 8,11</b>	<b>68,11 ± 18,78</b>	<b>74,56 ± 11,69</b>

A Tabela 4 apresenta os resultados da ViT (Seção 2.4.11). Dentre os quatro modelos avaliados, a ViT obteve o pior desempenho, com *F1-Score* médio de apenas 71,29% ±

8,11. Além do baixo valor absoluto, o elevado desvio-padrão em todas as métricas indica instabilidade do modelo entre os *folds*, sugerindo que a arquitetura baseada puramente em atenção global pode não ser adequada para esta tarefa com o volume de dados disponível.

Tabela 5 – Resultados da validação cruzada para a *SwinTr*.

Fold	Sensibilidade(%)	Precisão(%)	<i>F1-Score</i> (%)	Especificidade(%)	Acurácia(%)
1	97,38	92,18	94,64	95,19	95,99
2	94,93	93,06	93,84	96,45	95,95
3	97,31	90,86	93,58	92,78	94,51
4	95,37	95,88	95,50	97,64	96,81
5	98,69	88,54	93,08	93,86	95,33
<b>Média</b>	<b>96,74 ± 1,56</b>	<b>92,11 ± 2,71</b>	<b>94,13 ± 0,95</b>	<b>95,18 ± 1,95</b>	<b>95,72 ± 0,86</b>

Por fim, a Tabela 5 apresenta os resultados da *SwinTr* (Seção 2.4.11.2). O modelo obteve *F1-Score* médio de 94,13% ± 0,95, posicionando-se como a segunda melhor arquitetura avaliada. A *SwinTr* destacou-se pela maior sensibilidade média entre todos os modelos (96,74% ± 1,56), indicando boa capacidade de identificar corretamente as fatias que contêm o pâncreas. Contudo, sua precisão média (92,11% ± 2,71) foi a mais baixa entre os modelos com bom desempenho geral, refletindo uma taxa relativamente maior de falsos positivos em comparação à *EfficientNet*.

Tabela 6 – Comparativo de desempenho médio entre os modelos avaliados na etapa de filtragem de fatias. O melhor resultado por métrica está destacado.

Modelo	Sensibilidade(%)	Precisão(%)	<i>F1-Score</i> (%)	Especificidade(%)	Acurácia(%)
ViT	85,82 ± 4,30	66,60 ± 10,96	71,29 ± 8,11	68,11 ± 18,78	74,56 ± 11,69
<i>ResNet50</i>	91,88 ± 0,53	94,30 ± 1,10	91,29 ± 0,89	96,82 ± 0,48	95,09 ± 0,26
<i>SwinTr</i>	<b>96,74 ± 1,56</b>	92,11 ± 2,71	94,13 ± 0,95	95,18 ± 1,95	95,72 ± 0,86
<b><i>EfficientNet</i></b>	96,28 ± 0,93	<b>94,09 ± 1,01</b>	<b>95,02 ± 0,34</b>	<b>96,48 ± 0,52</b>	<b>96,40 ± 0,25</b>

A *EfficientNet* (Seção 2.4.10.2) destacou-se como a melhor arquitetura para a tarefa de filtragem de fatias, alcançando o maior *F1-Score* médio (95,02% ± 0,34) e o menor desvio-padrão entre os modelos, indicando não apenas alto desempenho, mas também consistência entre os *folds*. Inicialmente, a binarização das saídas do classificador foi realizada utilizando um limiar padrão de 0,5.

Embora a redução do limiar tenda a aumentar a taxa de falsos positivos, esse efeito é menos crítico no contexto do método proposto, uma vez que essas fatias adicionais podem ser posteriormente refinadas ou descartadas na etapa de pós processamento. Dessa forma, a escolha da *EfficientNet* também se justifica por seu bom equilíbrio entre sensibilidade e precisão, além de sua robustez frente a possíveis ajustes no limiar de decisão.

## 5.2.2 Experimento 2: Validação da Filtragem de Fatias

Após a seleção da *EfficientNet* como modelo de filtragem, foi avaliado o impacto do pós-processamento descrito na Seção 4.2, que consiste em considerar como positivas as

fatias situadas entre duas fatias classificadas como contendo o pâncreas. Essa etapa visa corrigir falsos negativos gerados pelo classificador, aproveitando a continuidade anatômica do órgão no volume tridimensional.

Para quantificar o ganho obtido, o modelo foi avaliado em um *fold* do conjunto de validação antes e após a aplicação do pós-processamento. Os resultados são apresentados na Tabela 7.

Tabela 7 – Impacto do pós-processamento no desempenho da filtragem de fatias.

Configuração	Sensibilidade(%)	Precisão(%)	F1-Score(%)	Especificidade(%)	Acurácia(%)
Sem pós-processamento	95,43	93,63	94,46	96,15	95,88
<b>Com pós-processamento</b>	<b>98,07</b>	<b>94,26</b>	<b>95,99</b>	<b>96,36</b>	<b>96,96</b>

A aplicação do pós-processamento melhorou o resultado em todas as métricas avaliadas. O ganho mais expressivo foi observado na sensibilidade, que passou de 95,43% para 98,07%, evidenciando a eficácia da estratégia na recuperação de fatias com pâncreas incorretamente descartadas pelo classificador. O *F1-Score* também apresentou melhoria, passando de 94,46% para 95,99%, confirmando que o ganho em sensibilidade não foi acompanhado de queda relevante na precisão. Esses resultados validam o uso do pós-processamento como etapa complementar à classificação, tornando a filtragem de fatias mais robusta e confiável para as etapas subsequentes do método.

### 5.2.3 Experimento 3: Segmentação Com e Sem Detecção de ROI

Após avaliar os filtros das fatias, o próximo passo é validar a eficácia da etapa de detecção da ROI (Seção 4.3). Para avaliar seu impacto, conduziu-se um experimento avaliando o desempenho da segmentação com e sem essa etapa, usando a rede *U-Net++* (Seção 2.4.10.4) com o *backbone ResNet50* (Seção 2.4.10.1).

Para a comparação dos experimentos, utilizam-se as seguintes configurações: um tamanho de lote de 16, função de perda BCE, otimizador *AdamW* e uma taxa de aprendizado de 0,0003. Os resultados estão apresentados na Tabela 8.

Tabela 8 – Desempenho de segmentação em porcentagem com e sem detecção de ROI.

Filtragem	Detecção de ROI	CNN	Dice médio (%)	IoU médio (%)	Recall médio (%)
Sim	Não	U-Net++	62,32	53,21	54,06
Sim	Sim	U-Net++	74,75	62,15	78,24

Após a aplicação do atlas, o desempenho teve uma melhoria de 15% nas métricas de *Dice* e *Recall*. Isso destaca a importância da etapa de detecção da ROI na filtragem de estruturas irrelevantes que podem induzir a geração de falsos positivos. Ao focar apenas na região do pâncreas, o processo de segmentação se torna mais preciso, reduzindo a interferência de estruturas anatômicas e informações não relevantes.

### 5.2.4 Experimento 4: Avaliação de diferentes arquiteturas de *backbone*

Neste experimento, foi realizada a avaliação de diferentes arquiteturas de *backbone* utilizadas no processo de extração de características das imagens.

O objetivo foi analisar o impacto de diferentes redes convolucionais na qualidade da segmentação do pâncreas. Para isso, foram testadas quatro arquiteturas amplamente utilizadas na literatura e apresentadas na Seção 4.4.

Todas as configurações foram treinadas sob as mesmas condições experimentais apresentadas no experimento anterior (Seção 5.2.3), garantindo uma comparação justa entre os modelos.

Para avaliar o desempenho dos backbones nas redes CNN, foi-se utilizada a U-Net++. Os resultados obtidos em termos das métricas *Dice* médio, *Recall* médio e IoU médio são apresentados na Tabela 9.

Tabela 9 – Comparação de segmentação em porcentagem com diferentes *backbones* convolucionais.

<b>Backbone</b>	<b><i>Dice</i> Médio (%)</b>	<b><i>Recall</i> Médio (%)</b>	<b>IoU Médio (%)</b>
<i>EfficientNetB3</i>	73,76	72,77	61,98
<i>ResNet50</i>	75,58	74,36	63,37
<i>EfficientNetB5</i>	75,18	75,65	63,31
<b><i>ResNet101</i></b>	<b>75,68</b>	<b>78,64</b>	<b>63,42</b>

Observa-se que as arquiteturas baseadas em *ResNet* apresentaram desempenho consistente, com destaque para a *ResNet101*, que obteve o maior valor médio de *Dice* (75,68%) e o maior IoU médio (63,42%). Esses resultados indicam que arquiteturas mais profundas podem capturar representações mais discriminativas das estruturas pancreáticas nas imagens.

Entre os modelos da família *EfficientNet*, a arquitetura *EfficientNetB5* apresentou desempenho superior em relação à *EfficientNetB3*, alcançando um *Dice* médio de 75,18%, *Recall* médio de 75,65% e IoU médio de 63,31%. Esse comportamento sugere que versões mais profundas dessa arquitetura também contribuem para melhorar a extração de características relevantes para a tarefa de segmentação.

Adicionalmente, foram avaliadas duas arquiteturas baseadas em *Transformer* para a tarefa de segmentação utilizadas como *backbone* para a rede de segmentação *SegFormer*, apresentada na Seção 2.4.11.1. A Tabela 10 apresenta os resultados obtidos pela *SwinTr* e pelo *ViT*.

Entre os modelos baseados em *Transformer*, a *SwinTr* obteve desempenho superior em todas as métricas avaliadas, alcançando *Dice* médio de 77,46%, *Recall* médio de 84,11% e IoU médio de 66,63%, superando inclusive os melhores resultados obtidos pelas

Tabela 10 – Comparação de segmentação em porcentagem com diferentes *backbones* baseados em *Transformer*.

<b>Backbone</b>	<b>Dice Médio (%)</b>	<b>Recall Médio (%)</b>	<b>IoU Médio (%)</b>
ViT	73,91	72,12	62,07
<b>SwinTr</b>	<b>77,46</b>	<b>84,11</b>	<b>66,63</b>

arquiteturas convolucionais. O *ViT*, por sua vez, apresentou desempenho inferior, com *Dice* médio de 73,91%, sugerindo que a modelagem de dependências locais e hierárquicas da *SwinTr* é mais adequada para a tarefa de segmentação do pâncreas do que a atenção global irrestrita do *ViT*.

De forma geral, os resultados indicam que a escolha do *backbone* influencia diretamente o desempenho da segmentação, sendo que arquiteturas mais profundas tendem a apresentar melhor capacidade de representação.

## 5.2.5 Experimento 5: Treinamento das Redes de Segmentação

Após verificar o desempenho dos *backbones*, as arquiteturas de segmentação semântica, descritas na Seção 4.4, foram avaliadas individualmente. Todas as redes foram treinadas utilizando apenas as fatias previamente filtradas pela etapa de classificação e após o recorte da ROI, assegurando que o processo de segmentação fosse aplicado exclusivamente a imagens contendo o pâncreas e com o mínimo de informações relevantes possíveis.

As arquiteturas foram treinadas sob o mesmo protocolo experimental apresentado na Seção 5.2.3, incluindo divisão dos dados, função de perda e critérios de avaliação, de modo a garantir uma comparação justa entre os modelos. A avaliação foi realizada no conjunto de teste utilizando as métricas *Dice*, *IoU* e *Recall*, apresentadas na Seção 2.5. Os resultados quantitativos são apresentados na Tabela 11.

Tabela 11 – Desempenho dos modelos individuais e do *ensemble* no conjunto de teste.

<b>Modelo</b>	<b>Dice(%)</b>	<b>IoU(%)</b>	<b>Recall(%)</b>
<i>DeepLabV3 + ResNet101</i>	75,70	64,73	81,61
<i>UNet++ + ResNet101</i>	75,94	65,19	82,46
<i>SegFormer + SwinTr</i>	77,46	66,63	84,11

Observa-se que o modelo *SegFormer + SwinTr* obteve o melhor desempenho individual, alcançando coeficiente *Dice* de 77,46%, *IoU* de 66,63% e *Recall* de 84,11%. Esse resultado evidencia a contribuição do modelo baseado em *Transformer*, capaz de capturar características de longo alcance, além de modelar melhor o contexto global da imagem. As arquiteturas *UNet++* e *DeepLabV3* também apresentaram resultados competitivos, porém, com desempenho um pouco inferior.

### 5.2.6 Experimento 6: Aplicação do *Ensemble*

No sexto experimento, foi aplicada uma estratégia de *ensemble* (Seção 2.4.13), combinando as predições das três arquiteturas de segmentação treinadas individualmente. O objetivo dessa etapa foi explorar a complementaridade entre modelos convolucionais e baseados em *Transformer*, reduzindo erros individuais e aumentando a robustez do método. O resultado foi obtido após a aplicação de uma busca em grade para otimização dos pesos de cada rede, onde os três melhores resultados estão presentes na Tabela 12.

Tabela 12 – Melhores distribuições de pesos do *ensemble* e seus respectivos desempenhos.

Rank	DeepLabV3	UNet++	SegFormer	Dice (%)	Recall (%)	IoU (%)
1º	0,2	0,3	0,5	<b>78,55</b>	84,41	<b>68,17</b>
2º	0,15	0,25	0,6	78,08	86,22	67,37
3º	0,10	0,3	0,6	78,08	<b>86,28</b>	67,37

A estratégia de *ensemble* resultou em uma melhoria consistente em todas as métricas avaliadas, atingindo, respectivamente, *Dice* e *IoU* médios de 78,55%, 68,17%, superando os resultados obtidos por qualquer modelo individualmente. Destaca-se que as melhores taxas de *Recall* foram observadas nas combinações em que a *U-Net++* recebeu peso relativo maior que a *DeepLabV3*, indicando sua maior capacidade de identificar regiões relevantes.

De forma geral, a utilização do *ensemble* permite explorar a complementaridade entre arquiteturas convolucionais e modelos baseados em *Transformer* resultando em ganhos consistentes tanto em desempenho quanto em robustez. Assim, a técnica se mostra uma estratégia promissora para tarefas de segmentação médica, especialmente em cenários complexos, como a segmentação do pâncreas em exames de tomografia computadorizada.

## 5.3 Comparação com trabalhos relacionados

Nesta seção, o desempenho do nosso método com abordagens existentes na literatura é comparado e exibido na Tabela 13. A avaliação é baseada na abordagem (3D, 2.5D e 2D) e no *Dice*, destacando as vantagens e limitações da nossa abordagem.

Em relação aos trabalhos descritos na Tabela 13, nosso método 2D proposto alcançou um *Dice* médio de 78,55% na base de dados MSD, representando um resultado competitivo dentro das abordagens 2D presentes na literatura. Embora métodos 2,5D e 3D, como o de (HE et al., 2025) (87,16%) e (ZHAO et al., 2019) (85,99%), tenham obtido pontuações superiores, é importante destacar que tais arquiteturas geralmente demandam maior custo computacional devido à sua capacidade de capturar informações espaciais tridimensionais.

Além disso, é importante ressaltar que as comparações entre os trabalhos devem ser analisadas com cautela devido às diferenças nas bases de dados utilizadas. Enquanto

Tabela 13 – Comparação de pontuações de *Dice* para diferentes abordagens.

Autores	Base de dados	Abordagem	Técnica	Dice (%)
Zhao et al. (2019)	NIH	3D	CNN 3D	<b>85,99</b>
Tian et al. (2021)	NIH	3D	3D + MCMC <i>U-Net</i>	78,13
Zhu et al. (2023)	NIH, MSD	3D	3D <i>U-Net</i>	75,43
Tondji et al. (2025)	MSD	3D	<i>U-net</i> 3D + Atenção	78,10
Yu et al. (2017)	NIH	2D	RSTN	<b>82,73</b>
Cai et al. (2019)	NIH	2D	CNN	76,10
Nishio, Noguchi e Fujimoto (2020)	NIH	2D	<i>Deep U-net</i>	<b>78,90</b>
Kurnaz et al. (2024)	TCIA	2D	Pascal <i>U-Net</i>	71,35
Juwita, Hassan e Datta (2025)	NIH, MSD	2D	<i>MOMUNet</i> + ECC	75,29
Zheng et al. (2020)	NIH	2.5D	CNN + contexto espacial	86,21
Li et al. (2021)	NIH, MSD	2.5D	<i>FCN</i>	71,36
Ferrara et al. (2024)	MSD	2.5D	SRSNet	66,80
He et al. (2025)	NIH	2.5D	<i>U-net</i> + Atenção	<b>87,16</b>
<b>Método proposto</b>	<b>MSD</b>	<b>2D</b>	<b>Filtragem de fatias + Atlas + Ensemble</b>	<b>78,55</b>

diversos trabalhos reportam resultados na base NIH ou em combinações entre NIH e MSD, o método proposto foi avaliado exclusivamente na base MSD, que apresenta características distintas em termos de variabilidade anatômica, contraste e qualidade das imagens, o que pode impactar diretamente o desempenho dos modelos.

Por fim, o *ensemble* permitiu agregar predições de múltiplos modelos, aumentando a robustez e a estabilidade da segmentação final. Embora as comparações diretas sejam desafiadoras devido às diferenças nos conjuntos de dados, os resultados demonstram a viabilidade e a eficácia de um *pipeline* de segmentação 2D bem otimizado para a segmentação do pâncreas.

### 5.3.1 Estudo de Caso

Para uma melhor avaliação do método proposto, serão apresentados estudos de caso que fornecem uma análise qualitativa dos resultados da segmentação. Esses estudos de caso ilustram como o método funciona em diferentes cenários e destacam sua capacidade de segmentar com precisão o pâncreas em imagens de TC médicas.

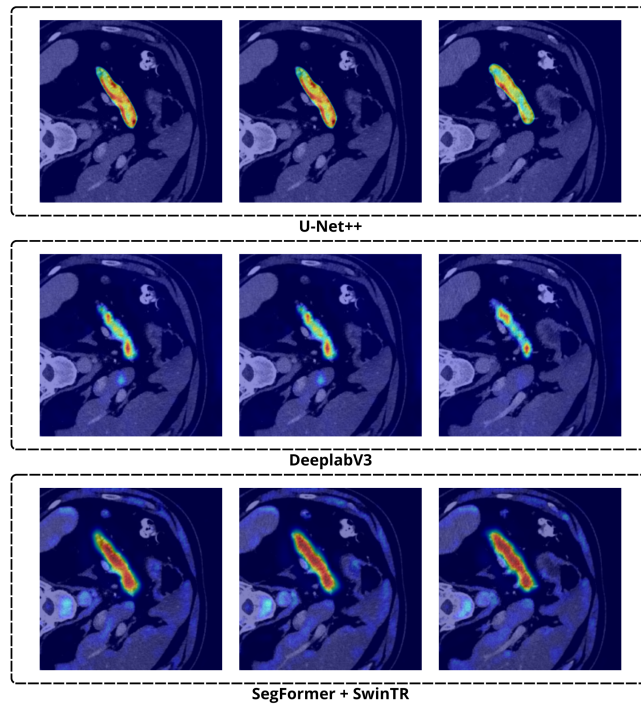
#### 5.3.1.1 Estudo de Caso 1

Inicialmente são apresentados os mapas de calor obtidos através de uma técnica de *Grad-Cam* (Seção 2.4.9) aplicada aos modelos utilizados no Experimento 5 (Seção 5.2.5). Os resultados obtidos estão na Figura 20.

Observa-se na Figura 20 que todas as arquiteturas conseguem direcionar a maior parte de sua atenção para a região do pâncreas, destacando corretamente a estrutura anatômica de interesse. Esse comportamento indica que os modelos conseguiram aprender características relevantes da região, contribuindo para uma segmentação mais precisa.

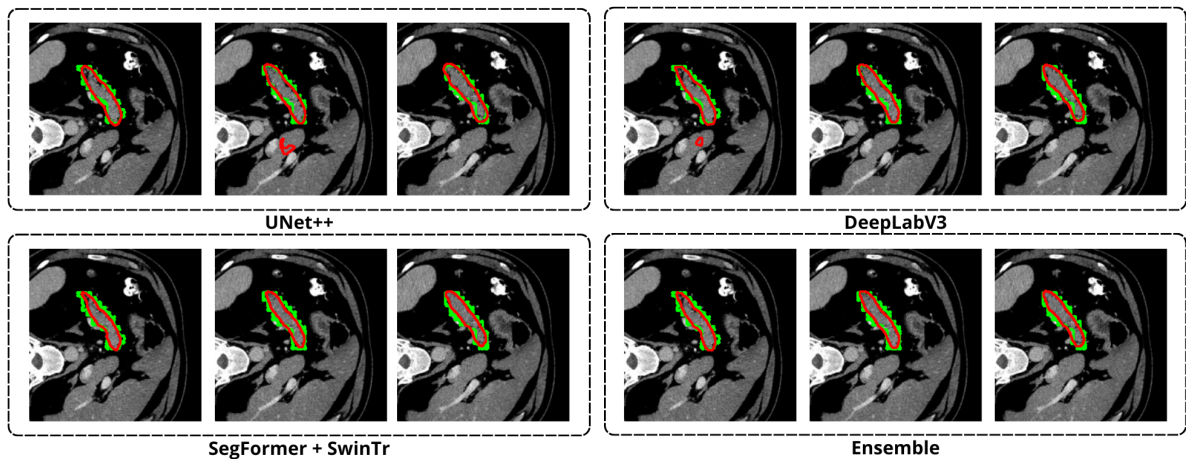
A Figura 21 mostra os melhores casos do conjunto de dados de teste, onde, nas fatias, a anotação do especialista está em verde e a segmentação está em vermelho.

Figura 20 – Mapa de calor do Cenário Favorável.



Fonte: Elaborado pelo autor.

Figura 21 – Segmentação do Cenário Favorável.



Fonte: Elaborado pelo autor.

Observa-se uma boa sobreposição entre as regiões segmentadas automaticamente e as anotações manuais, indicando que o método proposto é capaz de capturar adequadamente a forma e a localização do pâncreas mesmo em diferentes variações anatômicas e níveis de contraste presentes nas imagens de TC.

Comparando o mapa de calor com a segmentação final, percebe-se que algumas

redes, como a *U-Net++* e a *DeepLabV3*, cometeram erros ao considerar determinadas regiões da imagem que não correspondem ao pâncreas como se fossem parte do órgão. Esse comportamento pode ser observado nos mapas de calor, onde certas ativações aparecem em áreas que não pertencem à região anotada pelo especialista.

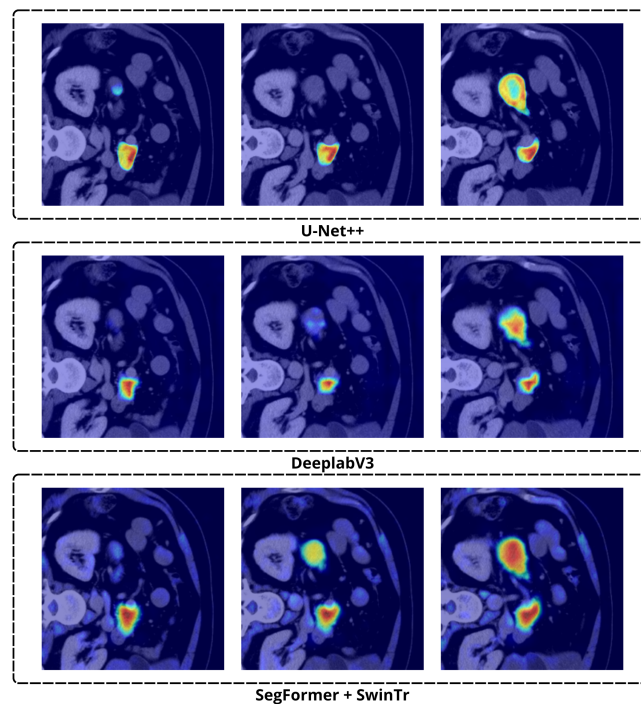
Entretanto, ao analisar o resultado final do ensemble, observa-se que esses erros não foram propagados para a segmentação final. Isso ocorre porque a combinação das predições das diferentes redes permite reduzir falhas individuais, priorizando regiões que foram corretamente identificadas pela maioria dos modelos.

Dessa forma, regiões que foram corretamente segmentadas por algumas redes acabam sendo reforçadas no resultado final, enquanto ativações equivocadas presentes em apenas um ou poucos modelos tendem a ser ignoradas. Esse comportamento evidencia a capacidade do ensemble de reduzir erros individuais e produzir segmentações mais consistentes.

### 5.3.1.2 Estudo de Caso 2

De forma semelhante ao cenário anterior, a Figura 22 apresenta os mapas de calor das redes para um caso mais desafiador do conjunto de teste.

Figura 22 – Mapa de calor do Cenário Desafiador.



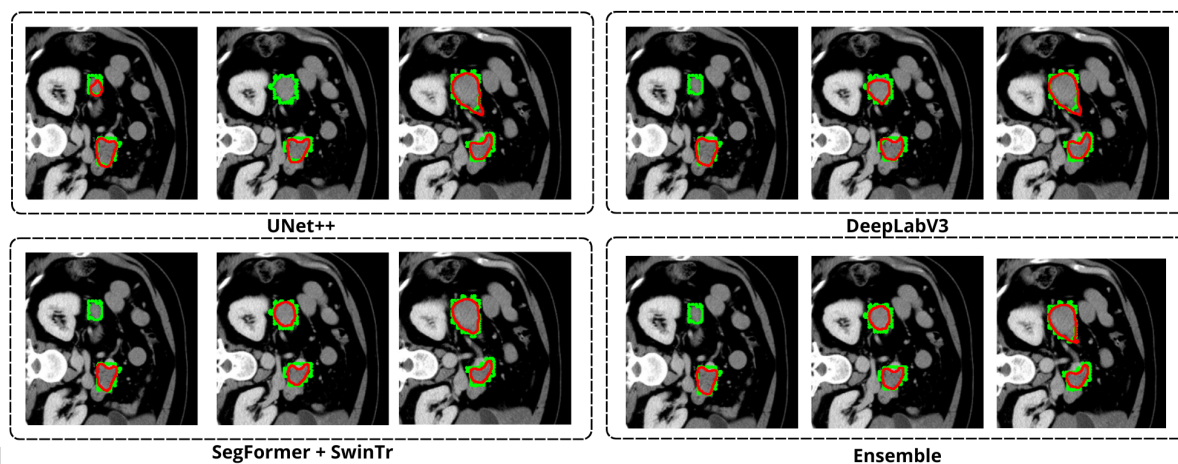
Fonte: Elaborado pelo autor.

Analisando a Figura 22, é possível perceber que algumas redes não conseguem focar em todas as regiões que correspondem ao pâncreas. Isso ocorre porque determinadas

estruturas próximas apresentam características visuais semelhantes, o que pode confundir o modelo durante o processo de segmentação. Além disso, esse comportamento também pode estar relacionado à aplicação do atlas probabilístico, uma vez que eventuais imprecisões na definição da região de interesse podem levar à exclusão parcial de áreas relevantes ou à inclusão de regiões adjacentes, impactando a capacidade do modelo de identificar corretamente todo o órgão.

Como consequência, os modelos deixam de apresentar ativações em regiões próximas, o que acaba influenciando o resultado final da segmentação, como pode ser visto na Figura 23.

Figura 23 – Segmentação do Cenário Desafiador.



Fonte: Elaborado pelo autor.

A Figura 23 apresenta um caso desafiador observado no conjunto de teste. Nessa situação, algumas redes apresentam dificuldades para delimitar corretamente as regiões do pâncreas, resultando em segmentações com erros mais evidentes quando comparadas às anotações do especialista.

Assim, em cenários mais desafiadores, como o apresentado na Figura 23, a combinação dos modelos não é suficiente para corrigir completamente as falhas presentes nas segmentações individuais. Mesmo assim, o ensemble ainda tende a apresentar resultados mais estáveis quando comparado às redes avaliadas separadamente.

## 5.4 Considerações Finais

Neste capítulo, foram apresentados os resultados obtidos a partir dos experimentos realizados para avaliar o método proposto para segmentação do pâncreas em imagens de TC. Inicialmente, foi descrito o ambiente experimental utilizado, seguido pela apresentação dos diferentes experimentos projetados para validar cada etapa do *pipeline* proposto.

## 6 Conclusão

O pâncreas é um órgão de grande importância clínica, sendo frequentemente analisado em exames de TC para o diagnóstico e acompanhamento de diversas doenças, incluindo pancreatite e câncer pancreático. A segmentação precisa desse órgão é essencial para aplicações como análise quantitativa, planejamento de tratamentos e desenvolvimento de sistemas de diagnóstico auxiliado por computador. Entretanto, a segmentação do pâncreas em exames de TC constitui uma tarefa desafiadora, mesmo para especialistas, devido à grande variabilidade anatômica do órgão, ao baixo contraste entre tecidos adjacentes e à presença de estruturas próximas com características visuais semelhantes. Nesse contexto, métodos computacionais baseados em aprendizado profundo têm sido amplamente investigados como forma de auxiliar especialistas na análise de imagens médicas.

Os resultados experimentais demonstraram que as arquiteturas baseadas em aprendizado profundo são capazes de identificar o pâncreas de forma consistente nas imagens de TC. Entretanto, os modelos individuais ainda apresentaram limitações, principalmente ao falhar na segmentação de regiões completas do órgão ou ao incluir áreas vizinhas com características semelhantes. Essas dificuldades estão diretamente relacionadas à complexidade anatômica do pâncreas e à grande variabilidade presente nos exames de TC.

Para lidar com essas limitações, foi proposta uma estratégia de combinação de modelos por meio de um método de *ensemble*. Essa abordagem permitiu explorar as características complementares das diferentes arquiteturas, resultando em segmentações mais robustas e estáveis. Os resultados obtidos demonstraram que o *ensemble* apresentou desempenho superior em comparação aos modelos individuais, alcançando melhores valores nas métricas de avaliação utilizadas.

Entretanto, mesmo após a aplicação do *ensemble*, o estudo ainda apresentou limitações relevantes. A técnica empregada não é capaz de agregar novas informações ao processo de aprendizado das redes, atuando principalmente na melhoria da consistência e do formato das máscaras segmentadas. Dessa forma, o método não contribui para a identificação de regiões que não foram previamente detectadas pelos modelos individuais. Além disso, a distribuição de pesos no *ensemble* exerce influência significativa no resultado final da segmentação. Em situações em que um modelo apresenta melhor desempenho em determinada amostra, mas possui um peso reduzido na combinação, há a possibilidade de degradação do resultado final para essa imagem.

Outra limitação enfrentada foi a necessidade de elevado poder computacional, o que ocasionou demora na execução dos experimentos e dificultou a realização de uma

validação estatística mais abrangente dos modelos de segmentação. O treinamento de redes neurais profundas requer recursos de hardware robustos, como unidades de processamento gráfico (GPUs, do inglês *Graphics Processing Unit*) de alta capacidade, e o tempo gasto em cada experimento restringiu o número de configurações e arquiteturas que puderam ser avaliadas dentro do escopo deste trabalho. Ainda assim, é importante destacar que o custo computacional associado à abordagem proposta, baseada em modelos 2D, permanece inferior ao observado em métodos que utilizam redes 3D, os quais demandam maior consumo de memória e maior tempo de processamento devido ao tratamento volumétrico das imagens.

De forma geral, os resultados obtidos indicam que a abordagem proposta apresenta potencial para auxiliar especialistas na análise de exames de TC abdominal. Considerando que a segmentação manual do pâncreas é uma tarefa demorada e sujeita a variações entre observadores, a utilização de métodos automáticos pode contribuir para aumentar a eficiência do processo de análise e apoiar o desenvolvimento de sistemas de diagnóstico auxiliado por computador.

## 6.1 Trabalhos Futuros

Apesar dos resultados promissores obtidos neste trabalho, diversas possibilidades podem ser exploradas para aprimorar o método proposto e ampliar as contribuições deste estudo.

Como perspectiva de trabalhos futuros, sugere-se a investigação de técnicas de ensemble mais sofisticadas, capazes de realizar a fusão das predições de forma adaptativa e orientada ao contexto da imagem. Abordagens que considerem mecanismos de atenção, seleção dinâmica de modelos ou aprendizado baseado em meta-modelos podem permitir uma combinação mais eficiente das saídas, contribuindo não apenas para o refinamento das máscaras, mas também para a identificação de regiões não segmentadas pelos modelos individuais. Além disso, a exploração de estratégias que ajustem automaticamente os pesos do ensemble de acordo com o desempenho local de cada rede pode representar um avanço significativo na qualidade das segmentações obtidas.

Outra direção promissora envolve a exploração de modificações arquiteturais nos modelos utilizados. A avaliação de novas arquiteturas baseadas em *transformers*, bem como a combinação de blocos de convolução tradicionais com mecanismos de atenção, pode contribuir para capturar de forma mais eficaz as características espaciais presentes nas imagens de TC.

Por fim, futuros trabalhos podem explorar técnicas avançadas de processamento de imagens com o objetivo de melhorar a qualidade visual dos exames, como métodos de realce de contraste e redução de ruído. Essas técnicas podem contribuir para aumentar a

visibilidade das estruturas anatômicas e facilitar o aprendizado dos modelos de segmentação.

## 6.2 Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, da Fundação de Amparo à Pesquisa do Maranhão (FAPEMA) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Além disso, reconhece-se o uso de *Large Language Models* (LLM) para verificação ortográfica, correção gramatical e auxílio na tradução de termos específicos.

## 6.3 Produções Científicas

Durante o período do estudo, os seguintes trabalhos foram publicados ou submetidos:

- ***Tooth Detection and Numbering in Panoramic Radiographs Using YOLOv8-Based Approach.***
  - *Conferência:* MobiHealth 2023
  - *Status:* Publicado em 28 de junho de 2024
  - *Papel:* Primeiro autor e correspondente
- ***Improving Colorectal Cancer Diagnosis Using MIRNet and InceptionV3 on Histopathological Images.***
  - *Conferência:* BRACIS 2024
  - *Status:* Publicado em 30 de janeiro de 2025
  - *Papel:* Coautor
- ***Atlas-Guided U-Net++ with EfficientNetB5 for Automatic Pancreas Segmentation in Abdominal CT Scans.***
  - *Conferência:* SBCAS 2025
  - *Status:* Publicado em 09 de junho de 2025
  - *Papel:* Primeiro autor e correspondente
- ***MultiSegUS: Sistema Web para Segmentação Multimodelo de Lesões Mamárias em Ultrassonografia.***
  - *Conferência:* SBCAS 2026
  - *Status:* Submetido em 09 de março de 2026

- *Papel*: Coautor
- ***Ensemble* Ponderado de Deep Learning para Segmentação de Lesões Mamárias em Ultrassonografia.**
  - *Conferência*: SBCAS 2026
  - *Status*: Submetido em 09 de março de 2026
  - *Papel*: Coautor
- **Segmentação do Pâncreas em Tomografia Computadorizada com *Ensemble* Guiado por Atlas e *Backbone Swin Transformer*.**
  - *Conferência*: SBCAS 2026
  - *Status*: Submetido em 09 de março de 2026
  - *Papel*: Primeiro autor

## Referências

- ABDALLA, M.; EL-GINDY, A. et al. Chest x-ray pneumothorax segmentation using u-net with efficientnet and resnet architectures. *PeerJ Computer Science*, PeerJ Inc., v. 7, p. e607, 2021. Citado na página 34.
- ABDELRAHMAN, L.; VIRIRI, S. EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images. *Frontiers in Computer Science*, Frontiers, v. 5, p. 1235622, 2023. Citado na página 52.
- AKULA, R.; GARIBAY, I. Interpretable multi-head self-attention architecture for sarcasm detection in social media. *Entropy*, v. 23, n. 4, 2021. ISSN 1099-4300. Disponível em: <<https://www.mdpi.com/1099-4300/23/4/394>>. Citado na página 30.
- ANTONELLI, M.; REINKE, A.; BAKAS, S.; FARAHANI, K.; KOPP-SCHNEIDER, A.; LANDMAN, B. A.; LITJENS, G.; MENZE, B.; RONNEBERGER, O.; SUMMERS, R. M. et al. The medical segmentation decathlon. *Nature communications*, Nature Publishing Group UK London, v. 13, n. 1, p. 4128, 2022. Citado na página 49.
- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 2010. Citado na página 39.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of machine learning research*, v. 13, n. 2, 2012. Citado na página 40.
- BOCKMAN, D. E. Anatomy of the pancreas. *The exocrine pancreas: biology, pathobiology and disease*. Raven Press, New York, p. 1–8, 1993. Citado na página 21.
- BRAY, F.; LAVERSANNE, M.; SUNG, H.; FERLAY, J.; SIEGEL, R. L.; SOERJOMATARAM, I.; JEMAL, A. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, Wiley Online Library, v. 74, n. 3, p. 229–263, 2024. Citado na página 17.
- BUZUG, T. M. Computed tomography. In: *Springer handbook of medical technology*. [S.l.]: Springer, 2011. p. 311–342. Citado 2 vezes nas páginas 17 e 22.
- CAI, J.; LU, L.; XING, F.; YANG, L. Pancreas segmentation in ct and mri via task-specific network design and recurrent neural contextual learning. In: *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*. [S.l.]: Springer, 2019. p. 3–21. Citado 3 vezes nas páginas 45, 47 e 61.
- CANTATORE, A.; MÜLLER, P. Introduction to computed tomography. *Springer handbook of medical technology*, DTU Mechanical Engineering, 2011. Citado na página 23.
- CAO, H.; WANG, Y.; CHEN, J.; JIANG, D.; ZHANG, X.; TIAN, Q.; WANG, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In: *SPRINGER. European conference on computer vision*. [S.l.], 2022. p. 205–218. Citado na página 26.
- CASAMITJANA, A.; MANCINI, M.; ROBINSON, E.; PETER, L.; ANNUNZIATA, R.; ALTHONAYAN, J.; CRAMPSIE, S.; BLACKBURN, E.; BILLOT, B.; ATZENI, A. et al.

A probabilistic histological atlas of the human brain for mri segmentation. *Nature*, Nature Publishing Group UK London, p. 1–8, 2025. Citado na página 25.

CHEN, L.-C.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. Citado na página 35.

CONCI, A.; AZEVEDO, E. *Computação Gráfica*. [S.l.]: Campus, 1998. Citado na página 24.

DANIEL, N.; FARINELLA, R.; CHATZIIOANNOU, A. C.; JENAB, M.; MAYÉN, A.-L.; RIZZATO, C.; BELLUOMINI, F.; CANZIAN, F.; TAVANTI, A.; KESKI-RAHKONEN, P. et al. Genetically predicted gut bacteria, circulating bacteria-associated metabolites and pancreatic ductal adenocarcinoma: a mendelian randomisation study. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 25144, 2024. Citado na página 17.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 248–255. Citado 3 vezes nas páginas 38, 50 e 54.

DINIZ, J. O.; FERREIRA, J. L.; SILVA, G. L. da; QUINTANILHA, D. B.; SILVA, A. C.; PAIVA, A. Segmentação de coração em tomografias computadorizadas utilizando atlas probabilístico e redes neurais convolucionais. In: SBC. *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2021. p. 83–94. Citado 2 vezes nas páginas 17 e 25.

DINIZ, J. O.; QUINTANILHA, D. B.; NETO, A. C. S.; SILVA, G. L. da; FERREIRA, J. L.; NETTO, S. M.; ARAUJO, J. D.; CRUZ, L. B. D.; SILVA, T. F.; MARTINS, C. M. da S. et al. Segmentation and quantification of covid-19 infections in ct using pulmonary vessels extraction and deep learning. *Multimedia Tools and Applications*, Springer, v. 80, n. 19, p. 29367–29399, 2021. Citado na página 17.

DOLENŠEK, J.; POHOREC, V.; RUPNIK, M. S.; STOŽER, A. Pancreas physiology. *Challenges in pancreatic pathology*, IntechOpen London, UK, 2017. Citado na página 22.

DOSOVITSKIY, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. Citado 3 vezes nas páginas 30, 35 e 36.

FERNANDES, A. G. S.; JUNIOR, G. B.; DINIZ, J. O. B.; SILVA, A. C.; MATOS, C. E. F. Efficientdeeplab for automated trachea segmentation on medical images. In: SPRINGER. *Brazilian Conference on Intelligent Systems*. [S.l.], 2023. p. 154–166. Citado 2 vezes nas páginas 17 e 49.

FERRARA, N.; ANDRIA, G.; SCARPETTA, M.; LANZOLLA, A. M. L.; ATTIVISSIMO, F.; NISIO, A. D.; RAMOS, D. 2d and 2.5 d pancreas and tumor segmentation in heterogeneous ct images of pdac patients. In: IEEE. *2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. [S.l.], 2024. p. 1–5. Citado 3 vezes nas páginas 46, 47 e 61.

FLECK, L.; TAVARES, M. H. F.; EYNG, E.; HELMANN, A. C.; ANDRADE, M. d. M. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, v. 1, n. 13, p. 47–57, 2016. Citado na página 26.

- FRANTZ, E.; SOUZA-MELLO, V.; LACERDA, C. Mandarim-de. Pancreas: Anatomy, diseases and health implications. *Pancreas Anatomy, Diseases and Health Implications*, v. 45, n. September, p. 1–52, 2012. Citado na página 22.
- FURTADO, M. I. V. Redes neurais artificiais: Uma abordagem para sala de aula. *Ponta Grossa, PR. Atena Editora*, v. 19, 2019. Citado na página 26.
- FUSCO, J.; EL-GOHARY, Y.; GITTES, G. K. Pancreas embryology, anatomy, and physiology. In: *Endocrine Surgery in Children*. [S.l.]: Springer, 2017. p. 143–160. Citado na página 22.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. [S.l.]: Editora Blucher, 2000. Citado na página 24.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. *Deep learning*. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado 2 vezes nas páginas 25 e 26.
- HAWTHORNE, J.; ISAACS, Y. Fine-tuning fine-tuning. *Knowledge, belief, and god: New insights in religious epistemology*, Oxford University Press Oxford, p. 136–168, 2018. Citado na página 38.
- HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2001. Citado na página 27.
- 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*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 31.
- HE, P.; KONG, Q. q.; CHEN, Y.; SHAO, C. B.; SU, Z. Pancreas segmentation based on multi-stage attention enhanced u-net. *International Journal of Imaging Systems and Technology*, Wiley Online Library, v. 35, n. 1, p. e70025, 2025. Citado 4 vezes nas páginas 46, 47, 60 e 61.
- HENDRYCKS, D. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. Citado na página 36.
- HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. *science*, American Association for the Advancement of Science, v. 313, n. 5786, p. 504–507, 2006. Citado na página 28.
- JUNIOR, D. A. D.; CRUZ, L. B. da; DINIZ, J. O. Classificação da camada lipídica do filme lacrimal usando k-means e deep learning. In: *SBC. Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. [S.l.], 2024. p. 1–12. Citado na página 17.
- JUWITA, J.; HASSAN, G. M.; DATTA, A. Pancreas segmentation in ct scans: A novel momunet based workflow. *Computers in Biology and Medicine*, Elsevier, v. 193, p. 110346, 2025. Citado 3 vezes nas páginas 45, 47 e 61.
- KURNAZ, E.; CEYLAN, R.; BOZKURT, M. A.; CEBECI, H.; KOPLAY, M. A novel deep learning model for pancreas segmentation: Pascal u-net. *Inteligencia Artificial*, v. 27, n. 74, p. 22–36, 2024. Citado 3 vezes nas páginas 45, 47 e 61.
- LAKSHMANAN, V.; GÖRNER, M.; GILLARD, R. *Practical machine learning for computer vision*. [S.l.]: "O'Reilly Media, Inc.", 2021. Citado na página 25.

- 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. Citado 2 vezes nas páginas 26 e 28.
- LEE, H.; KIM, Y. S.; KIM, M.; LEE, Y. Low-cost network scheduling of 3d-cnn processing for embedded action recognition. *IEEE Access*, Institute of Electrical and Electronics Engineers (IEEE), v. 9, p. 83901–83912, 2021. Citado na página 18.
- LI, J.; LIN, X.; CHE, H.; LI, H.; QIAN, X. Pancreas segmentation with probabilistic map guided bi-directional recurrent unet. *Physics in Medicine & Biology*, IOP Publishing, v. 66, n. 11, p. 115010, 2021. Citado 3 vezes nas páginas 46, 47 e 61.
- LIU, Z.; LIN, Y.; CAO, Y.; HU, H.; WEI, Y.; ZHANG, Z.; LIN, S.; GUO, B. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2021. p. 10012–10022. Citado 2 vezes nas páginas 37 e 38.
- LONGNECKER, D. S.; THOMPSON, E. D. Anatomy, histology, and fine structure of the pancreas. *The pancreas: an integrated textbook of basic science, medicine, and surgery*, Wiley Online Library, p. 9–22, 2023. Citado na página 21.
- MAHADEVAN, V. Anatomy of the pancreas and spleen. *Surgery (Oxford)*, Elsevier, v. 37, n. 6, p. 297–301, 2019. Citado na página 21.
- MAZONAKIS, M.; DAMILAKIS, J. Computed tomography: What and how does it measure? *European journal of radiology*, Elsevier, v. 85, n. 8, p. 1499–1504, 2016. Citado na página 23.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 27.
- MÜLLER, D.; SOTO-REY, I.; KRAMER, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, Springer, v. 15, n. 1, p. 210, 2022. Citado na página 41.
- NADERALVOJOD, B.; HERNANDEZ-BOUSSARD, T. Improving machine learning with ensemble learning on observational healthcare data. In: *AMIA Annual Symposium Proceedings*. [S.l.: s.n.], 2024. v. 2023, p. 521. Citado na página 52.
- NETO, C. M. S.; SILVA, A. L.; PESSOA, A. C.; QUINTANILHA, D. B.; ALMEIDA, J. D. de; JUNIOR, G. B.; DINIZ, J. O. Diagnóstico de tuberculose em imagens de radiografia utilizando cvt. In: *SBC. Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. [S.l.], 2024. p. 342–353. Citado na página 24.
- NISHIO, M.; NOGUCHI, S.; FUJIMOTO, K. Automatic pancreas segmentation using coarse-scaled 2d model of deep learning: usefulness of data augmentation and deep u-net. *Applied Sciences*, MDPI, v. 10, n. 10, p. 3360, 2020. Citado 3 vezes nas páginas 45, 47 e 61.
- PAN, X.; JIAO, K.; LI, X.; FENG, L.; TIAN, Y.; WU, L.; ZHANG, P.; WANG, K.; CHEN, S.; YANG, B. et al. Artificial intelligence-based tools with automated segmentation and measurement on ct images to assist accurate and fast diagnosis in acute pancreatitis.

*British Journal of Radiology*, Oxford University Press, v. 97, n. 1159, p. 1268–1277, 2024. Citado 2 vezes nas páginas 24 e 26.

PEDRAYES, O.; LEMA, D.; GARCIA, F. D.; USAMENTIAGA, R.; ALONSO, A. Evaluation of semantic segmentation methods for land use with spectral imaging using sentinel-2 and pnoa imagery. *Remote Sensing*, v. 13, p. 2292, 06 2021. Citado na página 35.

PIRES, P. B.; SANTOS, J. D.; PEREIRA, I. V. Artificial neural networks: history and state of the art. *Encyclopedia of Information Science and Technology, Sixth Edition*, IGI Global, p. 1–25, 2025. Citado na página 25.

POLIKAR, R. Ensemble learning. In: *Ensemble machine learning*. [S.l.]: Springer, 2012. p. 1–34. Citado na página 40.

PRECHELT, L. Automatic early stopping using cross validation: quantifying the criteria. *Neural networks*, Elsevier, v. 11, n. 4, p. 761–767, 1998. Citado na página 38.

PRECHELT, L. Early stopping-but when? In: *Neural Networks: Tricks of the trade*. [S.l.]: Springer, 2002. p. 55–69. Citado na página 39.

RAUBER, T. W. Redes neurais artificiais. *Universidade Federal do Espírito Santo*, v. 29, p. 39, 2005. Citado 3 vezes nas páginas 25, 26 e 27.

ROBERTS, D. R.; BAHN, V.; CIUTI, S.; BOYCE, M. S.; ELITH, J.; GUILLERA-ARROITA, G.; HAUENSTEIN, S.; LAHOZ-MONFORT, J. J.; SCHRÖDER, B.; THUILLER, W. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, Wiley Online Library, v. 40, n. 8, p. 913–929, 2017. Citado 2 vezes nas páginas 39 e 40.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241. Citado na página 33.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado na página 27.

RUMELHART, D. E.; HINTON, G. E.; MCCLELLAND, J. L. et al. A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*, Cambridge, MA: MIT Press, v. 1, n. 45-76, p. 26, 1986. Citado na página 28.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986. Citado na página 28.

SAMEI, E.; PELC, B. J. *Computed tomography*. [S.l.]: Springer, 2020. Citado 2 vezes nas páginas 17 e 22.

- 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*. [S.l.: s.n.], 2017. p. 618–626. Citado na página 31.
- SHAHHOSSEINI, M.; HU, G.; PHAM, H. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Machine Learning with Applications*, Elsevier, v. 7, p. 100251, 2022. Citado na página 41.
- SHIN, H.-C.; ROTH, H. R.; GAO, M.; LU, L.; XU, Z.; NOGUES, I.; YAO, J.; MOLLURA, D.; SUMMERS, R. M. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1285–1298, 2016. Citado na página 38.
- SOUZA, V.; SILVA, L.; SANTOS, A.; ARAÚJO, L. Análise comparativa de redes neurais convolucionais no reconhecimento de cenas. *Anais do Computer on the Beach*, v. 11, p. 419–426, 2020. Citado na página 32.
- TAN, M.; LE, Q. E. et al. Rethinking model scaling for convolutional neural networks. In: *Proceedings of the International conference on machine learning, Long Beach, CA, USA*. [S.l.: s.n.], 2019. v. 15. Citado 2 vezes nas páginas 32 e 33.
- TIAN, M.; HE, J.; YU, X.; CAI, C.; GAO, Y. Mcmc guided cnn training and segmentation for pancreas extraction. *IEEE Access*, IEEE, v. 9, p. 90539–90554, 2021. Citado 4 vezes nas páginas 44, 46, 47 e 61.
- TONDJI, I. C. T.; SCAPICCHIO, C.; LIZZI, F.; FANTACCI, M. E.; OLIVA, P.; RETICO, A. Deep learning model with attention mechanism for a 3d pancreas segmentation in ct scans. *Mathematics*, MDPI, v. 13, n. 24, p. 3942, 2025. Citado 3 vezes nas páginas 44, 47 e 61.
- VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. *Proceedings of the xxix conference on graphics, patterns and images*. [S.l.], 2016. v. 1, n. 4. Citado na página 29.
- VASWANI, A. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. Citado 4 vezes nas páginas 18, 26, 29 e 30.
- WOLPERT, D. H. Stacked generalization. *Neural networks*, Elsevier, v. 5, n. 2, p. 241–259, 1992. Citado na página 40.
- XIE, E.; WANG, W.; YU, Z.; ANANDKUMAR, A.; ALVAREZ, J. M.; LUO, P. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, v. 34, p. 12077–12090, 2021. Citado 3 vezes nas páginas 10, 36 e 37.
- YANG, Q.; YAN, P.; ZHANG, Y.; YU, H.; SHI, Y.; MOU, X.; KALRA, M. K.; ZHANG, Y.; SUN, L.; WANG, G. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 6, p. 1348–1357, 2018. Citado na página 17.
- YAO, X.; SONG, Y.; LIU, Z. Advances on pancreas segmentation: a review. *Multimedia Tools and Applications*, Springer, v. 79, n. 9, p. 6799–6821, 2020. Citado na página 18.

YU, Q.; XIE, L.; WANG, Y.; ZHOU, Y.; FISHMAN, E. K.; YUILLE, A. L. Saliency transformation network: Incorporating multi-stage visual cues for pancreas segmentation. *CoRR*, 2017. Citado 3 vezes nas páginas 45, 47 e 61.

ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. In: SPRINGER. *European conference on computer vision*. [S.l.], 2014. p. 818–833. Citado na página 29.

ZHANG, X.; SU, T.; ZHANG, Y.; CUI, H.; TAN, Y.; ZHU, J.; XIA, D.; ZHENG, H.; LIANG, D.; GE, Y. Transferring u-net between low-dose ct denoising tasks: a validation study with varied spatial resolutions. *Quantitative Imaging in Medicine and Surgery*, AME Publications, v. 14, n. 1, p. 640, 2024. Citado na página 38.

ZHAO, N.; TONG, N.; RUAN, D.; SHENG, K. Fully automated pancreas segmentation with two-stage 3d convolutional neural networks. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2019. p. 201–209. Citado 5 vezes nas páginas 44, 46, 47, 60 e 61.

ZHAO, Z.; CHUAH, J. H.; CHOW, C. O.; XIA, K.; TEE, Y.; HUM, Y.; LAI, K. W. Machine learning approaches in comparative studies for alzheimer’s diagnosis using 2d mri slices. *Turkish Journal of Electrical Engineering and Computer Sciences*, v. 32, p. 93–107, 02 2024. Citado 2 vezes nas páginas 10 e 37.

ZHENG, H.; QIAN, L.; QIN, Y.; GU, Y.; YANG, J. Improving the slice interaction of 2.5 d cnn for automatic pancreas segmentation. *Medical Physics*, Wiley Online Library, v. 47, n. 11, p. 5543–5554, 2020. Citado 4 vezes nas páginas 45, 46, 47 e 61.

ZHOU, Z.; SIDDIQUEE, M. M. R.; TAJBAKHSI, N.; LIANG, J. Unet++: A nested u-net architecture for medical image segmentation. In: SPRINGER. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. [S.l.], 2018. p. 3–11. Citado na página 34.

ZHOU, Z.-H. Ensemble learning. In: *Machine learning*. [S.l.]: Springer, 2021. p. 181–210. Citado na página 40.

ZHU, Y.; HU, P.; LI, X.; TIAN, Y.; BAI, X.; LIANG, T.; LI, J. An end-to-end data-adaptive pancreas segmentation system with an image quality control toolbox. *Journal of Healthcare Engineering*, Wiley Online Library, v. 2023, n. 1, p. 3617318, 2023. Citado 4 vezes nas páginas 44, 46, 47 e 61.