

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Bruno Ferreira Tomé

**Avaliação da Swin-Unet para Segmentação de
Lesões de Câncer de Endométrio em Imagens
de Tomografia Computadorizada.**

Uberlândia, Brasil

2026

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Bruno Ferreira Tomé

**Avaliação da Swin-Unet para Segmentação de Lesões de
Câncer de Endométrio em Imagens de Tomografia
Computadorizada.**

Trabalho de conclusão de curso apresentado
à Faculdade de Computação da Universidade
Federal de Uberlândia, como parte dos requi-
sitos exigidos para a obtenção título de Ba-
charel em Ciência da Computação.

Orientador: Fernanda Maria da Cunha Santos

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2026

Bruno Ferreira Tomé

Avaliação da Swin-Unet para Segmentação de Lesões de Câncer de Endométrio em Imagens de Tomografia Computadorizada.

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 13 de Fevereiro de 2026:

Fernanda Maria da Cunha Santos
Orientador

Professor

Professor

Uberlândia, Brasil
2026

Agradecimentos

Agradeço a Deus, cuja presença me concedeu forças e orientação ao longo de toda a minha trajetória, especialmente nos momentos mais desafiadores. À minha orientadora, Professora Fernanda Maria da Cunha Santos, pelo apoio, pela sabedoria compartilhada e pela excelência na condução deste trabalho, sempre com dedicação e incentivo. Agradeço os professores Maria Adriava e Leandro Nogueira, que, com tanta generosidade e atenção, aceitaram participar da avaliação e contribuir de forma tão valiosa para este trabalho. Aos meus pais, por me incentivarem e acreditarem em mim, mesmo nos momentos difíceis, à minha irmã, pelo seu exemplo de dedicação aos estudos, e à minha família, por todo o amor, apoio incondicional e por sempre acreditarem no meu potencial, sendo a base de tudo o que conquistei. Aos amigos que conquistei ao longo da graduação - Amanda, Huryel, Isabelli e Tainá - que se tornaram minha família durante essa jornada, tornando o caminho mais leve, com companheirismo, apoio e momentos inesquecíveis. À Amanda e Maria Fernanda, amigas desde a escola, que me acompanham nessa trajetória, pela amizade duradoura, pelo apoio constante e por estarem presentes em todas as fases da minha vida. Vocês me acolheram nos momentos em que mais precisei, e serei eternamente grato. Por fim, agradeço ao Wander, pelo apoio constante, por tornar meus dias mais felizes e por me mostrar, na prática, o significado de amar e ser amado.

Resumo

A segmentação automática de câncer de endométrio em imagens de Tomografia Computadorizada (TC) é uma tarefa bastante desafiadora, dificultada pelo baixo contraste entre os tecidos moles da região pélvica e pela morfologia irregular dos tumores. Este trabalho investiga a aplicação da arquitetura Swin-Unet, baseada em Vision Transformers, para superar as limitações das Redes Neurais Convolucionais (CNNs) tradicionais na tarefa de segmentação de imagens. Para validar a proposta realizou-se um estudo comparativo com a arquitetura U-Net utilizando backbones clássicos (VGG-16 e ResNet50). Visando garantir robustez, todos os modelos foram submetidos a execuções independentes. Os resultados demonstraram a superioridade da Swin-Unet, que alcançou um F1-Score médio de 0,8129 e uma Sensibilidade de 0,8297, superando as abordagens convolucionais. Além de apresentar o melhor desempenho, a arquitetura proposta demonstrou maior estabilidade frente à inicialização dos pesos, contrastando com a alta variabilidade observada na ResNet50. As análises visuais confirmaram que o mecanismo de atenção permitiu uma delimitação mais precisa das bordas da lesão, validando a eficácia da Swin-Unet para o auxílio no diagnóstico médico.

Palavras-chave: Segmentação de Imagens Médicas, Câncer de Endométrio, Deep Learning, Swin-Unet, Vision Transformers.

Lista de ilustrações

Figura 1 – Comparação entre útero saudável e útero com câncer de endométrio. ¹	14
Figura 2 – Arquitetura de uma rede Transformer, com N codificadores à esquerda e N decodificadores à direita. Fonte: Adaptado de Vaswani et al. (2017).	16
Figura 3 – Imagens de tomografia computadorizada (TC) e suas respectivas máscaras (GT) utilizadas neste trabalho.	22
Figura 4 – Fluxo das etapas de pré-processamento e pós-processamento das imagens do conjunto de dados.	23
Figura 5 – Exemplo dos dois mecanismos de atenção no Swin Transformer: à esquerda, o Window-based Multi-head Self-Attention (W-MSA), no qual a atenção é calculada apenas dentro de janelas locais fixas; à direita, o Shifted Window-based Multi-head Self-Attention (SW-MSA), onde as janelas são deslocadas para permitir a troca de informações entre diferentes regiões da imagem.	24
Figura 6 – Arquitetura Swin-Unet utilizada, composta por blocos de codificação e decodificação baseados em Swin Transformers, conexões de salto e camada <i>bottleneck</i> , aplicada à segmentação de lesões em imagens TC.	25
Figura 7 – Esquema de um Bloco Residual. A conexão de identidade (<i>skip connection</i>) permite que o gradiente flua diretamente, mitigando o problema de degradação e facilitando o treinamento de redes profundas. A função de ativação utilizada é a Unidade Linear Retificada (ReLU).	27
Figura 8 – Comparação das segmentações obtidas pela Swin-Unet.	32
Figura 9 – Matriz de Confusão - Swin-Unet	34
Figura 10 – Curva ROC - Swin-Unet	35
Figura 11 – Matriz de Confusão - ResNet50.	36
Figura 12 – Curva ROC - ResNet50.	37
Figura 13 – Matriz de Confusão - VGG-16	37
Figura 14 – Curva ROC - VGG-16	38
Figura 15 – Resultado Segmentação - ResNet50.	39
Figura 16 – Resultado Segmentação - VGG-16.	39
Figura 17 – Resultado Segmentação - Swin-Unet.	40

Lista de tabelas

Tabela 1 – Parâmetros do equipamento PET/TC	21
Tabela 2 – Resultados dos Experimentos na Swin-Unet, com Diferentes Tamanhos de Lote	31
Tabela 3 – Resultados das cinco execuções da Swin-Unet.	33
Tabela 4 – Swin-Unet - Média $\pm$ Desvio Padrão	33
Tabela 5 – Comparativo de desempenho entre as arquiteturas - Média $\pm$ Desvio Padrão de 5 execuções	35

Lista de abreviaturas e siglas

CE	Câncer no Endométrio
RM	Ressonância Magnética
TC	Tomografia Computadorizada
CAD	Diagnóstico Auxiliado por Computador
CNNs	Redes Neurais Convolucionais
ECPC-IDS	<i>(Endometrial Cancer PET/TC Image Dataset)</i>
GT	<i>Ground Truth</i>
RNNs	Redes Neurais Recorrentes
ViT	<i>Vision Transformer</i>
NLP	Processamento de Linguagem Natural
W-MSA	<i>Window-based Multi-Head Self-Attention</i>
SW-MSA	<i>Shifted Window Multi-Head Self-Attention</i>
SMP	<i>Segmentation Models Pytorch</i>
GPU	Unidade de Processamento Gráfico
VP	Verdadeiro Positivo
FP	Falso Positivo
FN	Falso Negativo
VN	Verdadeiro Negativo

Sumário

1	INTRODUÇÃO	9
1.1	Justificativa	10
1.2	Objetivo	11
1.3	Organização do Texto	11
2	REVISÃO BIBLIOGRÁFICA	13
2.1	Câncer no Endométrio	13
2.2	Redes Neurais Convolucionais e a U-Net	14
2.3	Redes Neurais baseadas em Transformers	15
2.3.1	Transformers em Visão Computacional e o <i>Vision Transformer</i> (ViT)	17
2.3.2	Swin Transformer	17
2.4	Trabalhos Correlatos	17
3	DESENVOLVIMENTO	20
3.1	Análise da Base de dados	20
3.2	Pré-Processamento	22
3.3	Segmentação das Lesões	23
3.3.1	Swin-Unet	23
3.3.2	U-Net	25
3.4	Treinamento do Modelo	27
3.4.1	Tamanho do lote (<i>Batch Size</i>)	28
3.4.2	Métricas de Avaliação	29
4	RESULTADOS	31
4.1	Análise do impacto do tamanho do <i>batch size</i>	31
4.2	Desempenho da Swin-Unet	33
4.3	Comparação de Desempenho entre Swin-Unet e U-Net	35
5	CONCLUSÃO	41
	REFERÊNCIAS	43

1 Introdução

O Câncer de Endométrio (CE) é um tipo de câncer que se origina no revestimento interno do útero e representa uma das doenças ginecológicas mais comuns globalmente. De acordo com [Chaves et al. \(2025\)](#), trata-se do sexto câncer mais frequente entre mulheres, com incidência crescente associada a fatores como obesidade e envelhecimento populacional. Mesmo com os avanços na detecção e no tratamento, o CE ainda constitui um desafio importante para a saúde feminina. O recidiva da doença, as dificuldades no diagnóstico precoce e a progressão do câncer continuam impactando de maneira significativa as taxas de sobrevivência e a qualidade de vida das pacientes ([CHAVES et al., 2025](#)).

O diagnóstico pode ser realizado por meio de exames como a Ressonância Magnética (RM), a Tomografia Computadorizada (TC) e as biópsias guiadas. No entanto, interpretar essas imagens ainda é um processo complexo e pode variar de um especialista para outro. [Kurata et al. \(2021\)](#) ressaltam que a segmentação manual das lesões, além de exigir muito tempo e trabalho, está sujeita à avaliação de um especialista. Somam-se a isso evidências recentes mostrando que métodos tradicionais para diagnósticos podem subestimar a presença da malignidade em até 45,3% dos casos, o que reforça a necessidade urgente de ferramentas computacionais, como a segmentação de imagens, para o diagnóstico mais preciso e eficiente na tomada de decisão clínica ([WANG et al., 2025](#)).

Nos últimos anos, ferramentas baseadas em visão computacional têm ganhado espaço no processamento de imagens médicas, mostrando resultados promissores em diversas aplicações clínicas ([CAO et al., 2021](#)). Principalmente, a etapa de segmentação tem se destacado como uma das mais importantes nesse processo, pois desempenha um papel fundamental na construção de sistemas de Diagnóstico Auxiliado por Computador (CAD). Uma segmentação consistente permite identificar e delimitar lesões com maior precisão, contribuindo diretamente para análises mais confiáveis.

Junto ao processo de segmentação, destacam a presença de arquiteturas de redes neurais com o propósito de atuar na classificação para identificar células cancerígenas ou até substituir o processo de segmentação, como ocorre em alguns trabalhos com as redes neurais profundas ([MINAEE et al., 2022](#)). As Redes Neurais Convolucionais (CNNs) também têm apresentado resultados sólidos em tarefas de segmentação e classificação, permitindo diferenciar tecidos saudáveis de regiões suspeitas em várias condições clínicas ([FIGUEREDO et al., 2023](#); [PINTO et al., 2024](#)). Além disso, estudos mostram que esses modelos podem reduzir a necessidade de procedimentos invasivos e apoiar o diagnóstico de forma mais eficiente.

Com o avanço das arquiteturas das redes neurais baseadas em Transformers, es-

pecialmente direcionado na segmentação de imagens médicas, novas abordagens têm sido propostas para superar limitações das CNNs. Pesquisas recentes evidenciam que Transformers hierárquicos, como o Swin Transformer (LIU et al., 2021), oferecem maior capacidade de modelar dependências de longo alcance, aprimorando o desempenho em cenários complexos de baixa definição de contraste, problema comum em exames de Tomografia Computadorizada.

Assim, o escopo deste trabalho tem como foco a segmentação semântica, voltada à delimitação precisa da lesão em imagens de tomografia computadorizada de pacientes com diagnóstico confirmado ou forte suspeita clínica de câncer endometrial. Nesse contexto, o objetivo é implementar modelos computacionais para aprender a segmentar a região tumoral. Ou seja, a segmentação semântica realiza uma classificação em nível de pixel, mesmo em imagens cuja maior parte dos pixels corresponde a tecidos saudáveis, e que serão tratados como classe de fundo (*background*). Dessa forma, o modelo aprende a diferenciar as características da lesão tumoral em relação ao tecido normal que está ao redor, conforme adotado em trabalhos como os de Kurata et al. (2021), Tang et al. (2024).

1.1 Justificativa

Apesar dos avanços recentes em técnicas de segmentação baseadas em *deep learning*, a pesquisa aplicada especificamente ao câncer de endométrio ainda é limitada, em grande parte pela escassez de bases de dados públicas e padronizadas. Essa lacuna dificulta o desenvolvimento e a comparação de modelos voltados para essa neoplasia, restringindo o avanço de soluções computacionais mais robustas para o apoio ao diagnóstico (TANG et al., 2024).

Um passo importante para mudar esse cenário foi a publicação do ECPC-IDS, o primeiro conjunto de dados público em larga escala contendo imagens PET/TC de pacientes com diagnóstico confirmado de câncer de endométrio. Desde sua disponibilização, a base vem sendo adotada em diferentes investigações, incluindo segmentação de lesões, análise de regiões hipermetabólicas e aplicações em radiogenômica, o que evidencia sua consolidação como referência experimental na área (TANG et al., 2024; DU et al., 2025).

No entanto, os resultados iniciais obtidos com o ECPC-IDS revelaram um desafio importante nas arquiteturas convolucionais tradicionais, como a U-Net, que apresentam desempenho limitado na segmentação das lesões em imagens de tomografia computadorizada. Fatores como o baixo contraste entre tecidos, a heterogeneidade tumoral e a variabilidade anatômica da região pélvica tornam essa tarefa particularmente difícil para modelos baseados apenas em convoluções (TANG et al., 2024). Trabalhos mais recentes têm buscado superar essas limitações por meio de arquiteturas que incorporam mecanismos de atenção e estratégias para ampliar o campo receptivo (PENG et al., 2025), o que

reforça que a segmentação de lesões endometriais em TC ainda é um problema em aberto.

Diante desse cenário, investigar modelos capazes de capturar informações contextuais de forma mais ampla e eficiente se torna necessário. Arquiteturas baseadas em Transformers, como a Swin-Unet, surgem como uma alternativa promissora, especialmente pela capacidade de modelar dependências de longo alcance em imagens com baixo contraste e estruturas complexas. Avaliar essa arquitetura sobre o ECPC-IDS pode contribuir tanto para o avanço dos métodos de segmentação aplicados ao câncer de endométrio quanto para fortalecer o uso da base como *benchmark* para pesquisas futuras.

1.2 Objetivo

Este trabalho tem como objetivo avaliar o desempenho da rede Transformers Swin-Unet, uma arquitetura baseada em redes neurais profundas, direcionada à imagens de tomografia computadorizada, visando a segmentação automática de lesões de câncer de endométrio. Também, verificar se essa abordagem é capaz de superar as limitações de desempenho observadas em modelos convolucionais tradicionais, como a U-Net.

Para alcançar os objetivos propostos, a metodologia estrutura-se a partir da utilização de uma base de dados pública especializada (*ECPC-IDS*), composta por exames de TC e suas respectivas anotações médicas. As imagens, compostas por cortes axiais 2D extraídos dos volumes de TC, passarão por uma etapa de pré-processamento que inclui redimensionamento para 224×224 pixels, conversão para tensores e aplicação de técnicas de aumento de dados com transformações rígidas, visando contornar a escassez de amostras e melhorar a robustez do treinamento. A implementação e o treinamento do modelo serão realizados em linguagem Python, utilizando o framework PyTorch (PINTO et al., 2024), com ajustes específicos em funções de perda e otimizadores para segmentação binária. Por fim, o desempenho da Swin-Unet será validado e comparado estatisticamente contra modelos convolucionais consolidados na literatura, utilizando métricas focadas no contexto médico, com destaque para o Coeficiente Dice (F1-Score).

1.3 Organização do Texto

Para facilitar a compreensão do projeto, este trabalho está estruturado em capítulos que abrangem todas as etapas e informações relevantes ao estudo. O Capítulo 1 apresenta o contexto, a motivação, justificativa e os objetivos deste projeto. O Capítulo 2 discute a fundamentação teórica, abordando os principais conceitos envolvidos, bem como trabalhos relacionados que contribuíram para a construção do referencial e para o alinhamento das ideias do estudo. O Capítulo 3 descreve detalhadamente todas as etapas de desenvolvimento do projeto, incluindo as decisões de implementação do modelo

Swin-UNET. Além disso, também apresenta as métricas utilizadas, os testes realizados e os critérios de comparação com a arquitetura U-Net. O [Capítulo 4](#) é dedicado à exposição e análise dos resultados obtidos, incluindo a apresentação dos experimentos realizados e a comparação final entre os modelos avaliados. Por fim, o [Capítulo 5](#) reúne as conclusões do estudo, destacando as contribuições alcançadas e sugerindo possíveis direções para pesquisas futuras.

2 Revisão Bibliográfica

Este capítulo apresenta os fundamentos teóricos necessários para a compreensão deste estudo. Além disso, é feita uma sucinta descrição de trabalhos relacionados ao tema proposto neste, contribuindo para uma melhor contextualização.

2.1 Câncer no Endométrio

O câncer de endométrio é uma neoplasia maligna que afeta o revestimento interno do útero, o endométrio, conforme ilustrado na Figura 1. É o tipo mais comum de câncer ginecológico e apresenta uma incidência crescente em países como o Brasil (FIDELIS; COSTA, 2024; PAULINO et al., 2018). A doença é mais frequente em mulheres na pós-menopausa, geralmente diagnosticada acima dos 50 anos de idade, embora também possa ocorrer em mulheres mais jovens (FIDELIS; COSTA, 2024).

No Brasil, estimativas indicam um aumento contínuo no número de novos casos. Segundo Paulino et al. (2018), projetava-se que o número de casos passasse de cerca de 6.000 em 2016 para mais de 9.000 em 2025. Esse aumento na incidência está associado a mudanças no perfil demográfico e de saúde da população, especificamente o envelhecimento e o aumento das taxas de obesidade e sobrepeso (PAULINO et al., 2018).

A obesidade é reconhecida como um dos fatores de risco mais importantes para o desenvolvimento do câncer de endométrio. Dados do IBGE destacam que a porcentagem de mulheres brasileiras com excesso de peso aumentou significativamente nas últimas décadas, criando um cenário propício para a elevação da incidência da doença (PAULINO et al., 2018). Além da obesidade, outros fatores de risco relevantes incluem:

- **fatores hormonais e reprodutivos:** nunca ter tido filhos e menopausa tardia aumentam a exposição aos estrogênios sem a oposição da progesterona, elevando o risco (FIDELIS; COSTA, 2024).
- **comorbidades:** hipertensão arterial é frequentemente associada à doença (FIDELIS; COSTA, 2024; PAULINO et al., 2018).
- **uso de medicamentos:** o uso de Tamoxifeno para tratamento de câncer de mama é um fator de risco conhecido (FIDELIS; COSTA, 2024).
- **predisposição genética:** a Síndrome de Lynch aumenta significativamente a predisposição ao desenvolvimento deste tipo de câncer (FIDELIS; COSTA, 2024).

O câncer de endométrio é dividido em dois tipos patogênicos. O tipo I (carcinoma endometriode) é o mais comum, sendo estrogênio-dependente e geralmente associado a um melhor prognóstico. Já o tipo II (não endometriode) é menos comum, não depende de estrogênio e tende a ser mais agressivo (FIDELIS; COSTA, 2024).

O sintoma mais característico da doença é o sangramento uterino anormal, especialmente na pós-menopausa, o que muitas vezes permite o diagnóstico em estágios iniciais (FIDELIS; COSTA, 2024). O diagnóstico precoce é determinante para o prognóstico, quando a doença é detectada em fase inicial e restrita ao útero, as taxas de cura são altas. No entanto, o diagnóstico tardio ou a falta de acesso adequado ao tratamento especializado, como radioterapia, podem impactar negativamente as taxas de mortalidade (PAULINO et al., 2018; FIDELIS; COSTA, 2024).

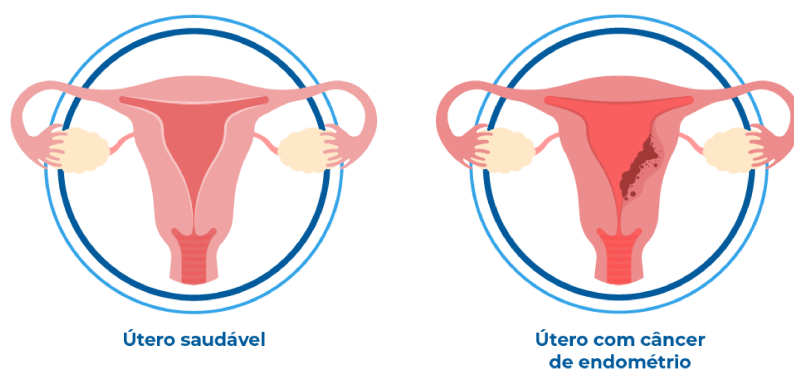


Figura 1 – Comparação entre útero saudável e útero com câncer de endométrio.¹

2.2 Redes Neurais Convolucionais e a U-Net

O avanço das redes neurais *Deep Learning* tem proporcionado resultados notáveis em tarefas de reconhecimento de imagens e detecção de objetos, superando técnicas convencionais que dependiam da extração manual de características (SHAZIYA; SHYAMALA, 2020). Nesse contexto, as CNNs se consolidaram como o modelo mais sucedido para o processamento de dados visuais, sendo capazes de aprender automaticamente hierarquias de características (como bordas e formas) diretamente a partir dos dados de entrada (SHAZIYA; SHYAMALA, 2020; ANBUDEVI; SUGANTHI, 2022).

Enquanto as CNNs tradicionais são excelentes para classificação, a análise de imagens médicas frequentemente exige a segmentação semântica, que consiste em classificar cada pixel da imagem individualmente para delinear com precisão as fronteiras de uma lesão ou órgão (FARIAS et al., 2025; ANBUDEVI; SUGANTHI, 2022).

¹ Disponível em: <<https://www.einstein.br/n/glossario-de-saude/cancer-de-endometrio>>. Acesso em: 28 nov. 2025.

Para atender a essa necessidade, [Ronneberger, Fischer e Brox \(2015\)](#) propuseram a U-Net, uma arquitetura de rede neural totalmente convolucional projetada especificamente para segmentação de imagens biomédicas. Diferente das CNNs de classificação, que apenas reduzem a imagem, a U-Net possui uma estrutura simétrica em forma de U, composta por dois caminhos principais ([OLIVEIRA et al., 2025](#); [SHAZIYA; SHYAMALA, 2020](#)):

- **Caminho de Contração (*Encoder*):** Segue a estrutura de uma CNN convencional, aplicando convoluções para capturar o contexto da imagem e extrair características de alto nível, reduzindo a resolução espacial ([OLIVEIRA et al., 2025](#)).
- **Caminho de Expansão (*Decoder*):** Realiza operações de convoluções transpostas ou interpolação para restaurar a resolução original da imagem e permitir a localização precisa das estruturas segmentadas ([SHAZIYA; SHYAMALA, 2020](#)).

Uma das inovações cruciais da U-Net são as conexões de salto (*skip connections*). Essas conexões transferem mapas de características diretamente do *encoder* para as camadas correspondentes no *decoder* ([ANBUDEVI; SUGANTHI, 2022](#)). Segundo [Farias et al. \(2025\)](#), esse mecanismo é fundamental para recuperar detalhes espaciais finos que seriam perdidos durante o processo de contração, permitindo uma segmentação mais precisa das bordas das lesões.

A flexibilidade da U-Net permite que o seu *encoder* seja substituído por arquiteturas de CNNs clássicas e pré-treinadas, conhecidas como *backbones*. [Farias et al. \(2025\)](#) destacam que arquiteturas profundas, como a VGG ou ResNet, continuam representando o estado da arte para extração de características.

Ao utilizar esses modelos como codificadores na U-Net, aplica-se a técnica de Transferência de Aprendizado (*Transfer Learning*). Isso permite que o modelo de segmentação inicie o treinamento com pesos que já aprenderam a extrair características robustas, como texturas e formas de grandes bancos de dados genéricos (exemplo: ImageNet), acelerando a convergência e melhorando o desempenho em tarefas médicas onde os dados rotulados são escassos ([FARIAS et al., 2025](#); [KURATA et al., 2021](#)).

2.3 Redes Neurais baseadas em Transformers

As redes Transformer são uma arquitetura de *deep learning* introduzida inicialmente por [Vaswani et al. \(2017\)](#), se destacando no processamento de linguagem natural e, recentemente em aplicações de visão computacional, incluindo na área médica ([CAO et al., 2021](#)). Essa arquitetura foi projetada para superar limitações de modelos anteriores, como

as redes neurais recorrentes (RNNs), que enfrentam problemas de gradientes e dificuldades na captura de dependências de longo prazo em sequências extensas (NASCIMENTO, 2023).

Para resolver essas limitações, as redes Transformer introduz o mecanismo de autoatenção, que permite analisar todos os tokens em paralelo, evitando o processamento estritamente sequencial das RNNs. A estrutura geral, ilustrada na Figura 2, segue o modelo codificador-decodificador, em que o codificador recebe vetores de entrada (*embeddings*) e os processa por múltiplas camadas compostas por dois módulos principais: o módulo de autoatenção e o módulo *feed-forward* (VASWANI et al., 2017).

O módulo de autoatenção identifica quais tokens são mais relevantes dentro da sequência, atribuindo pesos que destacam suas contribuições. Já o módulo *feed-forward* realiza processamento individual de cada token. Esse mecanismo é repetido várias vezes, resultando nas chamadas cabeças de atenção (*attention heads*), que utilizam vetores de consulta (Q), chave (K) e valor (V) para calcular, por meio de uma função *softmax*, o grau de importância de cada token (VASWANI et al., 2017).

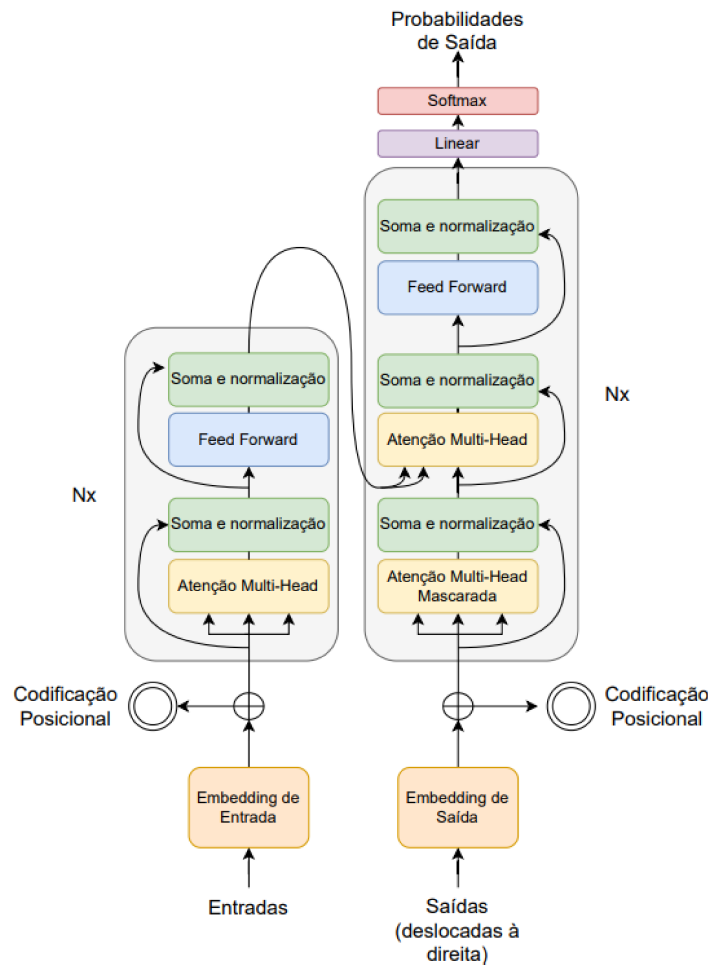


Figura 2 – Arquitetura de uma rede Transformer, com N codificadores à esquerda e N decodificadores à direita. Fonte: Adaptado de Vaswani et al. (2017).

2.3.1 Transformers em Visão Computacional e o *Vision Transformer* (ViT)

A capacidade dos Transformers de capturar relacionamentos globais e de longo prazo, impulsionou seu uso além do Processamento de Linguagem Natural (NLP). Em visão computacional, um marco importante foi a introdução do ViT, proposto por [Dosovitskiy et al. \(2021\)](#). No ViT, a imagem é dividida em pequenos recortes fixos (*patches*), que são linearizados, transformados em vetores e tratados como uma sequência de tokens. Assim, cada patch assume um papel análogo ao de uma palavra em texto.

Esse paradigma permite que o modelo capture dependências globais entre regiões distantes da imagem por meio da autoatenção. Porém, o ViT apresenta limitações práticas em imagens de alta resolução. O custo computacional da autoatenção cresce ao aumentar o número de *patches*, tornando sua aplicação direta em imagens médicas complexa e frequentemente inviável ([Liu et al., 2021](#)).

2.3.2 Swin Transformer

Para mitigar essa limitação, [Liu et al. \(2021\)](#) propuseram o *Swin Transformer* (*Shifted Window Transformer*), uma arquitetura que realiza autoatenção local em janelas fixas da imagem por meio do mecanismo *Window-based Multi-Head Self-Attention* (W-MSA). Em seguida, as janelas são deslocadas, permitindo que diferentes regiões se comuniquem entre si, através do *Shifted Window Multi-Head Self-Attention* (SW-MSA). Esse mecanismo hierárquico reduz o custo computacional e permite trabalhar com imagens maiores, tornando o *Swin Transformer* mais adequado para aplicações médicas do que o ViT convencional.

Arquiteturas derivadas, como a Swin-Unet, combinam essa eficiência dos Transformers com estruturas de segmentação inspiradas em redes convolucionais, demonstrando desempenho superior em tarefas médicas, como segmentação de órgãos e detecção de lesões ([CAO et al., 2021](#); [PINTO et al., 2024](#)).

No Capítulo 3, a arquitetura Swin-Unet destinada a segmentação do câncer endometrial será detalhada.

2.4 Trabalhos Correlatos

A aplicação de *Deep Learning* para a segmentação do Câncer de Endométrio tem sido explorada em diferentes modalidades de imagem, com resultados variados dependendo da complexidade do exame e da arquitetura utilizada. Uma revisão recente conduzida por [Aggarwal et al. \(2025\)](#) analisou as técnicas de segmentação automática mais utilizadas para esta patologia. Os autores concluíram que, embora existam avanços, a maioria dos estudos se concentram em imagens de Ressonância Magnética (RM), devido ao melhor

contraste de tecidos moles, enquanto a aplicação em Tomografia Computadorizada (TC) permanece menos explorada e ainda desafiadora.

Kurata et al. (2021) propuseram um método de segmentação automática de câncer de endométrio em imagens de RM utilizando uma CNN baseada na arquitetura U-Net. O estudo utilizou uma base de dados privada com 200 pacientes e focou em superar a subjetividade e o tempo exigido pela segmentação manual. Os autores alcançaram o coeficiente *F1-Score* médio de 0.806 no conjunto de teste, demonstrando que arquiteturas convolucionais clássicas conseguem obter bons resultados em imagens de RM.

Em contrapartida, Tang et al. (2024) publicaram o *dataset* ECPC-IDS, o primeiro *benchmark* público focado em imagens multimodais (PET/TC) destinada ao estudo de câncer de endométrio. O trabalho não apenas disponibilizou os dados, mas também realizou um estudo comparativo de modelos de segmentação semântica. Diferentemente dos resultados em RM observados por Kurata et al. (2021), Tang et al. (2024) apontaram que modelos convolucionais tradicionais, incluindo a U-Net e a SegNet, apresentaram um desempenho insatisfatório nas imagens de TC deste *dataset*. O coeficiente *F1-Score* para a segmentação em TC atingiu um máximo de apenas 0.510, mostrando a dificuldade das CNNs em capturar o contexto necessário para delimitar lesões com baixo contraste em exames de tomografia.

Diante das limitações das convoluções em capturar dependências globais, Cao et al. (2021) propuseram a *Swin-Unet*, uma arquitetura inovadora baseada em *Transformers*. O modelo adapta o *Swin Transformer* para uma estrutura *Encoder-Decoder* semelhante à U-Net, mas substituindo as operações de convolução por mecanismos de autoatenção com janelas deslocadas (*Shifted Windows*).

Os experimentos conduzidos por Cao et al. (2021) em *datasets* médicos demonstraram que a abordagem baseada em *Transformers* supera métodos baseados puramente em CNNs, como a U-Net em tarefas de segmentação complexas. O estudo concluiu que a capacidade da Swin-Unet de modelar interações semânticas de longo alcance resulta em uma melhor preservação de bordas e detalhes finos, características que são essenciais para o diagnóstico médico preciso.

A revisão dos trabalhos correlatos evidencia que, apesar dos avanços obtidos em ressonância magnética por Kurata et al. (2021), a segmentação em tomografia computadorizada ainda permanece como um problema em aberto. Os resultados apresentados por Tang et al. (2024) reforçam essa lacuna, mostrando que modelos baseados em CNNs ainda não alcançam um bom desempenho nesse tipo de imagem.

Nesse sentido, o objetivo deste estudo é explorar a aplicação da arquitetura Swin-Unet, proposta por Cao et al. (2021), avaliando se os mecanismos de atenção e a capacidade de capturar dependências de longo alcance das redes *Transformers* podem superar

as limitações existentes nas CNNs aplicadas ao *dataset* ECPC-IDS. Logo, o objetivo central é investigar se essa abordagem pode elevar a qualidade da segmentação do câncer de endométrio em imagens de tomografia computadorizada.

3 Desenvolvimento

Este capítulo detalha a metodologia do modelo computacional proposto para a segmentação automática de lesões em imagens de tomografia computadorizada, através da implementação de uma arquitetura de rede neural Swin-Unet. A implementação foi realizada com algumas adaptações no código disponibilizado em github.com/ashish-sbisht/SwinUnetArchitecturePytorch, seguindo a abordagem metodológica proposta por [Cao et al. \(2021\)](#). Assim, o método proposto será dividido em etapas que abrangem desde a preparação da base de dados até a validação dos resultados.

3.1 Análise da Base de dados

Para este trabalho, foi utilizado o conjunto de dados público ECPC-IDS (*Endometrial Cancer PET/TC Image Dataset*) ([GROUP, 2023](#)), descrito por [Tang et al. \(2024\)](#). Esta é um depósito de imagens composto por exames de tomografia por emissão de pósitrons (PET) e tomografia computadorizada (TC) de 155 pacientes com diagnóstico confirmado de câncer endometrial, de acordo com a classificação da Federação Internacional de Ginecologia e Obstetrícia ([BHATLA et al., 2018](#)).

As imagens foram obtidas entre os anos de 2013 e 2022 no *Shengjing Hospital*, afiliado à *China Medical University*, utilizando equipamento PET/TC modelo *Discovery 690* da *GE Healthcare*. É importante destacar que o ECPC-IDS foi coletado a partir de uma população específica, composta exclusivamente por pacientes atendidas na China. Essa característica implica que os dados podem refletir particularidades demográficas, clínicas e de aquisição de imagem próprias dessa população, como protocolos de exame, perfil etário e fatores epidemiológicos regionais. Embora isso não comprometa a validade dos experimentos realizados, é um aspecto relevante a ser considerado ao interpretar os resultados e ao avaliar a capacidade de generalização dos modelos treinados para populações de outras regiões geográficas.

A escolha do ECPC-IDS se deve à sua relevância e representatividade na área, sendo uma das poucas bases públicas que contém exames PET/TC de pacientes com câncer endometrial. Além disso, os dados já foram validados por profissionais clínicos ([AGGARWAL et al., 2025](#)), o que contribui para tornar as análises e experimentos mais confiáveis. As especificações técnicas do equipamento PET/TC utilizado estão apresentadas na Tabela 1.

O ECPC-IDS se configura como o primeiro *dataset* público em larga escala focado no câncer de endométrio, contendo imagens multimodais, nos formatos DICOM e PNG, e

Tabela 1 – Parâmetros do equipamento PET/TC

Parâmetro	Especificação
Fabricante	GE Healthcare
Modelo	Discovery 690
Resolução da TC	512×512
Espessura dos cortes (slices)	3,27 mm
Resolução da PET	192×192

Fonte: Adaptado de [Tang et al. \(2024\)](#).

arquivos XML com anotações de regiões de interesse. O conjunto de dados está organizado em dois subconjuntos: o Subconjunto A, voltado à segmentação semântica, com 7.159 imagens e suas respectivas máscaras, e o Subconjunto B, direcionado à detecção de objetos, contendo 3.579 imagens anotadas.

As imagens foram anotadas por radiologistas, utilizando o software *3D Slicer*, com segmentações realizadas camada a camada. A segmentação foi inicialmente feita sobre as imagens PET e, posteriormente, alinhada às imagens TC correspondentes, respeitando as diferenças de resolução espacial entre as modalidades.

Entretanto, é importante ressaltar que, embora o ECPC-IDS seja um *dataset* composto por diferentes tipos de dados, para este trabalho foram utilizadas somente as imagens TC e suas respectivas máscaras de segmentação do Subconjunto A, não sendo considerado o uso das imagens PET. Essa escolha foi feita porque o Subconjunto A é específico para a tarefa de segmentação de imagens, que corresponde ao mesmo objetivo desse trabalho, enquanto o Subconjunto B foi utilizado para a parte de detecção de objetos, não sendo relevante para a metodologia de segmentação utilizada nesse estudo.

Em relação ao formato das imagens, é importante destacar que tanto as imagens de tomografia computadorizada quanto as suas respectivas máscaras de segmentação são padronizadas para possuírem a mesma resolução espacial durante a etapa de pré-processamento. Para garantir a compatibilidade com a arquitetura da rede, ambas são processadas como matrizes bidimensionais e redimensionadas para dimensões idênticas de 224×224 pixels, assegurando o alinhamento pixel a pixel entre a anatomia do paciente e a anotação médica. A distinção entre elas reside exclusivamente na representação dos valores, enquanto as imagens de TC mantêm um formato de canal único com valores contínuos em tons de cinza (normalizados para a faixa de tensores da rede), as imagens segmentadas assumem um formato estritamente binário, onde o valor numérico 1 representa a área do tumor de endométrio e o valor 0 delimita o tecido de fundo e os órgãos adjacentes saudáveis.

A Figura 3 exibe exemplos de imagens TC e suas respectivas máscaras (GT),

utilizadas no treinamento e avaliação do modelo.

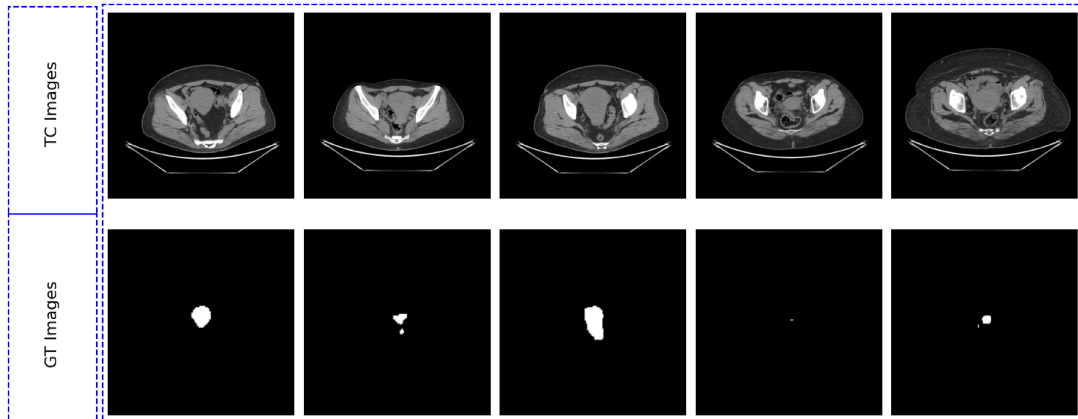


Figura 3 – Imagens de tomografia computadorizada (TC) e suas respectivas máscaras (GT) utilizadas neste trabalho.

O ECPC-IDS está disponível gratuitamente para uso no repositório Figshare: <https://figshare.com/articles/dataset/ECPC-IDS/23808258>.

3.2 Pré-Processamento

A etapa de pré-processamento foi fundamental para adequar os dados do *dataset* ECPC-IDS às necessidades do modelo Swin-Unet, garantindo a padronização dos dados de entrada. Inicialmente, foi realizado um processo de padronização em cada imagem de TC e sua respectiva máscara de segmentação. Este processo incluiu o redimensionamento das imagens para 224x224 pixels, afim de padronizar as dimensões de entrada exigidas pela Swin-Unet, a conversão para escala de cinza para garantir um formato de canal único e focar apenas nos dados de intensidade, que é o padrão para esse tipo de imagem, e a normalização dos valores de pixel para o intervalo de $[0, 1]$. Sendo essa última etapa, essencial para garantir a estabilidade do treinamento, prevenindo problemas como a explosão de gradiente e com isso, facilitando a uma convergência mais rápida da rede (GORGENS et al., 2009).

Após essa etapa, o conjunto de dados foi subdividido em três subconjuntos distintos utilizando a biblioteca *Scikit-learn*. A divisão seguiu a proporção de 70% para treinamento, 15% para validação e 15% para teste. O conjunto de treinamento foi utilizado no aprendizado dos parâmetros da Swin-Unet, o conjunto de validação para monitorar o comportamento do modelo durante o processo de otimização e prevenir o *overfitting* e por fim, o conjunto de teste para mensurar a capacidade de generalização do modelo. Essa estratégia busca garantir que o modelo seja treinado, ajustado e avaliado com amostras de dados independentes, garantindo maior confiabilidade nos resultados.

Com o objetivo de aumentar a variabilidade do conjunto de treinamento e tornar o modelo mais eficiente, foram aplicadas técnicas de aumento de dados. As técnicas utilizadas, no conjunto de treino, foram a transformação elástica, inversão horizontal e vertical e rotação de 180 graus em ambos os eixos. A aplicação dessas técnicas resultou no aumento do volume de imagens de treinamento, aumentando o conjunto de 390 para 1950 amostras, ajudando a ter uma melhor capacidade de generalização do modelo.

Por fim, todos os conjuntos de dados foram convertidos para tensores do *PyTorch*, o formato usado para o treinamento da arquitetura Swin-Unet por [Cao et al. \(2021\)](#). Uma dimensão adicional de canal foi inserida para adequar os dados ao formato esperado por redes convolucionais, sendo eles, canais, altura e largura, visando adequar as imagens ao formato de entrada requerido pelas redes. Posteriormente, os dados foram encapsulados em objetos *DataLoader*, responsáveis por organizar o fluxo de dados durante as etapas de treinamento e validação, gerenciando o carregamento em lotes (*batches*) de tamanho definido como 16, favorecendo a estabilidade do treinamento e a eficiência no processamento dos dados.

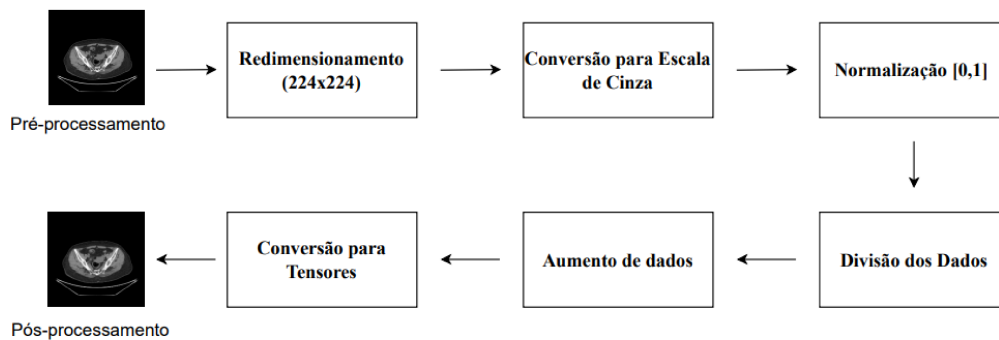


Figura 4 – Fluxo das etapas de pré-processamento e pós-processamento das imagens do conjunto de dados.

Fonte: Do Autor

A Figura 4 ilustra o fluxo completo das etapas de pré-processamento e a imagem resultante pós-processamento aplicadas às imagens do conjunto de dados, desde o redimensionamento inicial até a conversão final para tensores do *PyTorch*.

3.3 Segmentação das Lesões

3.3.1 Swin-Unet

Para a tarefa de segmentação das lesões de câncer de endométrio, foi adotada a arquitetura de rede neural Swin-Unet, um modelo que combina os princípios da U-Net com a eficiência das redes Transformers. A Swin-Unet foi escolhida por oferecer um equilíbrio entre desempenho, eficiência computacional e capacidade de generalização, sendo uma

das arquiteturas mais promissoras para tarefas de segmentação médica baseada em *deep learning* (PINTO et al., 2024).

Originalmente desenvolvidos para o processamento de linguagem natural, os Transformers têm demonstrado resultados promissores em tarefas de visão computacional (LIU et al., 2021). Sua abordagem consiste em dividir a imagem de entrada em pequenos recortes (*patches*), que são transformados em vetores e processados como uma sequência de *tokens*. Por meio de um mecanismo de autoatenção, a rede aprende as correlações entre diferentes regiões da imagem, permitindo uma compreensão contextual mais ampla (LIU et al., 2021)

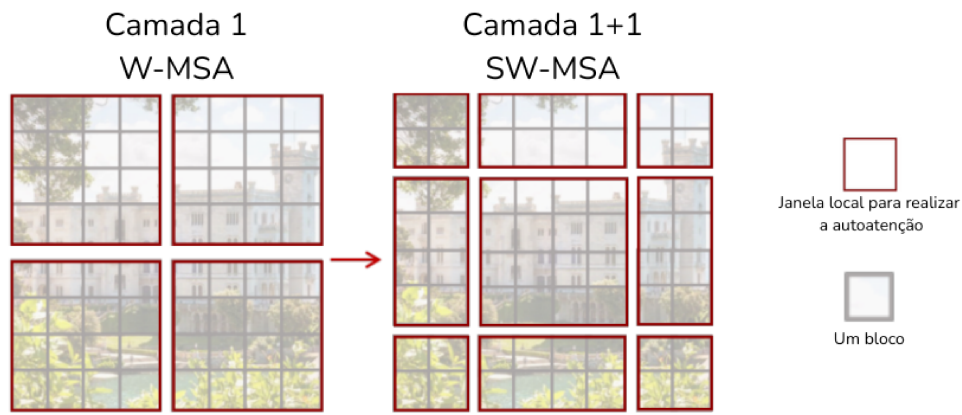


Figura 5 – Exemplo dos dois mecanismos de atenção no Swin Transformer: à esquerda, o Window-based Multi-head Self-Attention (W-MSA), no qual a atenção é calculada apenas dentro de janelas locais fixas; à direita, o Shifted Window-based Multi-head Self-Attention (SW-MSA), onde as janelas são deslocadas para permitir a troca de informações entre diferentes regiões da imagem.

A Swin-Unet é baseada, então, na arquitetura Swin Transformer, uma evolução mais eficiente do ViT proposta por Liu et al. (2021). Sua principal inovação reside na forma como a autoatenção é calculada. Em vez de relacionar todos os *patches* da imagem simultaneamente, o Swin Transformer utiliza dois mecanismos principais, como ilustrados na figura 5 e descrito na sequência:

- *Window-based Multi-head Self-Attention (W-MSA)*: calcula a autoatenção localmente, apenas dentro de janelas que agrupam *patches* vizinhos, focando no aprendizado de características detalhadas;
- *Shifted Window-based Multi-head Self-Attention (SW-MSA)*: permite a comunicação entre janelas deslocadas, promovendo a propagação da informação por toda a imagem e aumentando a percepção global do modelo.

A arquitetura Swin-Unet, detalhada por Cao et al. (2021) aplica esses blocos em uma estrutura de codificador-decodificador (*encoder-decoder*), semelhante à U-Net clássica.

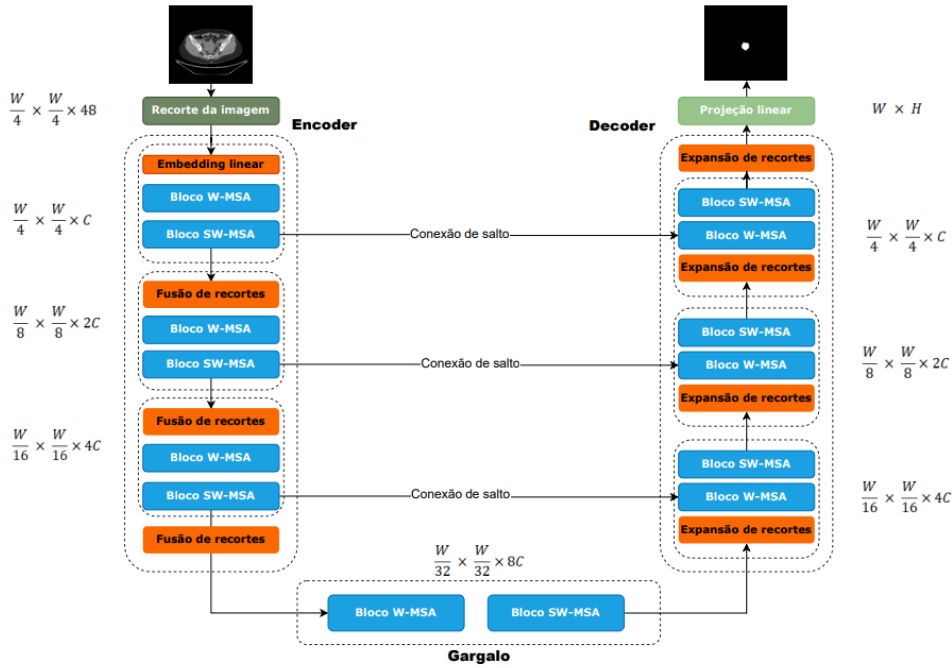


Figura 6 – Arquitetura Swin-Unet utilizada, composta por blocos de codificação e decodificação baseados em Swin Transformers, conexões de salto e camada *bottleneck*, aplicada à segmentação de lesões em imagens TC.

sica, sendo especialmente adequada para tarefas de segmentação. Tanto o codificador, responsável por extrair as características da imagem, quanto o decodificador, responsável por reconstruir a máscara de segmentação, são formados por blocos Swin Transformer. O modelo também utiliza conexões de salto (*skip connections*) para combinar as informações do codificador com as do decodificador, preservando detalhes da imagem, e incorpora uma camada intermediária do tipo *bottleneck*, ou gargalo. A arquitetura da rede é mostrada na figura 6. A saída da rede é um mapa de segmentação com um único canal, representando a região do câncer endometrial.

Neste trabalho, diferentemente de Pinto et al. (2024) e Cao et al. (2021), que utilizam transferência de aprendizado, foi optado por treinar o modelo do zero. Dessa forma, os pesos da rede não foram carregados de um modelo externo pré-treinado, mas sim inicializados por meio de um método otimizado, com a intenção de promover uma convergência mais estável e eficiente durante o processo de treinamento.

3.3.2 U-Net

Com o objetivo de avaliar o desempenho da Swin-Unet, foram implementadas duas arquiteturas de referência baseadas no modelo U-Net, amplamente consolidado na literatura para segmentação de imagens médicas (RONNEBERGER; FISCHER; BROX, 2015). Essas arquiteturas atuam como *baseline*, ou seja, modelos de referência que utilizam *backbones* (redes já consolidadas como extratores de características), permitindo comparar a

Swin-Unet com arquiteturas CNN em etapas de segmentação de lesões, reforçando a complexidade do problema e evidenciando o potencial da Swin-Unet em relação a abordagens convolucionais convencionais.

Para a implementação, foi utilizada a biblioteca *Segmentation Models Pytorch* (SMP), disponível publicamente em *GitHub*, a partir do repositório <https://github.com/qubvel/segmentation_models.pytorch>. Essa biblioteca fornece implementações estáveis e reproduzíveis de arquiteturas de segmentação, facilitando o uso de arquiteturas como a VGG16 e a ResNet50.

A principal vantagem de utilizar essa abordagem é a aplicação de aprendizado por transferência (*transfer learning*). Nela, os codificadores das redes são inicializados com pesos previamente treinados em grandes bases de dados, como a ImageNet, o que permite aos modelos aproveitar o conhecimento adquirido na extração de características de baixo nível (como bordas, texturas e formas) (SHIN et al., 2016). Com isso, o treinamento torna-se mais eficiente e os modelos alcançam melhor desempenho, algo essencial em cenários médicos.

A VGG16 é uma arquitetura clássica, caracterizada pela profundidade e pelo uso consistente de filtros convolucionais empilhados, extraíndo representações hierárquicas de maneira robusta (SIMONYAN; ZISSERMAN, 2015). Ao ser utilizada como *backbone* no codificador da U-Net, ela fornece mapas de características ricos e detalhados, auxiliando na reconstrução precisa de máscaras de segmentação. Essa abordagem vêm mostrando resultados notáveis em contextos médicos. Por exemplo, um modelo U-Net aprimorado com VGG-16 aplicado por Ghosh, Chaki e Santosh (2021), na segmentação de tumores cerebrais em imagens de ressonância magnética obteve ótimas precisões superando a U-Net original, mostrando que o VGG-16 pode melhorar significativamente a eficácia da segmentação de imagens médicas.

A Rede Neural Residual (ResNet) foi desenvolvida para superar a dificuldade encontrada no treinamento de redes com muitas camadas. Representando uma versão aprimorada das CNNs tradicionais, sua principal inovação são as conexões residuais, também chamadas de *skip connections*. Conforme ilustrado na Figura 7, nesse tipo de arquitetura a entrada original de um bloco convolucional é somada diretamente à sua saída, permitindo um fluxo mais eficiente de gradientes durante o treinamento. O objetivo fundamental desse mecanismo é reduzir o problema do desaparecimento do gradiente, um obstáculo comum que impede a convergência de redes muito profundas.

A arquitetura ResNet50, utilizada como referência no estudo proposto por Saifullah et al. (2025), é composta por 50 camadas de peso organizadas em blocos convolucionais principais. Cada um desses blocos, é formado por uma sequência de camadas de convolução e ao final da rede, um processo de agregação extrai o valor médio de cada mapa de características para a camada seguinte. Devido a essa capacidade de treinar arquiteturas

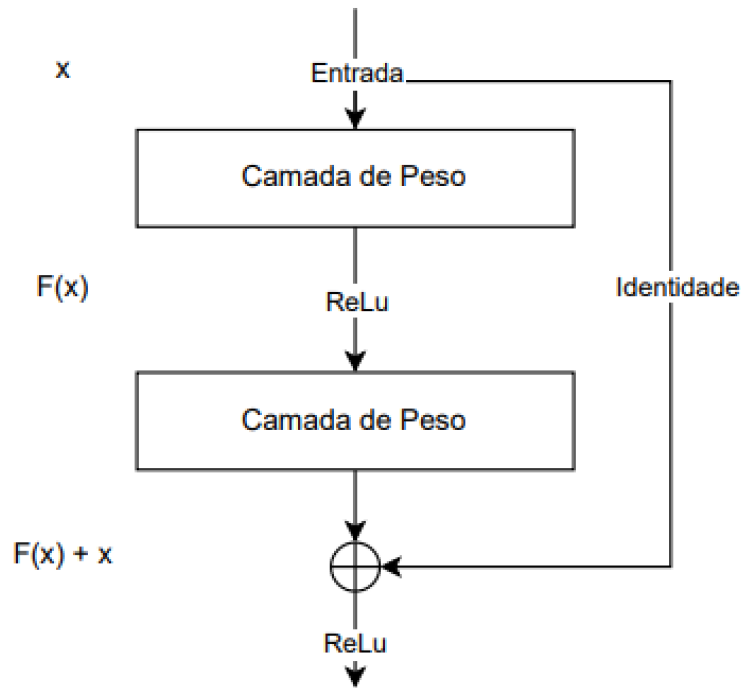


Figura 7 – Esquema de um Bloco Residual. A conexão de identidade (*skip connection*) permite que o gradiente flua diretamente, mitigando o problema de degradação e facilitando o treinamento de redes profundas. A função de ativação utilizada é a Unidade Linear Retificada (ReLU).

Fonte: Adaptado de [He et al. \(2015\)](#).

mais profundas e complexas de forma eficaz, a ResNet50 se estabeleceu como um backbone de referência, oferecendo uma capacidade de extração de características potencialmente mais poderosa para tarefas de visão computacional, como a segmentação de imagens médicas ([JAIN et al., 2024](#)).

3.4 Treinamento do Modelo

O treinamento foi realizado para que os modelos aprendessem a executar a tarefa de segmentação das imagens. Essa etapa consistiu em ajustar progressivamente os parâmetros internos da rede neural, de forma que as previsões se aproximassem cada vez mais das respostas corretas. O treinamento foi feito em linguagem Python, utilizando a biblioteca *PyTorch*, bastante utilizada em projetos de aprendizado profundo ([CAO et al., 2021](#)). Sempre que possível, foi utilizada uma GPU (Unidade de Processamento Gráfico) com suporte à plataforma CUDA, que permitiu acelerar os cálculos necessários.

Durante esse processo, foram adotadas algumas configurações que influenciam diretamente o desempenho do modelo, como:

- **Função de perda:** aplicação da *Binary Cross-Entropy with Logits*, que compara a máscara gerada pelo modelo com a máscara correta, avaliando o quão próximas elas estão.
- **Otimizador:** utilização do algoritmo Adam (*Adaptive Moment Estimation*), que ajusta a taxa de aprendizado individualmente para cada parâmetro. Este otimizador é amplamente recomendado para arquiteturas baseadas em Transformers devido à sua eficiência em lidar com gradientes esparsos e convergência rápida (KINGMA; BA, 2017).
- **Tamanho do lote:** definido como 16 imagens processadas simultaneamente a cada passo de aprendizado, escolhido após experimentos comparativos (detalhados na Seção 4.1), para garantir o equilíbrio entre estabilidade e generalização.

O processo de treinamento foi configurado com um limite máximo de 100 épocas. Diferente da abordagem de épocas fixas (200) utilizada no *benchmark* de Tang et al. (2024), este trabalho adotou a técnica de parada antecipada (*early stopping*). Esta escolha foi essencial para permitir uma comparação justa entre modelos com dinâmicas de aprendizado diferentes, transferência de aprendizado vs. treinamento do zero, prevenindo o *overfitting* nos modelos de convergência rápida (U-Net) e garantindo o treinamento adequado da arquitetura mais complexa (Swin-UNet). O critério de parada foi definido pela ausência de melhora na perda de validação por 5 épocas consecutivas, respeitando um período de aquecimento de 25 épocas.

3.4.1 Tamanho do lote (*Batch Size*)

O *batch size* é um parâmetro que define a quantidade de amostras, ou imagens, que serão utilizadas pelo modelo durante cada iteração do processo de treinamento, antes da atualização dos pesos. Ou seja, ele determina o número de imagens e máscaras processadas simultaneamente em uma única passagem de *forward* e *backward* da rede.

O valor deste parâmetro influencia diretamente o comportamento do modelo, pois quando o valor é muito baixo, o modelo realiza atualizações frequentes e com poucos exemplos, o que pode dificultar a convergência e não gerar resultados esperados. Por outro lado, tamanhos muito altos podem tornar o treinamento mais estável, mas com risco de convergir para mínimos locais de pior generalização (KANDEL; CASTELLI, 2020).

Estudos recentes mostram que o tamanho do lote impacta significativamente o comportamento e a capacidade de generalização de redes neurais. Segundo Radiuk (2017), o aumento do *batch size* tende a aumentar a precisão de reconhecimento e suavizar as curvas de aprendizado, porém gasta um maior tempo de treinamento e consumo operacional.

O autor apontou que lotes muito pequenos, resultam em curvas de aprendizado mais instáveis, enquanto valores mais elevados favorecem a estabilidade.

Por outro lado, [Kandel e Castelli \(2020\)](#) mostram que tamanho de lotes moderados, entre 16 e 32, tendem a proporcionar um melhor equilíbrio entre estabilidade, eficiência e capacidade de generalização. Os autores destacaram que lotes muito grandes podem reduzir o erro de treinamento, mas não necessariamente melhoram a generalização do modelo, podendo levar à convergência em mínimos locais menos representativos.

No contexto deste trabalho, a escolha adequada do *batch size* é especialmente relevante, dado o volume de imagens disponíveis no ECPC-IDS. Em cenários com poucos dados rotulados, como é comum em aplicações médicas, lotes muito grandes podem fazer com que o modelo memorize padrões específicos do conjunto de treinamento em vez de aprender representações generalizáveis. Por esse motivo, a definição desse parâmetro foi tratada como uma etapa experimental do trabalho, com o objetivo de identificar a configuração que melhor equilibra estabilidade de treinamento e capacidade de generalização para a tarefa de segmentação de lesões endometriais.

3.4.2 Métricas de Avaliação

Para analisar quantitativamente o desempenho dos modelos, foram utilizadas métricas baseadas na matriz de confusão calculada pixel a pixel. Neste contexto, cada pixel da imagem é tratado como uma instância de classificação, podendo ser categorizado em quatro situações distintas em relação à máscara de referência (*Ground Truth*):

- **Verdadeiro Positivo (VP):** Pixels corretamente classificados pelo modelo como pertencentes à região de interesse (lesão/tumor).
- **Falso Positivo (FP):** Pixels incorretamente classificados como lesão, quando na verdade pertencem ao fundo ou tecido saudável (alarme falso).
- **Falso Negativo (FN):** Pixels incorretamente classificados como fundo, quando na verdade pertencem à lesão (tumor não detectado).
- **Verdadeiro Negativo (VN):** Pixels corretamente classificados como não pertencentes à região de interesse.

Com base nessas definições, foram calculadas três métricas principais (Sensibilidade, Precisão e F1-Score), além da métrica de Erro utilizada no treinamento.

1. Sensibilidade (*Recall*): Mede a proporção de pixels de lesão que foram corretamente identificados pelo modelo. Em diagnóstico médico, esta é uma métrica crítica, pois indica a capacidade do sistema em não perder casos positivos (evitar falsos negativos). É calculada conforme a Equação 3.1:

$$Sensibilidade = \frac{VP}{VP + FN} \quad (3.1)$$

2. Precisão (*Precision*): Avalia a confiabilidade das detecções positivas. Indica a porcentagem de pixels classificados como tumor que realmente correspondem à doença, sendo fundamental para medir a taxa de alarmes falsos. É definida pela Equação 3.2:

$$Precisão = \frac{VP}{VP + FP} \quad (3.2)$$

3. F1-Score (Coeficiente Dice): Representa a média harmônica entre a Precisão e a Sensibilidade. O F1-Score é a métrica mais robusta para avaliar a segmentação em conjuntos de dados desbalanceados (onde o fundo é muito maior que a lesão), pois penaliza tanto o excesso quanto a falta de segmentação. É expresso pela Equação 3.3:

$$F1-Score = 2 \cdot \frac{Precisão \cdot Sensibilidade}{Precisão + Sensibilidade} = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (3.3)$$

4. Erro (*Loss*): Refere-se ao valor obtido pela função de custo *Binary Cross Entropy* (BCE), utilizada para treinar a rede. Diferentemente das métricas anteriores, quanto menor o valor do Erro, melhor o desempenho do modelo. O erro mede a divergência entre a probabilidade predita (p) e o rótulo real (y) para cada pixel, conforme a Equação 3.4:

$$Erro = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (3.4)$$

onde N é o número total de pixels, y_i é o valor real (0 ou 1) e p_i é a probabilidade predita pelo modelo.

4 Resultados

Esse capítulo apresenta e discute os resultados obtidos na etapa experimental do trabalho. O objetivo principal foi avaliar a eficácia da Swin-Unet na segmentação de lesões de câncer de endométrio, e comparar com modelos U-Net aplicados a imagens de tomografia computadorizada.

4.1 Análise do impacto do tamanho do *batch size*

Para investigar como *batch size* influencia o processo de segmentação realizado pela arquitetura Swin-Unet, foram realizados experimentos utilizando cinco configurações: 4, 8, 12, 16 e 32. A escolha desses valores foi pela necessidade de observar o comportamento do modelo em cenários de lotes pequenos, intermediários e maiores. Tal abordagem é de acordo com estudos como Radiuk (2017) e Kandel e Castelli (2020), que destacam que diferentes tamanhos de lote podem alterar de forma significativa a estabilidade do treinamento, a velocidade de convergência e a capacidade de generalização do modelo.

Os testes foram realizados sob as mesmas condições, alterando apenas o tamanho do lote. Ao final do treinamento, foram registradas as métricas de Erro (*Test Loss*), Precisão, Sensibilidade e *F1-Score*, para avaliar o desempenho global de cada configuração. Os resultados obtidos estão apresentados na Tabela 2.

Tabela 2 – Resultados dos Experimentos na Swin-Unet, com Diferentes Tamanhos de Lote

Batch Size	Erro	Precisão	Sensibilidade	F1-Score	Nº de Épocas
4	0.0040	0.8294	0.8449	0.8370	55
8	0.0039	0.7792	0.8256	0.8017	35
12	0.0059	0.8257	0.6004	0.6953	36
16	0.0037	0.8526	0.8467	0.8496	55
32	0.0046	0.8243	0.7971	0.8105	49

Analisando os resultados, é possível observar que o *batch size* de valor 16 apresentou o melhor desempenho global, alcançando a menor perda no conjunto de teste (0.0037) e as maiores taxas em todas as métricas, resultando o maior F1-Score (0.8496). Esse resultado sugere que o lote 16 proporcionou o equilíbrio ideal, suficiente para atuar como um regularizador e evitar o *overfitting*, e estável o bastante para garantir uma convergência precisa. Esse comportamento está alinhado com a literatura Kandel e Castelli (2020), que aponta que tamanhos intermediários de lote favorecem a estabilidade do treinamento e a capacidade de generalização.

Nos experimentos com lotes menores (4 e 8), observou-se um comportamento de alta variância. Embora o *batch size* 4 tenha alcançado uma Sensibilidade elevada (0.8449), o comportamento mais instável dos gradientes levou a uma Precisão menor e exigiu mais épocas de treinamento para convergir (55 épocas). Já o *batch size* 12 apresentou uma anomalia significativa, com a pior Sensibilidade (0.6004) e o menor F1-Score, o que indica que o modelo provavelmente convergiu para uma solução subótima, com desempenho limitado na detecção das lesões.

Por fim, o *batch size* 32, apesar da estabilidade no treinamento, não superou o desempenho do lote 16, obtendo um F1-Score de 0.8105. Esse comportamento é consistente com discussões propostas por [Kandel e Castelli \(2020\)](#), [Radiuk \(2017\)](#), que alertam que lotes maiores tendem a estimar gradientes muito suaves, levando o modelo a convergir para mínimos agudos que funcionam bem nos dados de treino, mas generalizam pior para dados de teste não vistos.

Portanto, com base nos resultados o *batch size* 16 foi validado como a configuração ideal para a etapa final de treinamento da Swin-Unet neste trabalho, por proporcionar estabilidade, eficiência e desempenho nas métricas de segmentação utilizadas.

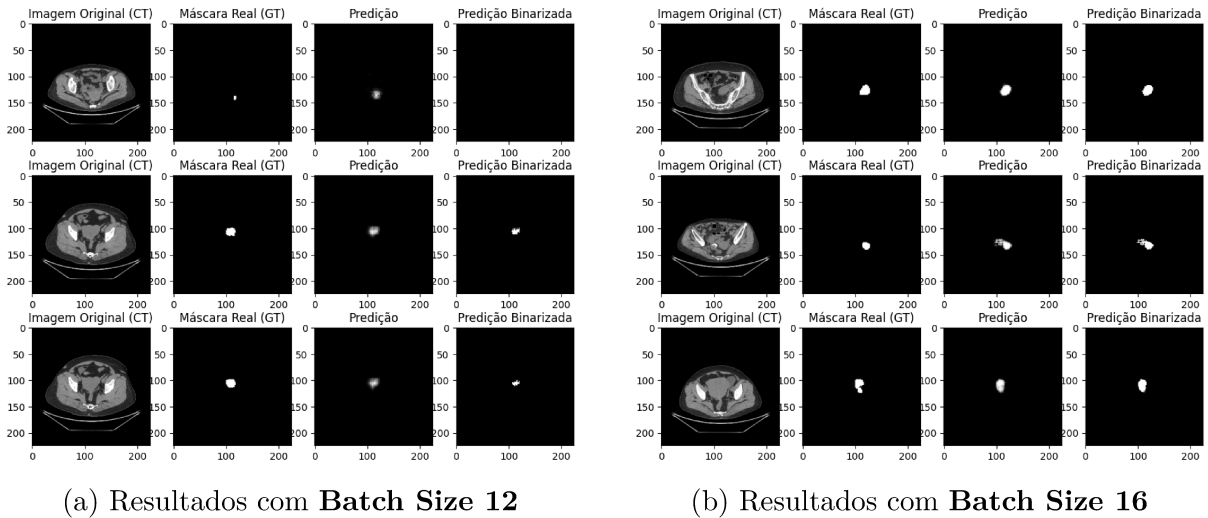


Figura 8 – Comparação das segmentações obtidas pela Swin-Unet.

Fonte: Do Autor

A análise detalhada dos dados apresentados na Tabela 2 é reforçada pelas predições mostradas na Figura 8. Devido à aleatoriedade na divisão do conjunto de teste, as imagens utilizadas nas Figuras 8a e 8b correspondem a pacientes diferentes. Ainda assim, foram selecionadas por representarem de forma consistente o comportamento típico observado em cada configuração.

Na Figura 8a (Batch Size 12, associado ao pior desempenho), é possível observar na terceira coluna (Predição) a formação de mapas de probabilidade com menor intensidade e contornos menos definidos. Esse padrão reflete maior incerteza do modelo, frequentemente

resultando em segmentações binárias incompletas ou na omissão de regiões menores da lesão, o que ajuda a explicar a baixa sensibilidade registrada (60,04%).

Em contrapartida, a Figura 8b (Batch Size 16, melhor desempenho) evidencia maior consistência nas predições. Os mapas gerados apresentam intensidades mais altas e limites mais nítidos, resultando em máscaras binarizadas (quarta coluna) mais próximas do *Ground Truth* (segunda coluna). Esse comportamento confirma a robustez dessa configuração e justifica os melhores valores obtidos de F1-Score (84,96%) e Sensibilidade (84,67%).

4.2 Desempenho da Swin-Unet

Para garantir que os resultados fossem corretos, o treinamento da Swin-Unet foi repetido cinco vezes. Em todas as execuções, foram mantidas exatamente as mesmas configurações e a mesma divisão dos dados descritas no Capítulo 3.

A Tabela 3 detalha como o modelo se comportou em cada uma das execuções. A arquitetura possui uma convergência estável, mesmo na execução com menor desempenho (Execução 3), e o modelo ainda apresentou uma sensibilidade superior a 85%. A Execução 2 se destacou com o melhor desempenho geral, atingindo o maior F1-Score (83%) e a maior detecção de lesões, com Sensibilidade de quase 87%.

Tabela 3 – Resultados das cinco execuções da Swin-Unet.

Execução	Erro	Precisão	Sensibilidade	F1-Score
1	0.0044	0.8268	0.7952	0.8107
2	0.0042	0.7969	0.8691	0.8314
3	0.0040	0.7320	0.8597	0.7907
4	0.0062	0.7954	0.8034	0.7994
5	0.0045	0.8444	0.8199	0.8320

A Tabela 4 traz a média e o desvio padrão das métricas. O desvio padrão baixo, principalmente no F1-Score ($\pm 0,0186$) e no Erro ($\pm 0,0009$), indica que o modelo consegue aprender e convergir para um bom resultado de forma estável, sem oscilar muito entre os treinamentos.

Tabela 4 – Swin-Unet - Média $\pm$ Desvio Padrão

Métrica	Média $\pm$ Desvio Padrão
Erro	0.005 $\pm$ 0.001
Precisão	0.799 $\pm$ 0.041
Sensibilidade	0.830 $\pm$ 0.030
F1-Score	0.813 $\pm$ 0.018

Hossin e M.N (2015) apontam que no diagnóstico de câncer, a métrica de Sensibilidade é mais priorizada que a Precisão, pois o custo associado a um Falso Negativo (não

diagnosticar um paciente doente) é maior do que o de um Falso Positivo (alarme falso). Com uma sensibilidade média de aproximadamente 83%, a Swin-Unet demonstrou uma capacidade superior de recuperar a área tumoral. Isso reforça a hipótese que o uso de *Transformers* e mecanismos de atenção permite que a rede compreenda o contexto global da imagem.

Para a análise visual do desempenho do modelo, foi selecionada a melhor execução (Execução 2), utilizando a Matriz de Confusão e a Curva ROC-AUC. As Figuras 9 e 10 apresentam, respectivamente, os resultados obtidos.

A análise conjunta da Matriz de Confusão e da Curva ROC mostra o alto poder discriminativo da Swin-Unet. A Matriz de Confusão (Figura 9) apresenta uma taxa de Verdadeiros Positivos de 86,91%, em concordância com a sensibilidade mostrada anteriormente e indicando que o modelo foi capaz de segmentar a maior parte da região tumoral. Os erros se concentram nos Falsos Negativos (13,09%), enquanto a taxa de Falsos Positivos foi de apenas 0,11%.

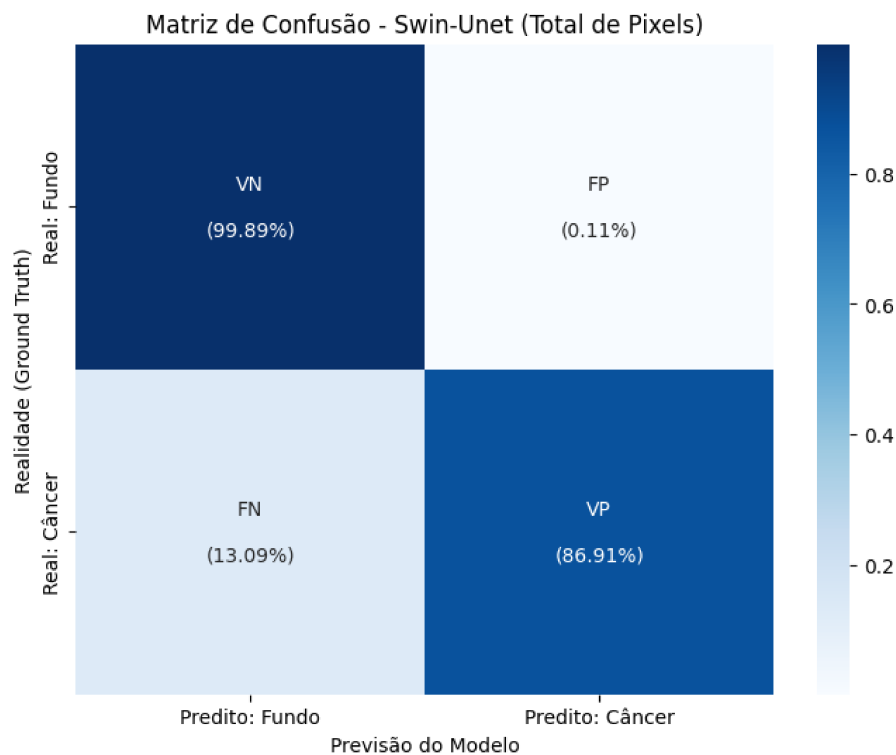


Figura 9 – Matriz de Confusão - Swin-Unet

Fonte: Do Autor

Em complemento, a Curva ROC (Figura 10) apresenta uma Área Sob a Curva (AUC) de 0,9995, reforçando a elevada capacidade do modelo de identificar as áreas. A curva próxima ao canto superior esquerdo indica que a Swin-Unet mantém altas taxas de verdadeiros positivos mesmo com níveis de falsos positivos baixo.

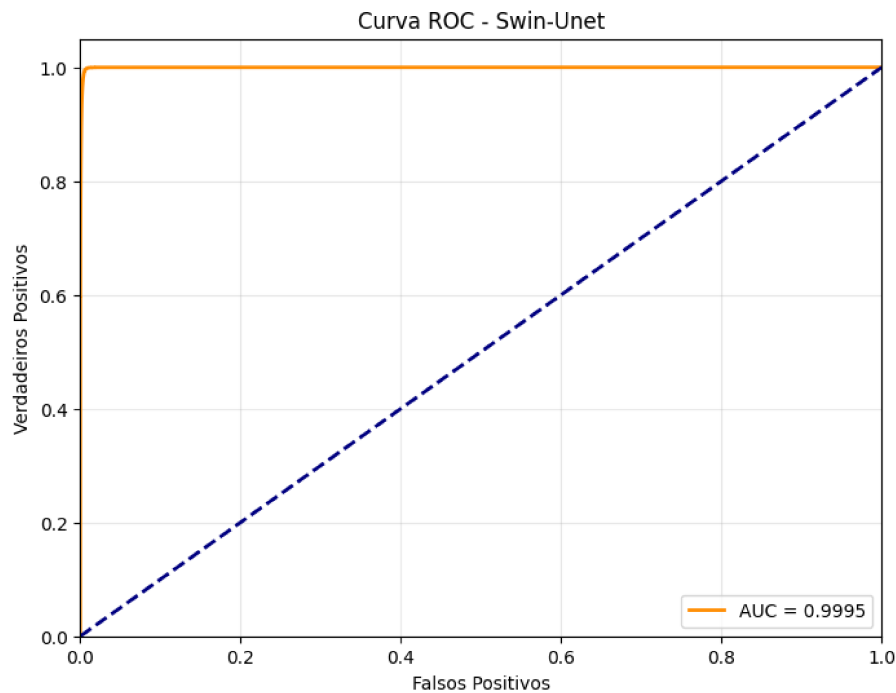


Figura 10 – Curva ROC - Swin-Unet

Fonte: Do Autor

4.3 Comparação de Desempenho entre Swin-Unet e U-Net

Para validar se a arquitetura proposta realmente é superior em relação aos métodos tradicionais, o desempenho da Swin-Unet foi comparado com duas CNNs U-Net com *backbone* VGG-16 e ResNet50.

Da mesma maneira que na Swin-Unet, para garantir uma comparação justa, todos os modelos passaram pelo mesmo protocolo de 5 execuções. A Tabela 5 apresenta a média e o desvio padrão das métricas resultantes.

Tabela 5 – Comparativo de desempenho entre as arquiteturas - Média $\pm$ Desvio Padrão de 5 execuções

Modelo	Erro	Precisão	Sensibilidade	F1-Score
ResNet50	0.006 ± 0.003	0.811 ± 0.128	0.683 ± 0.254	0.716 ± 0.197
VGG-16	0.005 ± 0.001	0.816 ± 0.044	0.794 ± 0.055	0.803 ± 0.027
Swin-Unet	0.005 ± 0.001	0.799 ± 0.041	0.830 ± 0.030	0.813 ± 0.018

A ResNet50 apresentou uma alta variabilidade entre as execuções, evidenciada pelo elevado desvio padrão da sensibilidade (25%), indicando instabilidade no processo de identificação do tumor. A VGG-16, mostrou resultados mais consistentes, com uma sensibilidade média de 79,4% e baixo desvio padrão. Já a Swin-Unet, obteve o melhor desempenho geral, alcançando a maior sensibilidade média (83%) e o maior F1-Score (81%), além do menor desvio padrão do F1-Score (0.018).

Para entender as diferenças de desempenho apresentadas na Tabela 5, foi analisado também o comportamento das redes através de suas matrizes de confusão e ROC-AUC.

A Figura 11 apresenta a Matriz de Confusão referente a uma das execuções instáveis da ResNet50. É possível observar uma proporção elevada de Falsos Negativos (FN), superior a 70%, indicando que o modelo teve dificuldade em identificar corretamente as regiões com tumor.

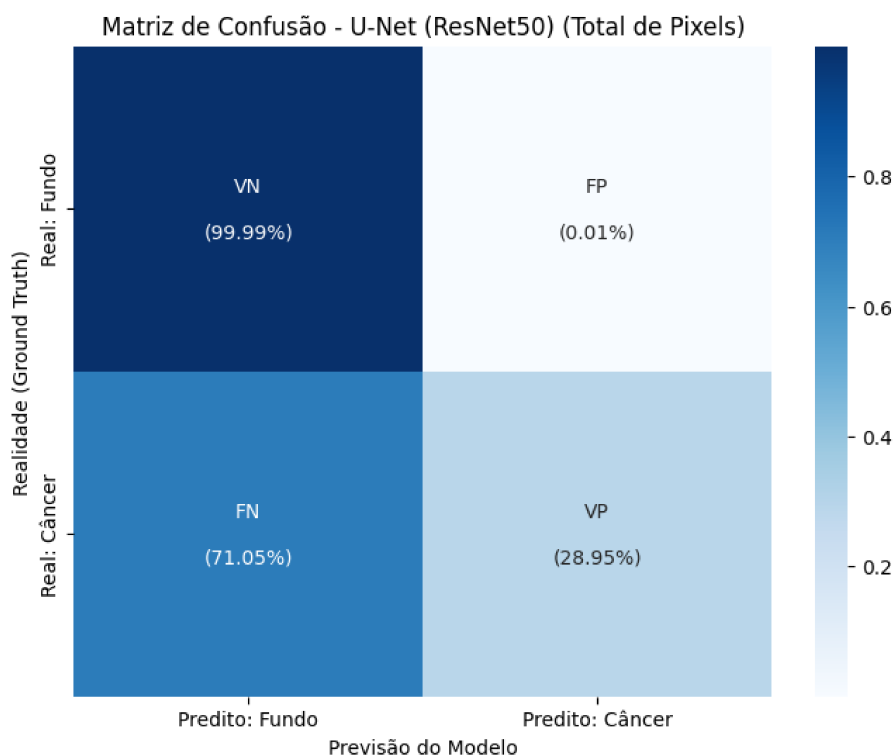


Figura 11 – Matriz de Confusão - ResNet50.

Fonte: Do Autor

Esse comportamento compromete a utilidade clínica do modelo, uma vez que grande parte dos tumores não é detectada. Esse padrão também pode ser observado na Curva ROC apresentada na Figura 12, que se mostra praticamente achatada, mostrando a baixa capacidade da rede em separar adequadamente as classes.

Por outro lado, a VGG-16 apresentou um comportamento mais equilibrado ao longo das execuções. Conforme ilustrado na Matriz de Confusão (Figura 13), o modelo foi capaz de identificar a presença da lesão na maioria dos casos, embora ainda seja observada uma taxa relevante de Falsos Negativos, variando entre 20% e 27%. Indicando que, apesar do bom desempenho geral, parte das regiões tumorais não foi completamente segmentada.

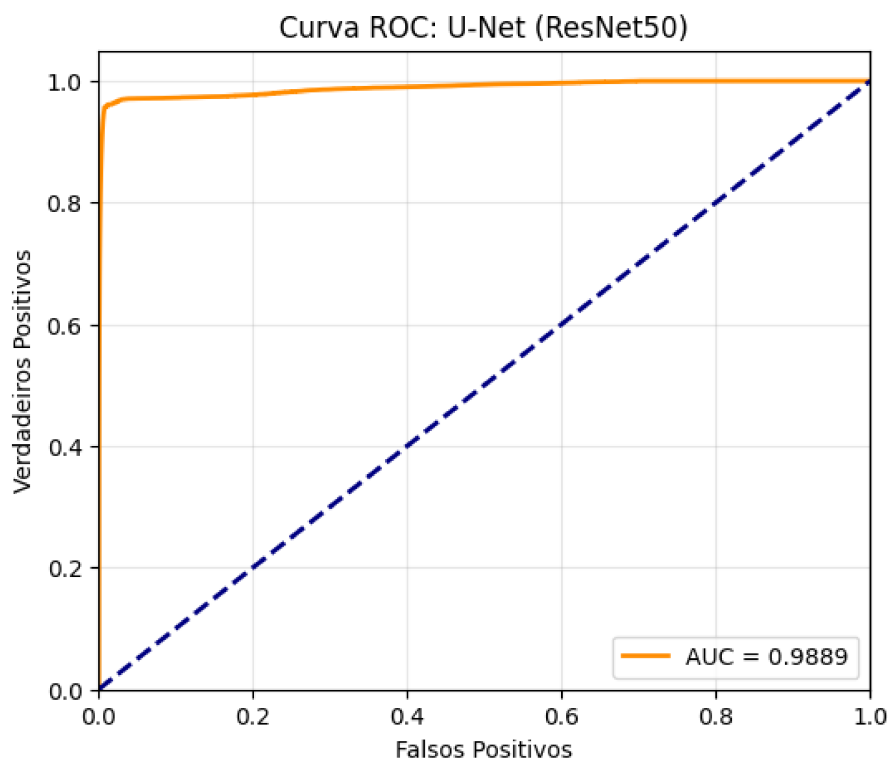


Figura 12 – Curva ROC - ResNet50.

Fonte: Do Autor

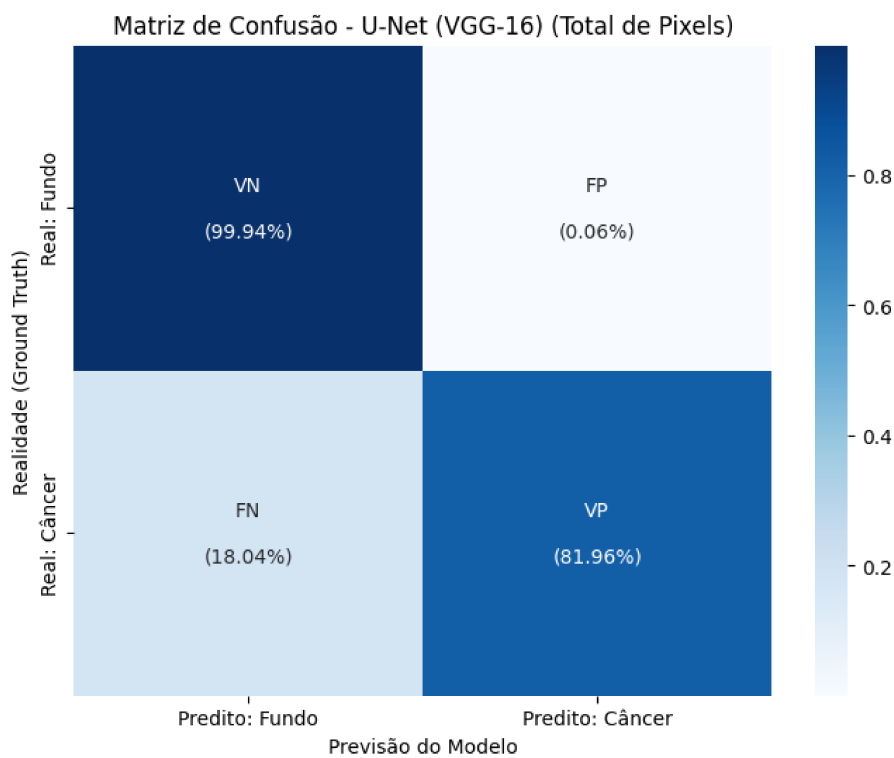


Figura 13 – Matriz de Confusão - VGG-16

Fonte: Do Autor

A Curva ROC (Figura 14) ilustra um formato consistente e desempenho superior ao observado nas execuções instáveis da ResNet50, porém ainda inferior ao de arquiteturas mais avançadas. Em conjunto, as duas análises indicam que a VGG-16 funciona adequadamente como um baseline estável, mas apresenta certas limitações na captura de detalhes mais finos das lesões.

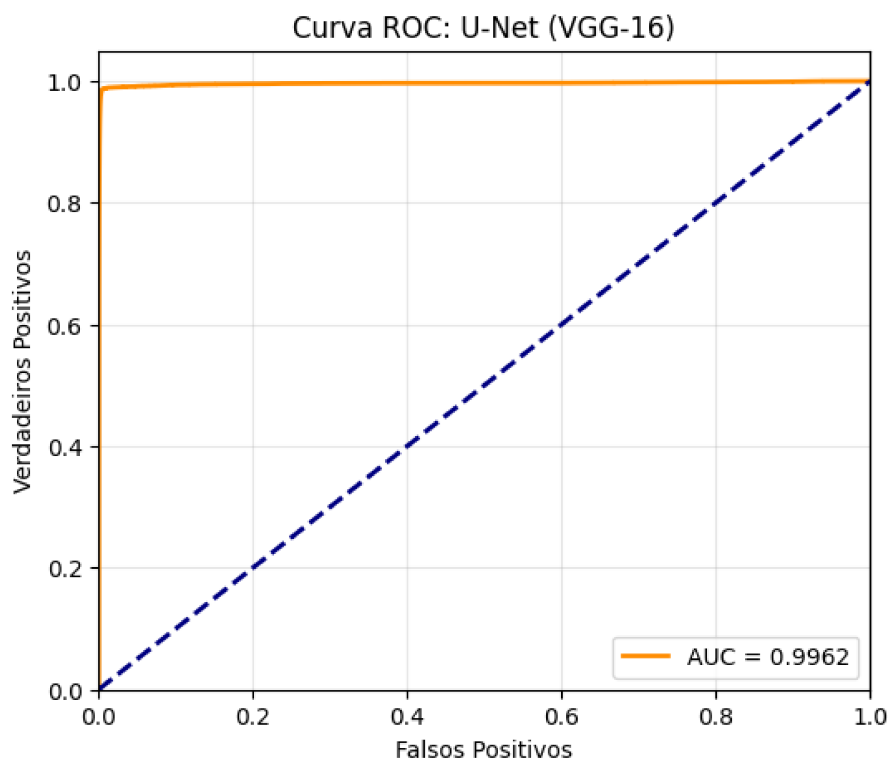


Figura 14 – Curva ROC - VGG-16

Fonte: Do Autor

As Figuras 15, 16 e 17 mostram exemplos dos resultados da segmentação obtidos pelos modelos. É importante ressaltar que, como já mencionado no capítulo anterior, devido à divisão aleatória do conjunto de teste em cada execução, as imagens apresentadas correspondem a pacientes distintos. Ainda assim, elas foram escolhidas por representarem o comportamento típico de cada arquitetura, bem como os principais padrões de acerto e erro observados nas métricas apresentadas anteriormente.

A Figura 15 ilustra uma limitação bastante comum observada na ResNet50. Com a elevada taxa de Falsos Negativos identificada anteriormente, é possível notar que o modelo tende a subestimar a extensão da lesão. No exemplo, a rede consegue identificar apenas uma pequena região do tumor, e não consegue acompanhar corretamente suas bordas, resultando em uma segmentação parcial da área.

A Figura 16 apresenta a VGG-16, em comparação com a ResNet50, o modelo demonstra uma melhor capacidade de localizar a lesão. No entanto, ao comparar a predição

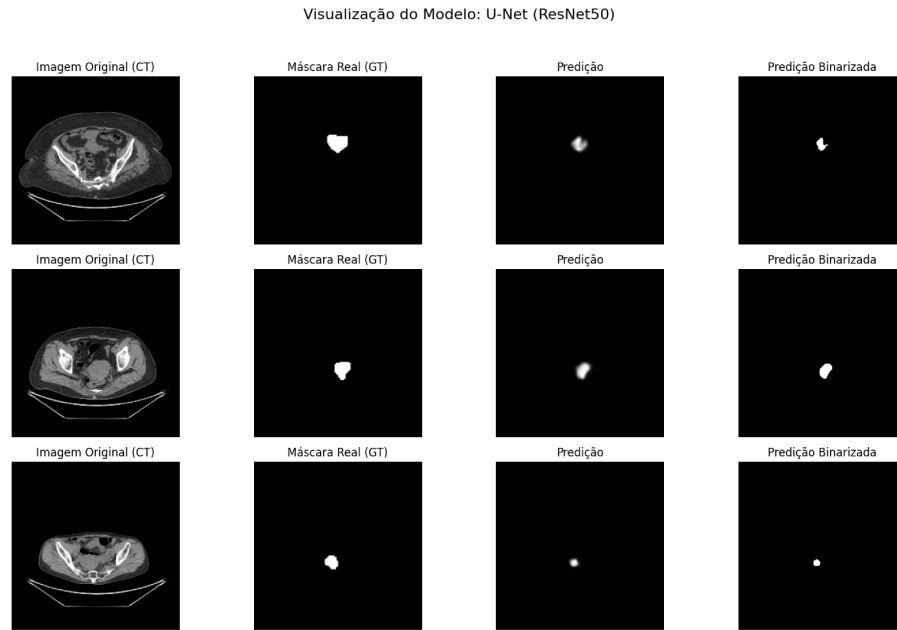


Figura 15 – Resultado Segmentação - ResNet50.

Fonte: Do Autor

com o GT, percebe-se que a segmentação simplifica o formato do tumor. As bordas aparecem mais suaves e arredondadas com perda de detalhes, indicando dificuldade do em capturar contornos mais finos e complexos.

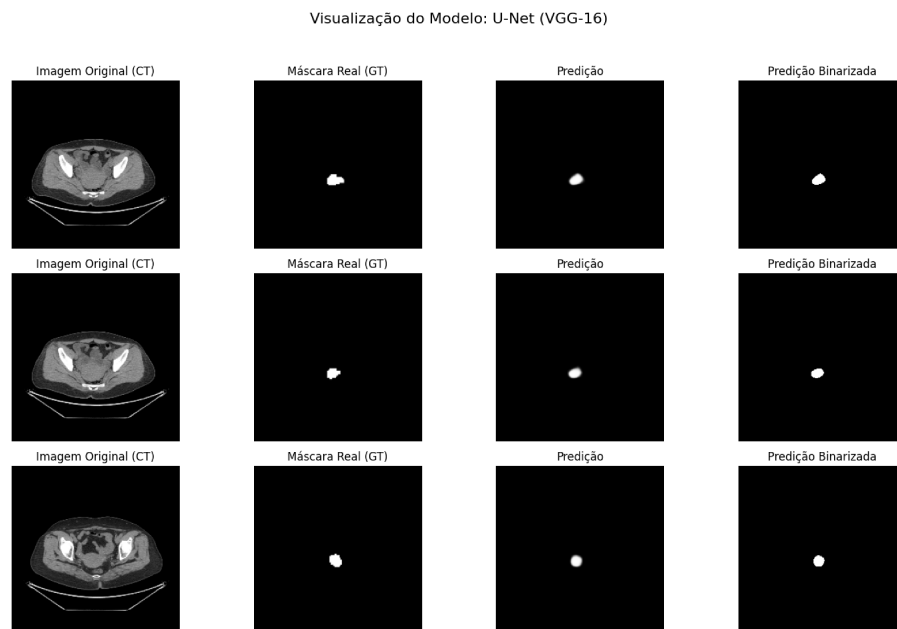


Figura 16 – Resultado Segmentação - VGG-16.

Fonte: Do Autor

Por fim, a Figura 17 exibe um resultado da arquitetura Swn-Unet. É possível observar uma aderência relativamente superior ao GT em comparação aos padrões das

CNNs. O modelo compreende melhor o contexto anatômico, resultando em uma segmentação que preenche o interior da lesão de forma correta e define suas bordas com uma boa fidelidade, superando a subestimação da ResNet50 quanto a perda de detalhes da VGG-16.

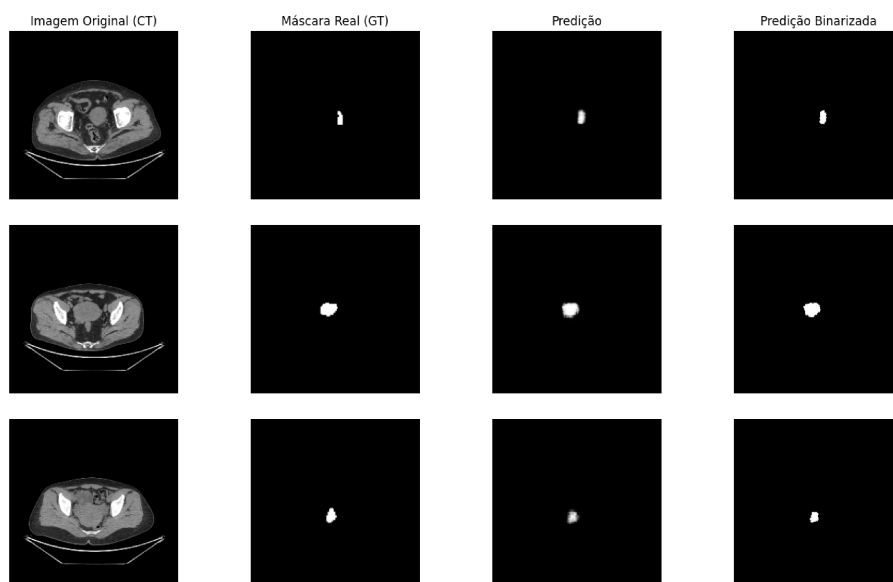


Figura 17 – Resultado Segmentação - Swin-Unet.

Fonte: Do Autor

5 Conclusão

A segmentação do câncer de endométrio em imagens de TC ainda é uma tarefa complexa e de bastante relevância, especialmente por seu papel no apoio ao diagnóstico precoce. Diante desse cenário, esse trabalho avaliou o uso da Swin-Unet, com objetivo de verificar se a capacidade do modelo de analisar informações em uma escala mais ampla da imagem, por meio do mecanismo de Atenção Global, poderia trazer ganhos em relação às CNNs clássicas. Para isso, o desempenho da Swin-Unet foi comparado ao de U-Nets com *backbones* VGG-16 e ResNet50.

Os experimentos foram conduzidos com múltiplas execuções independentes, o que possibilitou uma avaliação confiável dos modelos. A partir dos resultados obtidos, é possível destacar as seguintes conclusões:

- **Swin-Unet:** A arquitetura proposta apresentou os melhores resultados gerais, com uma Sensibilidade média de 83,0% e um F1-Score de 81,3%. Além dos valores médios, a Swin-Unet mostrou comportamento estável entre as execuções, com baixo desvio padrão, indicando maior confiabilidade do modelo.
- **VGG-16 e ResNet50:** A U-Net com ResNet50 mostrou instabilidade em algumas execuções, com falhas na detecção da lesão, refletidas pelo aumento de falsos negativos. Já a VGG-16 manteve um comportamento mais estável, mas com desempenho mais baixo, especialmente na delimitação precisa das bordas do tumor.

A análise visual das segmentações mostrou que o mecanismo de *Shifted Windows* da Swin-Unet contribuiu para melhor interpretação das imagens. Como resultado, as máscaras geradas pela Swin-Unet se aproximaram mais do GT, reduzindo a subestimação da área tumoral, um aspecto fundamental para minimizar o risco de diagnósticos incorretos.

Dessa forma, podemos concluir que o uso da Swin-Unet, representa um avanço relevante para a segmentação de câncer de endométrio em imagens de tomografia computadorizada. A arquitetura proposta demonstrou um melhor equilíbrio entre sensibilidade e precisão comparada às U-Nets tradicionais, reforçando seu potencial para aplicações em sistemas de apoio ao diagnóstico.

Algumas disciplinas do curso de Ciência da Computação apresentaram grande contribuição para o desenvolvimento do trabalho, como inteligência computacional e mineração de dados, que trouxe o conhecimento sobre redes neurais convolucionais e o processo de treinamento dessas redes, e inteligência artificial, que ajudou a entender os diferentes tipos de aprendizado de máquina.

Apesar de bons resultados, o estudo aponta possibilidades de aprofundamento em trabalhos futuros. Destaca-se a adaptação da Swin-Unet para o processamento de imagens 3D, afim de explorar a continuidade espacial entre as fatias da tomografia. Também é recomendada a ampliação da base de dados com imagens de pacientes com e sem tumor, permitindo avaliar não apenas a segmentação, mas também a capacidade de classificação dos modelos. Além disso, o uso de imagens de um mesmo paciente para a comparação entre as arquiteturas pode contribuir para análises mais controladas e consistentes, reduzindo a variabilidade entre os dados.

Referências

- AGGARWAL, K.; SINGHAL, A.; SINGH, S.; NAQVI, N. Z. Advancements in automated segmentation techniques for endometrial cancer: A comprehensive review. In: **2025 10th International Conference on Signal Processing and Communication (ICSC)**. [S.l.: s.n.], 2025. p. 218–227. Citado 2 vezes nas páginas 17 e 20.
- ANBUDEVI, M. Krithika alias; SUGANTHI, K. Review of semantic segmentation of medical images using modified architectures of unet. **Diagnostics**, v. 12, n. 12, 2022. ISSN 2075-4418. Disponível em: <<https://www.mdpi.com/2075-4418/12/12/3064>>. Citado 2 vezes nas páginas 14 e 15.
- BHATLA, N.; AOKI, D.; SHARMA, D. N.; SANKARANARAYANAN, R. Cancer of the cervix uteri. **International Journal of Gynecology & Obstetrics**, v. 143, n. S2, p. 22–36, 2018. Disponível em: <<https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/ijgo.12611>>. Citado na página 20.
- CAO, H.; WANG, Y.; CHEN, J.; JIANG, D.; ZHANG, X.; TIAN, Q.; WANG, M. **Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation**. 2021. Disponível em: <<https://arxiv.org/abs/2105.05537>>. Citado 9 vezes nas páginas 9, 15, 17, 18, 20, 23, 24, 25 e 27.
- CHAVES, C.; NICOLAU-NETO, P.; SIMÃO, T.; SOUZA-SANTOS, P. d.; BERGMANN, A.; BREWER, L.; MOREIRA, F.; REIS, B.; MOREIRA, M.; PINTO, L. Esr1 overexpression is a biomarker of relapse and worse prognosis in stage i endometrioid endometrial carcinoma. **Brazilian Journal of Medical and Biological Research**, Associação Brasileira de Divulgação Científica, 2025. Disponível em: <<https://doi.org/10.1590/1414-431X2025e14494>>. Citado na página 9.
- DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S.; USZKOREIT, J.; HOULSBY, N. **An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale**. 2021. Disponível em: <<https://arxiv.org/abs/2010.11929>>. Citado na página 17.
- DU, T.; JIANG, T.; LI, X.; RAHAMAN, M. M.; GRZEGORZEK, M.; LI, C. Prediction of tp53 mutations across female reproductive system pan-cancers using deep multimodal pet/ct radiogenomics. **Frontiers in Medicine**, Volume 12 - 2025, 2025. ISSN 2296-858X. Disponível em: <<https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2025.1608652>>. Citado na página 10.
- FARIAS, M.; FIGUEREDO, W.; SILVA, A.; PAIVA, A.; SALOMÃO, A.; OLIVEIRA, M. Segmentação de endometriose profunda em imagens de ressonância magnética utilizando redes convolucionais como codificadores para arquitetura u-net. In: **Anais do XXV Simpósio Brasileiro de Computação Aplicada à Saúde**. Porto Alegre, RS, Brasil: SBC, 2025. p. 874–884. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/35553>>. Citado 2 vezes nas páginas 14 e 15.

- FIDELIS, L. M. N. d. V.; COSTA, L. L. Câncer de endométrio - diagnóstico precoce e estratégias de prevenção. **Brazilian Journal of Health Review**, v. 7, n. 1, p. 1343–1355, Jan. 2024. Disponível em: <<https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/66345>>. Citado 2 vezes nas páginas 13 e 14.
- FIGUEREDO, W.; SILVA, I.; DINIZ, J.; SILVA, A.; PAIVA, A.; SALOMÃO, A.; OLIVEIRA, M. Abordagem computacional baseada em deep learning para o diagnóstico de endometriose profunda através de imagens de ressonância magnética. In: **Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde**. Porto Alegre, RS, Brasil: SBC, 2023. p. 138–149. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/25284>>. Citado na página 9.
- GHOSH, S.; CHAKI, A.; SANTOSH, K. Improved u-net architecture with vgg-16 for brain tumor segmentation. **Physical and Engineering Sciences in Medicine**, v. 44, 05 2021. Citado na página 26.
- GORGENS, E. B.; LEITE, H. G.; SANTOS, H. d. N.; GLERIANI, J. M. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, Sociedade de Investigações Florestais, v. 33, n. 6, p. 1141–1147, nov. 2009. ISSN 0100-6762. Disponível em: <<https://doi.org/10.1590/S0100-67622009000600016>>. Citado na página 22.
- GROUP, M. ECPC-IDS. 7 2023. Disponível em: <<https://figshare.com/articles/dataset/ECPC-IDS/23808258>>. Citado na página 20.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. **Deep Residual Learning for Image Recognition**. 2015. Disponível em: <<https://arxiv.org/abs/1512.03385>>. Citado na página 27.
- HOSSIN, M. bin; M.N, S. A review on evaluation metrics for data classification evaluations. **International Journal of Data Mining Knowledge Management Process**, v. 5, p. 01–11, 03 2015. Citado na página 33.
- JAIN, A.; BHARDWAJ, A.; MURALI, K.; SURANI, I. **A Comparative Study of CNN, ResNet, and Vision Transformers for Multi-Classification of Chest Diseases**. 2024. Disponível em: <<https://arxiv.org/abs/2406.00237>>. Citado na página 27.
- KANDEL, I.; CASTELLI, M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. **ICT Express**, v. 6, n. 4, p. 312–315, 2020. ISSN 2405-9595. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2405959519303455>>. Citado 4 vezes nas páginas 28, 29, 31 e 32.
- KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. 2017. Disponível em: <<https://arxiv.org/abs/1412.6980>>. Citado na página 28.
- KURATA, Y.; NISHIO, M.; MORIBATA, Y.; KIDO, A.; HIMOTO, Y.; OTANI, S.; FUJIMOTO, K.; YAKAMI, M.; MINAMIGUCHI, S.; MANDAI, M.; NAKAMOTO, Y. Automatic segmentation of uterine endometrial cancer on multi-sequence mri using a convolutional neural network. **Scientific Reports**, v. 11, p. 14440, 07 2021. Citado 4 vezes nas páginas 9, 10, 15 e 18.

- 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: **2021 IEEE/CVF International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2021. p. 9992–10002. Citado 3 vezes nas páginas 10, 17 e 24.
- MINAEE, S.; BOYKOV, Y.; PORIKLI, F.; PLAZA, A.; KEHTARNAVAZ, N.; TERZOPOULOS, D. Image segmentation using deep learning: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 44, n. 7, p. 3523–3542, 2022. Citado na página 9.
- NASCIMENTO, O. L. do. **Estudo comparativo entre modelos transformers aplicados ao desenvolvimento de chatbots**. 2023. Disponível em: <<https://repositorio.ufrn.br/handle/123456789/58117>>. Citado na página 16.
- OLIVEIRA, M.; MATOS, C.; JÚNIOR, G. B.; ALMEIDA, J. Dual-scale se unet: Uma rede neural convolucional aprimorada com se blocks para segmentação renal em imagens de tc. In: **Anais do XXV Simpósio Brasileiro de Computação Aplicada à Saúde**. Porto Alegre, RS, Brasil: SBC, 2025. p. 760–771. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/35543>>. Citado na página 15.
- PAULINO, E.; NOGUEIRA-RODRIGUES, A.; GOSS, P. E.; FARONI, L.; GUITMANN, G.; STRASSER-WEIPPL, K.; BUKOWSKI, A. Endometrial cancer in brazil: Preparing for the rising incidence. **Revista Brasileira de Ginecologia e Obstetrícia**, Federação Brasileira das Sociedades de Ginecologia e Obstetrícia, v. 40, n. 10, p. 577–579, Oct 2018. ISSN 0100-7203. Disponível em: <<https://doi.org/10.1055/s-0038-1673644>>. Citado 2 vezes nas páginas 13 e 14.
- PENG, R.; ZENG, Q.; HE, B.; LIU, J.; XIAO, Z. Endometrial cancer lesion region segmentation based on large kernel convolution and combined attention. **Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi**, v. 42, p. 928–935, 10 2025. Citado na página 10.
- PINTO, D.; FIGUEREDO, W.; SILVA, I.; SILVA, A.; PAIVA, A.; SALOMÃO, A.; OLIVEIRA, M. Segmentação automática de endometriose profunda em imagens de ressonância magnética baseada em swin-unet. In: **Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde**. Porto Alegre, RS, Brasil: SBC, 2024. p. 471–482. ISSN 2763-8952. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/28841>>. Citado 5 vezes nas páginas 9, 11, 17, 24 e 25.
- RADIUK, P. Impact of training set batch size on the performance of convolutional neural networks for diverse datasets. **Information Technology and Management Science**, v. 20, p. 20–24, 12 2017. Citado 3 vezes nas páginas 28, 31 e 32.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-Net: Convolutional Networks for Biomedical Image Segmentation**. 2015. Disponível em: <<https://arxiv.org/abs/1505.04597>>. Citado 2 vezes nas páginas 15 e 25.
- SAIFULLAH, S.; DREŻEWSKI, R.; YUDHANA, A.; SURYOTOMO, A. P. Automatic brain tumor segmentation: Advancing u-net with resnet50 encoder for precise medical image analysis. **IEEE Access**, v. 13, p. 43473–43489, 2025. Citado na página 26.

- SHAZIYA, H.; SHYAMALA, K. Pulmonary ct images segmentation using cnn and unet models of deep learning. In: **2020 IEEE Pune Section International Conference (PuneCon)**. [S.l.: s.n.], 2020. p. 195–201. Citado 2 vezes nas páginas 14 e 15.
- 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**, v. 35, n. 5, p. 1285–1298, 2016. Citado na página 26.
- SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. 2015. Disponível em: <<https://arxiv.org/abs/1409.1556>>. Citado na página 26.
- TANG, D.; LI, C.; DU, T.; JIANG, H.; MA, D.; MA, Z.; GRZEGORZEK, M.; JIANG, T.; SUN, H. Ecpc-ids: A benchmark endometrial cancer pet/ct image dataset for evaluation of semantic segmentation and detection of hypermetabolic regions. **Computers in Biology and Medicine**, v. 171, p. 108217, 2024. ISSN 0010-4825. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0010482524003019>>. Citado 5 vezes nas páginas 10, 18, 20, 21 e 28.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. u.; POLOSUKHIN, I. Attention is all you need. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>. Citado 3 vezes nas páginas 5, 15 e 16.
- WANG, W.; CAI, Y.; GUO, Z.; ZHAO, A.; MA, W.; WANG, W.; WANG, S.; ZHU, X.; DU, X.; SHEN, W. A deep learning-based computer-aided diagnosis system for detecting atypical endometrial hyperplasia and endometrial cancer through hysteroscopy, *iscience*. **iScience**, 07 2025. Citado na página 9.