



UNIVERSIDADE FEDERAL DO MARANHÃO

Programa de Pós-Graduação em Ciência da Computação

Daniel Moreira Pinto

***Segmentação Automática de Endometriose Profunda em
Imagens de Ressonância Magnética Baseada em Aprendizado
Profundo***

**São Luís
2025**

Daniel Moreira Pinto

Segmentação Automática de Endometriose Profunda em Imagens de Ressonância Magnética Baseada em 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. Aristófanés Corrêa Silva

São Luís - MA

2025

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

Moreira Pinto, Daniel.

Segmentação Automática de Endometriose Profunda em
Imagens de Ressonância Magnética Baseada em Aprendizado
Profundo / Daniel Moreira Pinto. - 2025.

80 f.

Orientador(a): Aristófanês Corrêa Silva.

Dissertação (Mestrado) - Programa de Pós-graduação em
Ciência da Computação/ccet, Universidade Federal do
Maranhão, Auditório do Nca Ufma, São Luís, Ma, 2025.

1. Endometriose. 2. Ressonância Magnética. 3.
Aprendizado Profundo. 4. Processamento de Imagens. 5.
Segmentação. I. Corrêa Silva, Aristófanês. II. Título.

Daniel Moreira Pinto

Segmentação Automática de Endometriose Profunda em Imagens de Ressonância Magnética Baseada em 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 APROVADO. São Luís - MA, 17 de Março de 2025:

Prof. Dr. Aristófanês Corrêa Silva
Orientador
Universidade Federal do Maranhão

Prof. Dr. Anselmo Cardoso de Paiva
Examinador Interno
Universidade Federal do Maranhão

Prof. Dr. Luana Batista da Cruz
Examinador Externo
Universidade Federal do Cariri

São Luís - MA
2025

Dedico este trabalho ao conhecimento científico público.

Agradecimentos

Agradeço à minha família por sempre ter apoiado meu desenvolvimento educacional e me fornecido todas as condições para dedicar-me aos estudos. Agradeço a meus amigos por me acompanharem e ajudarem durante toda a minha trajetória acadêmica. Agradeço aos meus professores, gigantes que permitiram-me subir em seus ombros para olhar novos horizontes em minha carreira. Um agradecimento especial a meu orientador, professor Dr. Aristófanês Corrêa Silva, por esses anos pavimentando meu caminho na Academia e me direcionando ao sucesso.

*"Se o escritor dessas linhas teve sucesso em prover algum material para o esclarecimento
deste problema, ele poderá perceber seu trabalho como frutífero."*

(Vladimir Ilitch Ulianov)

Resumo

A endometriose é a doença caracterizada pela presença do endométrio fora da cavidade uterina. Ela é declarada profunda quando a lesão se estende por mais de 5 milímetros. Essa doença atinge dezenas de milhões de mulheres de várias faixas etárias ao redor do mundo. Ela afeta a qualidade de vida das pacientes e pode apresentar muitos sintomas diferentes, além de haver a possibilidade de evoluir para problemas mais sérios. Dessa forma, o diagnóstico eficaz da doença é essencial para iniciar o tratamento. Uma das principais áreas de ocorrência de lesões devido à endometriose é o reto e sigmoide. Através da ressonância magnética (RM), especialistas podem avaliar a presença da doença de forma não invasiva, porém a um alto custo de tempo, devido às lesões apresentarem-se em diversos formatos, tamanhos e regiões de ocorrência. Assim, esse método de diagnóstico é extremamente desafiador. Neste contexto, métodos computacionais podem auxiliar os especialistas no diagnóstico, apresentando uma indicação da possível ocorrência das lesões. Porém, a literatura sobre a segmentação da endometriose em RM utilizando aprendizado profundo é extremamente limitada. Portanto, este trabalho propôs um método baseado na aplicação da Swin-Unet, uma rede de aprendizado profundo baseada em *transformer*, para a segmentação da endometriose profunda em imagens de RM do reto e sigmoide. Além disso, é proposta uma etapa de refinamento da segmentação, baseada em processamento de imagens, para compor o método. Foram obtidos uma precisão de 45,6%, sensibilidade 61,9%, Dice de 47,7% e Jaccard de 36,2%. Foi segmentada com boa qualidade ao menos uma imagem por paciente em 17 dos 18 pacientes utilizados para teste. Assim, o método desenvolvido demonstra-se promissor para auxiliar o diagnóstico da doença.

Palavras-chave: Endometriose, Ressonância Magnética, Aprendizado Profundo, Processamento de Imagens, Segmentação.

Abstract

Endometriosis is a disease characterized by the presence of endometrial tissue outside the uterine cavity. It is classified as deep endometriosis when the lesion extends more than 5 millimeters. This condition affects tens of millions of women of various age groups worldwide. It impacts patients' quality of life and can present a wide range of symptoms, with the potential to progress into more serious complications. Therefore, effective diagnosis of the disease is crucial for initiating treatment. One of the primary areas where endometriosis-related lesions occur is the rectum and sigmoid colon. Through magnetic resonance imaging (MRI), specialists can assess the presence of the disease non-invasively, but at a high time cost, as the lesions can appear in various shapes, sizes, and locations. Thus, this diagnostic method is extremely challenging. In this context, computational methods can assist specialists in diagnosis by providing an indication of the possible presence of lesions. However, the literature on deep learning-based segmentation of endometriosis in MRI is extremely limited. Therefore, this study proposes a method based on the application of Swin-Unet, a deep learning network based on *transformers*, for the segmentation of deep endometriosis in MRI images of the rectum and sigmoid colon. Additionally, a refinement step for the segmentation, based on image processing, is proposed to complement the method. The results achieved were an accuracy of 45.6%, sensitivity of 61.9%, Dice coefficient of 47.7%, and Jaccard index of 36.2%. At least one image per patient was segmented with good quality in 17 out of the 18 patients used for testing. Thus, the developed method shows promise in assisting the diagnosis of the disease.

Keywords: Endometriosis, Magnetic Resonance Imaging, Deep Learning, Image Processing, Segmentation.

Lista de ilustrações

Figura 1 – Representação da endometriose na pelve. 1. Peritônio; 2. útero; 3. vagina; 4. cólon sigmoide; 5. lesão de endometriose; 6. fundo de posterior; 7. retossigmoide; 8. reto; 9. septoretovaginal.	28
Figura 2 – Ilustração do fluxo da menstruação retrógrada.	29
Figura 3 – Principais estruturas pélvicas. (a) Imagem de RM (b) Imagem com estruturas pélvicas destacadas: bexiga em verde, útero em lilás, sigmoide em vermelho e reto em azul.	30
Figura 4 – Exemplo de lesões de endometriose. (a), (b) e (c) são as imagens originais. (d), (e) e (f) são as respectivas imagens com as lesões marcadas.	30
Figura 5 – Equipamento para a captura de RM.	31
Figura 6 – Neurônio biológico.	35
Figura 7 – Neurônio artificial.	35
Figura 8 – Ilustração das definições aprendidas pelas camadas de uma arquitetura de aprendizado profundo.	37
Figura 9 – Arquitetura da CNN LeNet-5.	39
Figura 10 – Arquitetura da U-Net.	40
Figura 11 – Atenção por produto escalar escalonado.	42
Figura 12 – Autoatenção de múltiplas cabeças	42
Figura 13 – Arquitetura da Transformer.	43
Figura 14 – Arquitetura da ViT e sua camada de codificação.	44
Figura 15 – Arquitetura da Swin Transformer e sua camada de codificação.	46
Figura 16 – Arquitetura da Swin-Unet.	47
Figura 17 – Esquema do método proposto.	49
Figura 18 – Fatias consecutivas de uma paciente, que apresentam lesão. As marcações das lesões pelo especialista estão em verde.	50
Figura 19 – Imagem da base com a região de interesse marcada em azul.	50
Figura 20 – <i>Template</i> utilizado na etapa de <i>template matching</i>	51
Figura 21 – Exemplo de ROI primária binarizada.	51
Figura 22 – Exemplo de ROI refinada.	52
Figura 23 – Arquitetura da Swin-Unet.	53
Figura 24 – Máscaras binárias das imagens com lesão de um paciente e sua sobreposição.	55
Figura 25 – Remoção dos falsos positivos.	55
Figura 26 – Simulação do comportamento da eliminação de falsos positivos.	56
Figura 27 – Protocolo de marcação da base de imagens.	59
Figura 28 – Imagens e suas respectivas ROIs delimitadas pelo método.	60
Figura 29 – Exemplos da segmentação inicial das lesões.	61

Figura 30 – Melhores segmentações de diferentes pacientes da base.	62
Figura 31 – Estudo de Caso 1	63
Figura 32 – Estudo de Caso 2.	66
Figura 33 – Estudo de Caso 3.	67
Figura 34 – Estudo de Caso 4.	68
Figura 35 – Estudo de Caso 5.	68

Lista de tabelas

Tabela 1	–	Visão geral dos trabalhos relacionados. “N/A” significa que a categoria não se aplica.	27
Tabela 2	–	Resultado médio da predição do modelo sobre o conjunto de teste. . .	62
Tabela 3	–	Resultado médio do <i>k-fold cross-validation</i> de cada modelo sobre o conjunto de teste.	64
Tabela 4	–	Comparação de resultados da etapa de segmentação das lesões com a de eliminação de falsos positivos	65
Tabela 5	–	Resultados de erro e acerto por imagem.	66
Tabela 6	–	Comparação quantitativa do método proposto com o método de Figueredo et al. (2024).	69

Lista de abreviaturas e siglas

CAD	<i>Diagnóstico Auxiliado por Computador</i>
CNN	<i>Rede Neural Convolucional</i>
C3D	<i>Convolutional 3D</i>
Det	<i>Detecção</i>
ENID	<i>Endometrial Implants Dataset</i>
FN	<i>Falso Negativo</i>
FP	<i>Falso Positivo</i>
GLEND A	<i>Gynecologic Laparoscopy Endometriosis Dataset</i>
GT	<i>Ground Truth</i>
mAP	<i>Média da Precisão Média</i>
MLP	<i>Multilayer Perceptron</i>
MSU _{net}	<i>Multi-scale Sensitive U-Net</i>
N/A	<i>Não se Aplica</i>
R-CNN	<i>Region-based Convolutional Neural Network</i>
ResNet	<i>Residual Neural Network</i>
RM	<i>Ressonância Magnética</i>
RNA	<i>Rede Neural Artificial</i>
ROI	<i>Região de Interesse</i>
Seg	<i>Segmentação</i>
SW-MSA	<i>Shifted Window-based Multi-head Self-attention</i>
USTV	<i>Ultrassonografia Transvaginal</i>
VGG	<i>Visual Geometry Group</i>
ViT	<i>Vision Transformer</i>

VN	<i>Verdadeiro Negativo</i>
VP	<i>Verdadeiro Positivo</i>
W-MSA	<i>Window-based Multi-head Self-attention</i>
XGBoost	<i>Extreme Gradient Boosting</i>

Sumário

1	INTRODUÇÃO	17
1.1	Objetivos	18
1.1.1	Objetivos Específicos	19
1.1.2	Contribuições	19
1.2	Organização do Trabalho	19
2	TRABALHOS RELACIONADOS	20
2.1	Diagnóstico de Endometriose em Imagens de Ressonância Magnética	20
2.2	Diagnóstico de Endometriose em Laparoscopia	22
2.3	Diagnóstico de Endometriose em Ultrassonografia Transvaginal	24
2.4	Outros Métodos de Diagnóstico Utilizando Aprendizado de Máquina	25
2.5	Considerações Finais	26
3	FUNDAMENTAÇÃO TEÓRICA	28
3.1	Endometriose	28
3.1.1	Reto e Sigmoide	29
3.2	Ressonância Magnética	31
3.3	Processamento de Imagens Digitais	32
3.3.1	Convolução	32
3.3.2	<i>Template Matching</i>	33
3.3.3	Binarização de Imagem por Otsu	33
3.4	Aprendizado Profundo	34
3.4.1	Aprendizado de Máquina	34
3.4.2	Redes Neurais Artificiais	34
3.4.3	<i>Perceptron</i> de Múltiplas Camadas	36
3.4.4	<i>Backpropagation</i>	36
3.4.5	Características Profundas	36
3.4.6	Transferência de Aprendizado	37
3.4.7	Aumento de Dados	38
3.4.8	Redes Neurais Convolucionais	38
3.4.9	U-Net	39
3.4.10	Autoatenção de Múltiplas Cabeças	41
3.4.11	Transformer	42
3.4.12	Vision Transformer	44
3.4.13	Swin Transformer	45
3.4.14	Swin-Unet	46

3.5	Métricas de Avaliação	47
3.6	Considerações Finais	48
4	MÉTODO PROPOSTO	49
4.1	Extração da Região de Interesse	49
4.2	Segmentação das Lesões	52
4.3	Eliminação de Falsos Positivos	54
5	RESULTADOS	58
5.1	Base de Imagens	58
5.1.1	Protocolo de Marcação das Imagens	58
5.2	Descrição dos Experimentos	59
5.3	Resultados da Extração da Região de Interesse	60
5.4	Resultados da Segmentação das Lesões	60
5.4.1	Comparação com Outras Arquiteturas	63
5.5	Resultados da Eliminação de Falsos Positivos	64
6	CONCLUSÃO	70
6.1	Limitações e Trabalhos Futuros	71
6.2	Produção Científica	72
	REFERÊNCIAS	73

1 Introdução

O endométrio é o tecido que reveste o interior do útero. A endometriose é uma doença caracterizada pela presença do endométrio fora da cavidade uterina, causando lesões nos órgãos em volta. Essas lesões são responsáveis por diversos sintomas que prejudicam a qualidade de vida das mulheres. A doença é declarada profunda quando a lesão se estende por mais de 5 milímetros (AGARWAL; SUBRAMANIAN, 2010; SCHNEIDER et al., 2016; CARAÇA et al., 2011). É de alta incidência, estimando-se que 190 milhões de mulheres ao redor do mundo sejam afetadas pela endometriose, sendo 8 milhões no Brasil, com a doença causando grande parte das hospitalizações ginecológicas em países industrializados (BELLELIS; GIACOMETTI, 2023).

Diversas faixas etárias estão suscetíveis aos efeitos da endometriose. Entre os sintomas mais comuns estão a dor pélvica, alterações intestinais e urinárias, cólica menstrual acentuada, dor ao ter relações sexuais e infertilidade (BERBEL; PODGAEC; ABRÃO, 2008; NOGUEIRA et al., 2018). Também há estudos que apontam a doença como causadora de fadiga e até mesmo ansiedade e depressão (ANDREUCCI et al., 2014; TAYLOR, 2019). Os órgãos mais comumente afetados são os da região pélvica, como o útero e o intestino, mas em casos extremos afeta órgãos mais distantes, como os pulmões e o cérebro (SARMA et al., 2004; MARCHIORI et al., 2012; BERBEL; PODGAEC; ABRÃO, 2008). Estima-se que até 37% dos casos de endometriose profunda ocorrem no trato gastrointestinal. Desses casos, 90% das ocorrências envolvem o reto e sigmoide, órgão que encontra-se no final do intestino, tornando-o uma região de grande interesse de análise para diagnóstico (MOURA et al., 2019; AGARWAL; SUBRAMANIAN, 2010).

O tratamento da doença frequentemente envolve cirurgia, que, dependendo do local de ocorrência das lesões, pode ser difícil e extensiva. Portanto é necessário um método preciso para diagnóstico e análise de custo-benefício para a operação (SCHNEIDER et al., 2016). Os métodos mais comuns para o diagnóstico de endometriose são a laparoscopia, a ultrassonografia e a ressonância magnética (RM) (MANGANARO et al., 2012). A laparoscopia é considerada o padrão-ouro de diagnóstico, porém é invasivo, consistindo de uma pequena cirurgia, apesar de ser bem menos complicada e menos demorada que o tratamento (PONTES; CLAUDINO, 2021; TAYLOR, 2019; NOGUEIRA et al., 2018).

Há muita evidência que RM é o melhor método não invasivo para avaliar o grau de presença da endometriose profunda, se aferido por um especialista experiente, chegando a alcançar acurácia, sensibilidade e especificidade de até 90% (SCHNEIDER et al., 2016; GIUSTI et al., 2012). Essa técnica permite uma avaliação tridimensional, observando os planos coronal, sagital e axial com alta resolução espacial (ANDREUCCI et al., 2014).

Ainda assim, essa análise é notavelmente desafiadora, pelo fato das lesões serem visualmente semelhantes ao tecido à sua volta, além de seu formato ser variável. Assim, é um método de diagnóstico que acaba demandando muito tempo de análise do especialista (SCHNEIDER et al., 2016; LEIBETSEDER et al., 2022a).

Detecção e diagnóstico auxiliado por computador (CAD e CADx, do inglês *Computer Aided Detection* e *Computer Aided Diagnosis*, respectivamente) auxilia o especialista a interpretar as informações de imagens médicas, além de identificar a localização e formato de potenciais lesões para que o profissional possa validá-las como presença da doença (KIMORI, 2011; DINIZ et al., 2021). Tarefas como essa são utilizadas como base para estudos de visão computacional, subárea da inteligência artificial. Nessa perspectiva, o aprendizado profundo tem sido extensamente estudado para aplicação em tarefas de visão computacional, demonstrado, também, grande potencial quando aplicado a imagens de RM (LUNDERVOLD; LUNDERVOLD, 2019).

Hoje, redes neurais baseadas em *transformer* fazem parte do estado-da-arte da visão computacional (CAO et al., 2022). A Vision Transformer (ViT) é uma arquitetura de rede de aprendizado profundo baseada em *transformer* introduzida por Dosovitskiy et al. (2020). A partir dela, Liu et al. (2021) propuseram a Swin Transformer, que é uma arquitetura aprimorada de ViT, mais eficiente que suas antecessoras. Expandindo essa tecnologia, o trabalho de Cao et al. (2022) compara com outras arquiteturas de redes convolucionais e *transformers* o desempenho da Swin-Unet, uma ViT baseada em Swin Transformer, para a segmentação de imagens médicas de tomografia computadorizada. Os autores demonstraram sua maior eficácia em comparação às outras arquiteturas.

Trabalhos de segmentação da endometriose nesse tipo de imagem, baseados em aprendizado de máquina, são escassos. Essa escassez é ainda mais evidente ao buscar trabalhos que utilizem aprendizado profundo nesse contexto. Assim este trabalho tem como objetivo propor um método que utilize técnicas de processamento de imagens e aprendizado profundo para a tarefa de segmentação automática da endometriose profunda na região do reto e sigmoide a partir de imagens de RM.

1.1 Objetivos

O objetivo geral deste trabalho consiste em propor um método automático baseado em processamento de imagens e aprendizado profundo para a tarefa de segmentação da endometriose profunda na região do reto e sigmoide a partir de imagens de ressonância magnética.

1.1.1 Objetivos Específicos

Tendo em vista o objetivo geral deste trabalho, busca-se os seguintes objetivos específicos aplicados ao problema:

- Estudar as características da doença e das imagens, para utilizá-las a favor do modelo de aprendizado profundo;
- Aplicar e avaliar redes de aprendizado profundo para a tarefa de segmentação da endometriose profunda em imagens de ressonância magnética;
- Obter uma segmentação que seja útil para o auxílio de especialistas no diagnóstico da doença;
- Avaliar técnicas de processamento de imagens e aumento de imagens como contribuição à eficácia da rede profunda utilizada.

1.1.2 Contribuições

Destacam-se como principais contribuições:

- Segmentação das lesões de endometriose profunda em imagens de ressonância magnética com precisão e sensibilidade promissoras, além de segmentar com alta qualidade ao menos uma lesão de cada exame;
- Determinação da rede de aprendizado profundo Swin-Unet para a segmentação das lesões de endometriose, após essa arquitetura demonstrar-se a mais eficaz nos experimentos;
- Desenvolvimento de uma etapa de refinamento das segmentações por técnica de comparação da conectividade dos pixels das mesmas, a qual foi capaz de eliminar muitas segmentações incorretas sem grandes prejuízos às corretas, aumentando a precisão do método.

1.2 Organização do Trabalho

Este trabalho está estruturado da seguinte forma: no Capítulo 2 é apresentada a literatura referente ao diagnóstico da endometriose utilizando as técnicas mais comuns; no Capítulo 3 são apresentados os fundamentos teóricos importantes para a compreensão do presente trabalho; no Capítulo 4 é apresentado o método proposto para a segmentação automática da endometriose; no Capítulo 5 são apresentados e discutidos os resultados obtidos pelo método; e, por fim, no Capítulo 6, são descritas as conclusões do trabalho, com suas contribuições e limites, bem como as sugestões de trabalhos futuros.

2 Trabalhos Relacionados

Neste capítulo serão apresentados os trabalhos relacionados, sintetizando seus métodos e resultados. Esta bibliografia foi utilizada como referência para o desenvolvimento deste estudo. Os trabalhos abordam diversas formas de diagnóstico da endometriose. A principal abordagem envolve simplesmente a detecção da presença da endometriose. Nos métodos de diagnóstico com imagens, é possível realizar a detecção conjuntamente com a indicação da posição da lesão nas imagens. A abordagem menos frequente envolve a segmentação precisa da área da lesão nas imagens.

Os trabalhos estão, em sua maioria, divididos em três tipos de diagnóstico da endometriose: utilizando ressonância magnética (RM), utilizando laparoscopia e utilizando ultrassonografia transvaginal (USTV). Além desses métodos, são apresentadas outras formas de aplicação de aprendizado de máquina para o diagnóstico de endometriose, menos frequentes na literatura.

2.1 Diagnóstico de Endometriose em Imagens de Ressonância Magnética

O trabalho de [Figueredo et al. \(2023\)](#) apresenta uma abordagem promissora utilizando aprendizado profundo para classificação automática de imagens de RM de pacientes que apresentam endometriose no reto e sigmoide, bem como pacientes saudáveis. Foram aplicadas técnicas de processamento de imagens e uma modificação da rede neural convolucional VGG-16 ([SIMONYAN; ZISSERMAN, 2015](#)) para obter a classificação. É importante destacar que os autores utilizaram a mesma base de imagens que foi utilizada neste trabalho. O método proposto alcançou acurácia de 83,89%, sensibilidade de 84,15% e especificidade de 83,86%.

Ao final da execução deste trabalho, um método derivado do apresentado em [Figueredo et al. \(2023\)](#) foi proposto em [Figueredo et al. \(2024\)](#). Este método já realiza a segmentação automática da endometriose profunda nas imagens de RM, após classificá-las como tendo ou não tendo a doença. O método apresentado é complexo, composto de várias etapas e combinando várias técnicas. Abordando o trecho de segmentação do método proposto pelos autores, primeiro é realizada uma segmentação inicial das lesões através da rede profunda TransUNet ([CHEN et al., 2021](#)). Em seguida, as segmentações iniciais obtidas são utilizadas para, em cada paciente, recortar a região da imagem que englobe apenas estas segmentações. Essa etapa reduz a região de interesse das imagens, para que seja realizada a segmentação final. Esta última etapa também utiliza a TransUNet, porém

segmenta apenas a área recortada pela etapa anterior. Os resultados são então refinados, removendo segmentações muito pequenas. Os autores utilizaram a mesma base privada apresentada em [Figueredo et al. \(2023\)](#).

[Yang et al. \(2021\)](#) também abordam a classificação da endometriose profunda baseada em aprendizado profundo, em volumes 3D de RM. Os autores propuseram uma versão modificada da rede neural C3D ([TRAN et al., 2015](#)), chamada de IC3D, para classificar os volumes. O método proposto obteve uma acurácia de 99,2%. Os autores compararam os resultados da arquitetura proposta com o desempenho de outras arquiteturas comumente utilizadas. Foi observado que, mesmo nos casos em que a IC3D não obteve acurácia muito superior a outras redes, ela foi mais eficiente. A base de imagens utilizada é privada. A mesma contém 118 pacientes com endometriose e 206 pacientes com outras doenças ginecológicas.

Na revisão sistemática feita por [Moura et al. \(2019\)](#) são apresentados estudos sobre o diagnóstico manual de endometriose no reto e sigmoide em imagens de RM e USTV, com o objetivo de comparar as duas técnicas. As conclusões do trabalho indicam que ambas as técnicas são eficazes em equivalência, mesmo observando suas vantagens e desvantagens. Para RM, os trabalhos obtêm acurácia de 80% a 96%, sensibilidade de 73% a 100% e especificidade de 50% a 100%.

Já em [Bazot e Daraï \(2017\)](#), é proposta a revisão da laparoscopia como padrão-ouro, comparando-a com técnicas de imagem, entre elas a RM e a ultrassonografia. Os autores argumentam através de uma revisão da literatura, que obtêm uma sensibilidade média de 92% e especificidade média de 96% para o diagnóstico de endometriose no reto e sigmoide. É observado que a detecção da doença nessa região é realizada ao visualizar o desaparecimento de tecido gorduroso entre ela e o útero, com o surgimento de uma nova massa de tecido. Além disso, o desaparecimento do sinal hipointenso emitido pela parede anterior do reto e sigmoide também é evidência de endometriose, no caso de RM do tipo T2.

O estudo conduzido por [Scardapane et al. \(2013\)](#) utilizou uma base de imagens de RM privada para diagnosticar manualmente a endometriose profunda. Das 143 pacientes na base, 119 apresentavam endometriose. Na região do intestino, foi obtida uma acurácia de 96% no diagnóstico, apontando presença da doença em 33 pacientes. Dessas, em 28 a lesão encontrava-se na região do reto e sigmoide.

Em [Krüger et al. \(2013\)](#) foi feito um estudo da eficácia da utilização da RM para o diagnóstico manual da endometriose, em vários locais de comum ocorrência da doença. Dois radiologistas interpretaram vários órgãos da região pélvica de 152 RMs. Para o reto e sigmoide foram obtidas sensibilidade e especificidade de 80% e 78%, respectivamente. Essa região foi a terceira com menor sensibilidade dentre todas as analisadas. Os autores constataram que, para o diagnóstico manual da doença, a RM é confiável dependendo da

região de ocorrência. A base utilizada é privada.

Giusti et al. (2012) propuseram segmentar de forma semi-automática os órgãos apresentados nos volumes de RM, para então reconstruí-los em forma tridimensional para diagnosticar a endometriose profunda. Para isso foi utilizado o software ITK-SNAP (YUSHKEVICH et al., 2006). A base utilizada pelos autores também é privada. As reconstruções foram avaliadas por um radiologista experiente, levando em consideração o quão eficazes foram em apresentar a localização exata das lesões. A sensibilidade do diagnóstico da endometriose no reto e sigmoide utilizando o método foi de 79%, enquanto a por simples uso da RM foi de 58%.

Já o estudo de Saccardi et al. (2012) comparou o diagnóstico da endometriose através de RM com os realizados através de análise clínica, sonografia transvaginal e sonovaginografia com contraste. As RMs foram analisadas por um único especialista experiente. A sonovaginografia e a RM foram os métodos mais eficazes para detecção da presença e sua localização precisa na região pélvica. Na região retal, ambos os métodos alcançaram a mesma sensibilidade, mas a especificidade da RM foi maior. A base de imagens de RM não está disponível com anotações.

O trabalho de Bazot et al. (2004) analisou a acurácia do diagnóstico da endometriose profunda em imagens de RM realizado por dois radiologistas experientes. Uma das regiões de interesse do estudo foi a do reto e sigmoide. A base utilizada é privada e possui 195 pacientes, cuja presença de endometriose foi confirmada posteriormente através de cirurgia. No reto e sigmoide, a acurácia obtida foi de quase 95%, com uma sensibilidade de 88% e especificidade de quase 98%.

No estudo conduzido por Stratton et al. (2003), dois radiologistas experientes tiveram como objetivo detectar a presença e localização da endometriose em RMs de 48 pacientes que apresentavam dores pélvicas. A base de imagens utilizada é privada. A sensibilidade e especificidade obtidas foram de 69% e 75%, respectivamente. A principal dificuldade do uso das RMs foi para encontrar lesões pequenas.

2.2 Diagnóstico de Endometriose em Laparoscopia

Saadati e Amirmazlaghani (2024) propuseram um método de laparoscopia realizada automaticamente por um robô, a partir da segmentação da endometriose nas imagens. Foi utilizada a base de imagens ENID (LEIBETSEDER et al., 2022b), com 160 imagens extraídas de vídeos de laparoscopia. Os autores aplicaram U-Nets (RONNEBERGER; FISCHER; BROX, 2015) com *backbone* de ResNet50 (HE et al., 2016) para a detectar o útero e os ovários. Em seguida, a segmentação das lesões em cada órgão é realizada por uma junção de três U-Nets com *backbones* diferentes, sendo que a seleção de *backbones* muda entre útero e ovários. Os *backbones* utilizados foram MobileNet, InceptionV3, EfficientNetb7,

ResNet50, ResNet101 e VGG19 (HOWARD, 2017; SZEGEDY et al., 2016; TAN; LE, 2019; HE et al., 2016; SIMONYAN; ZISSERMAN, 2015). Foi obtido um Jaccard de aproximadamente 97% nos ovários e de até 99% no útero. Os autores observaram a importância de usar transferência de aprendizado e aumentar a base artificialmente.

Visalaxi e Sudalaimuthu (2022) propuseram um método de análise de similaridade estrutural da endometriose profunda a partir de duas abordagens, uma utilizando análise estatística e outra utilizando aprendizado profundo. A abordagem de aprendizado profundo aplicou uma U-Net, que alcançou Dice de 74% na segmentação. A base de imagens utilizada é a GLENDa (LEIBETSEDER et al., 2020), precursora da ENID. Essa base possui mais de 13 mil imagens extraídas de vídeos de laparoscopia, sendo 350 anotadas com lesões de endometriose. 55 das imagens anotadas apresentam a lesão no reto e sigmoide. A base foi aumentada artificialmente.

Leibetseder et al. (2022b) utilizaram aprendizado profundo para a detecção da endometriose no trabalho em que propuseram a base de laparoscopia ENID. Para isso, aplicaram as redes Faster R-CNN e Mask R-CNN (REN et al., 2016; HE et al., 2017). A média da precisão média (mAP) obtida foi de 32% para uma intersecção de mais de 50% da predição com a localização real da doença. O mAP obtido para uma intersecção de até 50% foi de 64%. Os autores desconsideraram a localização da doença nos órgãos, valorizando em vez disso a similaridade visual das estruturas presentes nas imagens. Foi evidenciada a importância do aumento de dados para a maior eficácia dos modelos. Além disso, foi utilizada transferência de aprendizado nos dois modelos.

Em Visalaxi e Muthu (2021), o estudo também foi conduzido a partir da base de imagens GLENDa. Os autores propuseram um método de classificação das imagens baseado em aprendizado profundo, para identificar quais apresentam endometriose profunda. A arquitetura utilizada foi a ResNet50. A rede foi pré-treinada na ImageNet (DENG et al., 2009), para que fosse feita a transferência de aprendizado ao refinar o treinamento da rede na base de imagens. Foi obtida acurácia de 90%, sensibilidade de 82% e precisão de 83%. Menos da metade da base foi utilizada.

No trabalho de Visalaxi, Punnoose e Muthu (2021) é proposta a identificação da endometriose através de técnicas de processamento de imagens. Utilizaram o filtro gaussiano para suavizar ruídos, então aplicaram o filtro laplaciano para detectar bordas e, em seguida, limiar adaptativo para binarizar a imagem (GONZALEZ; WOODS, 2010). Os resultados obtidos se limitam à identificação da intensidade das áreas de lesão dentro dos contornos obtidos sobre as imagens de base, tendo valor médio de 53%.

Já Fernando et al. (2013) analisaram a confiabilidade do diagnóstico visual, por laparoscopia, de endometriose em 431 pacientes. Cinco cirurgiões foram postos à tarefa de diagnosticar a doença. Esses resultados foram posteriormente comparados com a análise histológica da doença, onde amostras do tecido das lesões foram coletadas para

confirmar a endometriose. O diagnóstico foi bem-sucedido em 49,7% dos casos, com a endometriose profunda sendo mais facilmente identificável. Lesões pequenas resultaram mais frequentemente em erros.

Mettler et al. (2003) também realizaram um estudo objetivando analisar a eficácia do diagnóstico de endometriose por laparoscopia. 164 pacientes foram diagnosticados através desse exame. A presença da doença foi confirmada por análise histológica. Ao final, 84% das pacientes realmente apresentavam endometriose. Em algumas partes da região pélvica, menos de 50% dos diagnósticos por laparoscopia estavam corretos. Assim, os autores concluem que o exame visual realizado através de laparoscopia pode resultar em erro e, por conseguinte, tratamento desnecessário.

2.3 Diagnóstico de Endometriose em Ultrassonografia Transvaginal

Podda et al. (2024) abordaram a USTV como meio de segmentação da endometriose. A base de dados utilizada é privada e é composta de 75 imagens de pacientes com endometriose. A base foi aumentada artificialmente. Os autores propuseram um método que aplica uma junção de quatro U-Nets com diferentes resoluções de entrada, indo de 64x64 a 256x256. As segmentações geradas passam por uma etapa de união de seus pixels e uma etapa paralela de intersecção. A intersecção é utilizada para, através de operações de dilatação e erosão (GONZALEZ; WOODS, 2010), aproximar a área da intersecção à área da união. Assim, são eliminados falsos positivos da união, porém não perde-se muita área de lesão. Foi obtido um Dice de 82%, superior ao obtido pelas arquiteturas DeepLabV3+ e MSUnet (CHEN et al., 2018; SU et al., 2021).

Em Szabó et al. (2023), o diagnóstico da endometriose em USTV é realizado de forma manual. A base contém 537 pacientes, cujas USTVs foram analisadas por três ginecologistas experientes, porém não analisaram em conjunto nenhuma paciente. 52% da base apresentava a doença no reto e sigmoide. Foi alcançada uma acurácia de 94% na identificação da endometriose.

O trabalho de Liu et al. (2022) também utiliza a USTV para a segmentação de lesões de endometriose de forma automática. Os autores aplicaram uma ResNet50, convertida no formato codificador-decodificador, para a tarefa de segmentação. Foi obtido um Dice médio de 81%. Além da segmentação das lesões, o trabalho tinha como objetivo medir a espessura da endometriose, utilizando a área do maior círculo inscrito na segmentação. Foi obtida uma taxa de erro satisfatória, com maior dificuldade de acerto em lesões muito pequenas.

No estudo de Wang et al. (2022) as USTVs foram segmentadas a partir de seu formato tridimensional. Os autores aplicaram, separadamente, duas técnicas de remoção de ruído nos volumes, para então combinar as duas versões pré-processadas com os respectivos

volumes originais. Foi utilizada uma U-Net com convoluções tridimensionais para a segmentação. Os autores demonstraram que utilizar o volume inteiro das USTVs com a U-Net 3D alcançava melhores resultados do que usando as fatias do volume com a arquitetura original. Foi utilizada uma técnica baseada em esqueletização da segmentação (GONZALEZ; WOODS, 2010) para medir a espessura das lesões. Foi alcançado um Dice de quase 91% na segmentação e 94% das medidas de espessura dentro da taxa de erro aceitável.

Em Goncalves et al. (2021) a endometriose é detectada manualmente, comparando a eficácia da USTV com a laparoscopia. As laparoscopias e USTVs foram analisadas por dois cirurgiões e dois radiologistas, respectivamente. A base utilizada é composta por 120 pacientes. Foi observado que os dois métodos atingem resultados similares na maioria das áreas da região pélvica. Entretanto, a laparoscopia não foi capaz de detectar endometriose vaginal, diferentemente da USTV. No reto e sigmoide, a laparoscopia teve sensibilidade de no máximo 6% para a detecção, enquanto a USTV alcançou 96%.

Em Ferrero et al. (2019), também é realizado o diagnóstico manual da endometriose a partir de imagens de USTV. Essa tarefa é executada por dois ginecologistas. O objetivo do trabalho é determinar se o procedimento de preparo do intestino, antes do exame, auxilia esse método de diagnóstico. Das 155 pacientes incluídas no estudo, 92 apresentavam a doença na região do reto e sigmoide. Os experimentos determinaram uma melhora pouco significativa no diagnóstico utilizando o preparo do intestino. A acurácia obtida, nesse caso, foi de aproximadamente 90%.

2.4 Outros Métodos de Diagnóstico Utilizando Aprendizado de Máquina

No estudo conduzido por Tore et al. (2023), os casos clínicos de 627.566 pacientes de várias faixas etárias foram analisados, dos quais 82% haviam sido diagnosticadas com endometriose. Os autores propuseram colocar em foco as comorbidades apresentadas pelas pacientes como forma de detecção da doença através de aprendizado de máquina. O modelo obtido pelo XGBoost (CHEN; GUESTRIN, 2016) foi o mais eficaz, dentre os comparados, alcançando sensibilidade de 69% e especificidade de 63%. Outras características extraídas dos casos clínicos foram incluídas como entrada para o modelo de aprendizado. De todas as características utilizadas, a análise de importância realizada pelos autores aponta para a idade, infertilidade, ansiedade, rinite alérgica e fibroides uterinos como as mais significativas. Tendo em vista o desbalanceamento da base, os resultados obtidos são promissores.

Já Goldstein e Cohen (2023) propõem aplicar aprendizado de máquina para detectar a endometriose a partir de dados e sintomas informados pelas próprias pacientes, sem a necessidade de exames invasivos. Os dados foram obtidos através de uma pesquisa *online*, em espaços virtuais de mulheres que sofrem da doença. 56 sintomas comuns em

casos de endometriose foram considerados. O melhor modelo foi o AdaBoost (FREUND; SCHAPIRE; ABE, 1999), que obteve sensibilidade de 93% e especificidade de 95%. Foi analisado o grau de importância de cada sintoma para cada modelo experimentado, com o AdaBoost sofrendo maior influência de 24 sintomas. Os dados não foram analisados com base em características demográficas, como faixa etária.

Bendifallah et al. (2022) objetivaram diagnosticar a endometriose aplicando técnicas de aprendizado de máquina a uma base de dados clínicos de pacientes com endometriose. 1.834 pacientes foram utilizados no estudo, com 1.226 apresentando a doença. Especialistas selecionaram 16 tipos de dado clínico para serem utilizados como entrada para os modelos de aprendizado. Os modelos mais eficazes foram a Random Forest e XGBoost (BREIMAN, 2001; CHEN; GUESTRIN, 2016), obtendo um valor de *F1-Score* de 92%. Os resultados são promissores, porém os autores realçam o problema do conjunto de validação ser pequeno, com apenas 100 pacientes, o que torna limitada a análise da generalização do modelo.

2.5 Considerações Finais

Foram apresentados trabalhos que abordam diversos métodos para o diagnóstico da endometriose. Dentre os resultados observados, é possível ver que a tarefa de detecção e segmentação da endometriose são desafiadoras e que não existem muitas bases de imagens públicas. No caso do uso da RM, tem-se acesso apenas à base utilizada neste trabalho. Além disso, é visível que a literatura com RM é escassa em trabalhos de segmentação automática com aprendizado profundo, enquanto já existem alguns trabalhos que abordam a segmentação em laparoscopia e USTV. No geral, a utilização de aprendizado profundo na tarefa ainda é pouco explorada. Resultados dos trabalhos mais robustos, obtidos com aprendizado de máquina, usualmente são menores do que os obtidos manualmente. É muito procurado um método de diagnóstico confiável, porém não invasivo, como alternativa ao padrão-ouro.

Os trabalhos percorridos neste capítulo apresentam semelhanças a esta pesquisa e de algum modo contribuíram no desenvolvimento da mesma. Na Tabela 1 mostra-se o resumo dos trabalhos apresentados neste capítulo. É realçado o método de diagnóstico, o objetivo do trabalho e o acesso à sua base de imagens. Os objetivos foram divididos em dois tipos. Detecção (Det, na tabela) são os trabalhos que objetivam o diagnóstico da endometriose, além de poder incluir sua localização específica nos órgãos. Segmentação (Seg) é para trabalhos como este, que almejam demarcar precisamente os limites das lesões nas imagens. É possível observar o domínio de métodos manuais, dependentes de um especialista, para o diagnóstico da endometriose em RM. Assim, este trabalho contribui à literatura propondo um método automático para essa tarefa.

Tabela 1 – Visão geral dos trabalhos relacionados. “N/A” significa que a categoria não se aplica.

	Trabalho	Método de Diagnóstico	Objetivo	Base
RM	Figueredo et al. (2023)	Aprendizado profundo	Det	Mesma deste trabalho
	Figueredo et al. (2024)	Aprendizado profundo	Seg	Mesma deste trabalho
	Yang et al. (2021)	Aprendizado profundo	Det	Privada
	Moura et al. (2019)	N/A	Det	N/A
	Bazot e Darai (2017)	N/A	Det	N/A
	Scardapane et al. (2013)	Especialista	Det	Privada
	Krüger et al. (2013)	Especialista	Det	Privada
	Giusti et al. (2012)	Especialista	Det	Privada
	Saccardi et al. (2012)	Especialista	Det	Privada
	Bazot et al. (2004)	Especialista	Det	Privada
	Stratton et al. (2003)	Especialista	Det	Privada
Laparoscopia	Saadati e Amirmazlaghani (2024)	Aprendizado profundo	Seg	ENID
	Visalaxi e Sudalaimuthu (2022)	Aprendizado profundo	Seg	GLENDIA
	Leibetseder et al. (2022b)	Aprendizado profundo	Det	ENID
	Visalaxi e Muthu (2021)	Aprendizado profundo	Det	GLENDIA
	Visalaxi, Punnoose e Muthu (2021)	Processamento de Imagens	Seg	GLENDIA
	Fernando et al. (2013)	Especialista	Det	Privada
	Mettler et al. (2003)	Especialista	Det	Privada
USTV	Podda et al. (2024)	Aprendizado profundo	Seg	Privada
	Szabó et al. (2023)	Especialista	Det	Privada
	Liu et al. (2022)	Aprendizado profundo	Seg	Privada
	Wang et al. (2022)	Aprendizado profundo	Seg	Privada
	Goncalves et al. (2021)	Especialista	Det	Privada
	Ferrero et al. (2019)	Especialista	Det	Privada
Outros	Tore et al. (2023)	Aprendizado de máquina	Det	Privada
	Goldstein e Cohen (2023)	Aprendizado de máquina	Det	Privada
	Bendifallah et al. (2022)	Aprendizado de máquina	Det	Privada

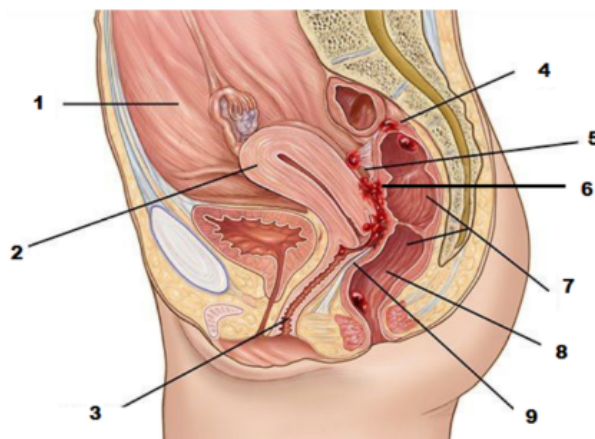
3 Fundamentação Teórica

Este capítulo apresenta os conceitos cuja compreensão é necessária para aprofundar-se neste estudo. Os componentes essenciais de cada conceito são explicados detalhadamente. Estes conceitos são a endometriose, a ressonância magnética, o processamento de imagens digitais, o aprendizado profundo e as métricas utilizadas para a avaliação do método proposto.

3.1 Endometriose

A endometriose é uma doença que caracteriza-se pela presença de células do endométrio fora da cavidade uterina. É uma condição inflamatória e afeta principalmente os órgãos localizados na região pélvica, como o útero, o reto e o sigmoide, conforme demonstrado pelas lesões vermelhas na Figura 1. Embora a manifestação da endometriose seja predominante nesses órgãos, casos mais raros podem apresentar a presença da doença em locais atípicos do corpo, como o cérebro e os pulmões. No início as lesões têm cor avermelhada, ficando mais escuras conforme a resposta inflamatória avança. Após a lesão ocorre a fibrose, que faz o tecido em que encontra-se a endometriose perder a função (SARMA et al., 2004; MARCHIORI et al., 2012; ANDREUCCI et al., 2014).

Figura 1 – Representação da endometriose na pelve. 1. Peritônio; 2. útero; 3. vagina; 4. cólon sigmoide; 5. lesão de endometriose; 6. fundo de posterior; 7. retossigmoide; 8. reto; 9. septoretovaginal.

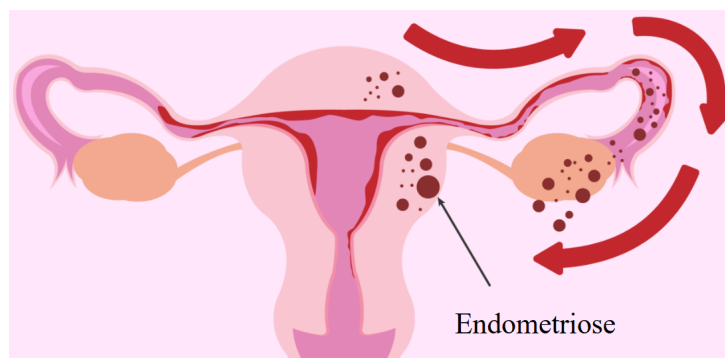


Fonte: [Levine \(2022\)](#).

Ainda há discussão entre as teorias propostas para a origem da endometriose. Um possível motivo pode ser por mudanças hormonais, que causam a implantação de células do endométrio em locais não convencionais do corpo. A teoria da menstruação retrógrada

é a mais aceita para a endometriose na região pélvica. Esta condição é caracterizada pelo deslocamento do fluxo menstrual no sentido oposto ao usual. Assim, células do endométrio são implantadas em órgãos ao redor da cavidade uterina, na região pélvica e abdominal (AGARWAL; SUBRAMANIAN, 2010; BERBEL; PODGAEC; ABRÃO, 2008; JUNIOR et al., 2008). Na Figura 2 é representado o fluxo da menstruação retrógrada.

Figura 2 – Ilustração do fluxo da menstruação retrógrada.



Fonte: Miranda et al. (2022).

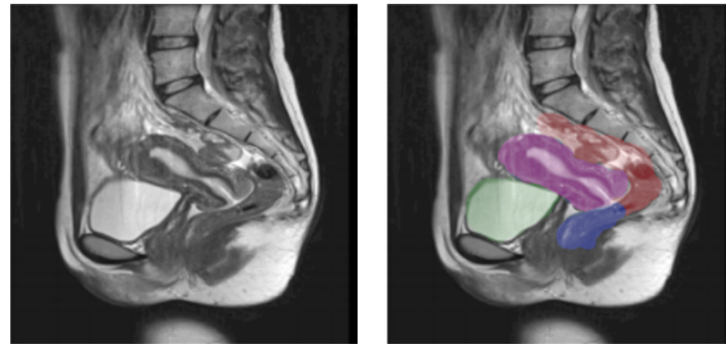
Entre os sintomas mais comuns estão a dor pélvica, alterações intestinais e urinárias, cólica menstrual acentuada, dor ao ter relações sexuais e infertilidade (BERBEL; PODGAEC; ABRÃO, 2008; NOGUEIRA et al., 2018). A doença também pode ser causadora de fadiga e até mesmo ansiedade e depressão (ANDREUCCI et al., 2014; TAYLOR, 2019). Lesões maiores que 5 milímetros são consideradas endometriose profunda. Essa é a forma mais grave da doença. Os métodos mais comuns para o diagnóstico de endometriose são a laparoscopia, a ressonância e a ultrassonografia (MANGANARO et al., 2012). A laparoscopia é considerada o padrão-ouro de diagnóstico, porém é invasivo, consistindo de uma pequena cirurgia, apesar de ser de baixa complexidade (PONTES; CLAUDINO, 2021; TAYLOR, 2019; NOGUEIRA et al., 2018). Métodos como ressonância magnética e ultrassonografia não são invasivos.

3.1.1 Reto e Sigmoide

O reto e o sigmoide são órgãos integrantes do sistema digestivo humano. O sigmoide, também conhecido como cólon sigmoide, apresenta uma configuração em forma de “S” e constitui uma porção do cólon que conecta o intestino grosso ao reto. Sua principal função é absorver água e nutrientes da comida não digerida, em uma escala menor que as áreas anteriores do intestino. Por sua vez, o reto é a última parte do intestino grosso, estabelecendo a ligação entre o cólon sigmoide e o canal anal. Sua principal função consiste em armazenar as fezes até que sejam eliminadas do corpo (HARKINS; SAJJAD, 2023). O reto e o sigmoide são as áreas mais afetadas pela endometriose no intestino grosso (AGARWAL; SUBRAMANIAN, 2010; ZWAS; LYON, 1991). A Figura 3 ilustra a

localização aproximada dessas regiões e de outras estruturas pélvicas femininas importantes, em uma ressonância magnética.

Figura 3 – Principais estruturas pélvicas. (a) Imagem de RM (b) Imagem com estruturas pélvicas destacadas: bexiga em verde, útero em lilás, sigmoide em vermelho e reto em azul.

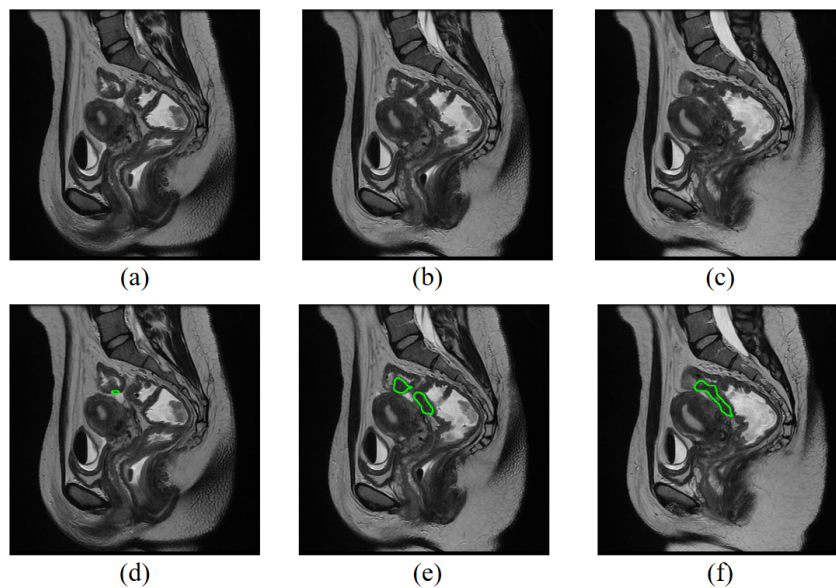


(a)

(b)

Fonte: Elaborado pelo autor.

Figura 4 – Exemplo de lesões de endometriose. (a), (b) e (c) são as imagens originais. (d), (e) e (f) são as respectivas imagens com as lesões marcadas.



(a)

(b)

(c)

(d)

(e)

(f)

Fonte: Elaborado pelo autor.

Na Figura 4 são apresentados exemplos de lesões de endometriose localizadas na região do reto e do sigmoide em um mesmo paciente. Nestas imagens, os especialistas delinearam as áreas afetadas em verde, enquanto a imagem original sem marcação evidencia as distintas formas das lesões e a dificuldade associada à sua visualização. Como ilustrado, as lesões nem sempre apresentam contornos nítidos, seja devido à sua textura ou à área circundante, e, por vezes, podem surgir em regiões onde a definição do reto e do sigmoide é

pouco perceptível. Além disso, é possível observar a falta de padrão de formato e tamanho das lesões, bem como podem haver lesões bem pequenas.

3.2 Ressonância Magnética

A ressonância magnética (RM) é um método de diagnóstico não invasivo, ou seja, que tem mínimo contato com o corpo do paciente. É um método altamente eficaz, e amplamente empregado para a visualização de várias áreas do corpo humano. Exemplos dessas áreas são o cérebro, a coluna vertebral, os músculos e os ossos. A resolução e o contraste da RM viabilizam a obtenção de imagens claras e detalhadas das estruturas internas do corpo. Sua aplicação é fundamental no auxílio ao diagnóstico de uma variedade de condições de saúde, tais como tumores, lesões, doenças cardíacas e distúrbios musculoesqueléticos (HAGE; IWASAKI, 2009).

Os primeiros estudos em RM remontam a 1946, quando dois grupos independentes, liderados por Purcell em Harvard e Bloch em Stanford, iniciaram investigações sobre a RM para diversas aplicações. Enquanto Purcell concentrava-se no estudo da RM em sólidos, a equipe liderada por Bloch explorava aplicações em líquidos para análises químicas dessas estruturas (BLOCH, 1946; PURCELL; TORREY; POUND, 1946). Em um marco significativo na história da RM, em 1972, o pesquisador Lauterbour, da Universidade de Illinois, obteve as primeiras imagens por meio dessa técnica. Em 1976, em Mansfield, na Universidade de Nottingham, realizou-se as primeiras imagens de uma parte do corpo humano utilizando a RM (HAGE; IWASAKI, 2009).

Figura 5 – Equipamento para a captura de RM.



Fonte: National Institute of Biomedical Imaging and Bioengineering. Disponível em: <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>. Acesso em: 12 de Dezembro de 2024.

A RM é uma técnica de imagem médica que utiliza campos magnéticos e ondas de rádio para gerar imagens do interior do corpo. O paciente é posicionado dentro de um campo magnético intenso. Então, a partir da interação entre este campo e os átomos de hidrogênio, o tecido corporal reage a esse campo magnético, gerando sinais de RM. Em seguida, ondas de rádio são empregadas para mensurar esses sinais e criar imagens detalhadas do interior do corpo. Essas imagens consistem em volumes compostos de várias fatias ([HAGE; IWASAKI, 2009](#); [JÚNIOR; YAMASHITA, 2001](#)). Na Figura 5 é apresentado o equipamento utilizado em procedimentos de RM.

3.3 Processamento de Imagens Digitais

O processamento de imagens digitais consiste na utilização de técnicas computacionais para interpretar uma imagem e extrair informações da mesma. As operações são divididas em três níveis: baixo, médio e alto. As de nível baixo são mais primitivas, como pré-processamento 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, 2010](#)).

A resolução de problemas por meio do processamento de imagens geralmente envolve os seguintes passos: aquisição, pré-processamento, segmentação, representação e reconhecimento. A aquisição refere-se à obtenção da imagem, o pré-processamento visa aprimorá-la, a segmentação divide a imagem em partes significativas, a representação seleciona atributos que caracterizam as informações relevantes e, por fim, o reconhecimento atribui rótulos e significados aos resultados. Uma base de conhecimento relacionada ao problema tratado orienta o funcionamento de cada etapa ([FILHO; NETO, 1999](#); [GONZALEZ; WOODS, 2010](#)). Embora todas essas etapas possam ser empregadas, um método específico pode incluir apenas um subconjunto delas, dependendo do problema a ser solucionado.

3.3.1 Convolução

A convolução é uma operação matemática muito empregada no processamento de imagens digitais. É caracterizada pela técnica que cria uma nova imagem a partir da imagem de entrada, utilizando um filtro, ou núcleo, deslizando na imagem. As dimensões do filtro devem ser menores que as da imagem original. Ao deslizar por cada pixel, cálculos matemáticos são executados sobre os pixels da região que o filtro engloba, criando uma nova imagem. Essa nova imagem pode ser uma versão aprimorada da imagem original, ou um resultado que destaca características relevantes da original. Além da dimensão do filtro,

é importante decidir como os pixels das bordas da imagem serão tratados. Comumente esses pixels são ignorados ou é aplicada alguma borda ao redor da imagem (GONZALEZ; WOODS, 2010).

A equação matemática que descreve a operação de convolução do filtro w sobre a imagem f , para obter a imagem g , é apresentada na Equação 3.1. O x e o y variam de forma que cada pixel em w percorre todos os pixels em f . a é dado por $(m - 1)/2$ e b é $(n - 1)/2$, onde m e n são as dimensões do filtro.

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x - s, y - t) \quad (3.1)$$

3.3.2 Template Matching

Essa técnica utiliza um *template* para buscar a região mais semelhante a ele em imagens maiores. Isto é feito deslizando o *template* sobre a imagem, como uma convolução 2D, até encontrar a região mais semelhante. A operação é repetida um certo número de vezes, testando cada vez com um tamanho diferente da imagem de *template*. A maior semelhança é indicada por um valor de similaridade atribuído a cada pixel da imagem sendo processada, que indica a probabilidade desse pixel ser o canto superior esquerdo da área que contém o *template* (GONZALEZ; WOODS, 2010).

3.3.3 Binarização de Imagem por Otsu

Algoritmos de limiarização, são utilizados para dividir as imagens em regiões. Esses algoritmos utilizam os níveis de intensidade de pixel, ou relações matemáticas entre eles, para definir os limiares. O método de limiarização de imagem de Otsu é um dos mais populares no processamento de imagens (GONZALEZ; WOODS, 2010). Ele foi proposto por Otsu et al. (1975). Ele é capaz de determinar o valor ideal que separe fundo e frente de uma imagem, atribuindo, usualmente, a cor branca para a frente e a cor preta para o fundo. Isso torna a imagem binária, com apenas dois valores de cor, portanto esse processo pode ser chamado de binarização de imagem.

A frente da imagem é normalmente o objeto de interesse que deseja-se extrair com a binarização. No caso deste trabalho, Otsu é utilizado para dividir áreas homogêneas das imagens, de forma que seja possível detectar diferentes regiões dos exames de RM. Otsu realiza essa tarefa iterando por todos os valores possíveis de limiar para a imagem, buscando aquele que minimiza a soma da variância dentro de cada conjunto de pixels, dos dois que serão gerados (brancos e pretos). Porém, esse cálculo de variâncias é computacionalmente custoso. Para corrigir esse problema, Otsu et al. (1975) buscou comprovar que é possível obter o mesmo resultado de limiarização minimizando a variância entre as duas classes.

3.4 Aprendizado Profundo

Aprendizado profundo é atualmente a área mais eficaz da inteligência artificial e vem sendo muito explorado para métodos de reconhecimento de texto, áudio, imagens, entre outros. É uma técnica robusta, capaz de processar dados que possuem muitas dimensões, fator limitante para as técnicas de aprendizado raso. 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; TAYE, 2023; PACHECO; PEREIRA, 2018). Nesta seção será explicado o que é aprendizado de máquina, além de quais são os componentes do aprendizado profundo. Ao final, serão apresentados os componentes que são utilizados neste trabalho.

3.4.1 Aprendizado de Máquina

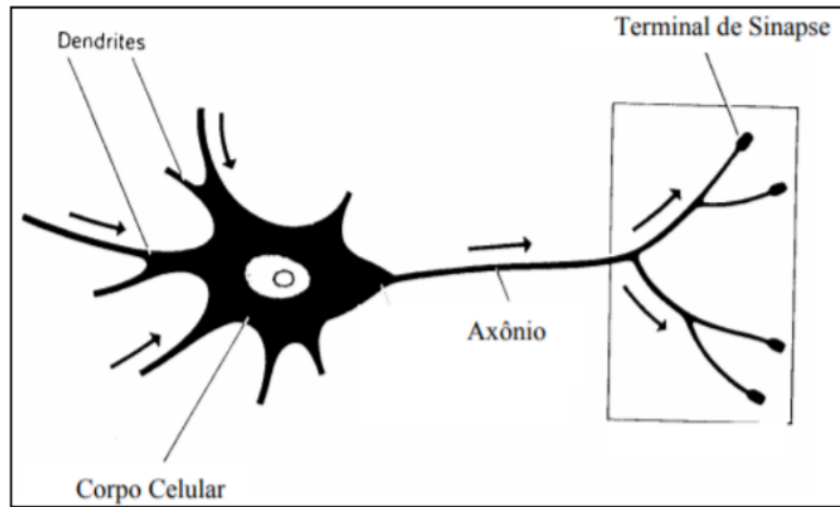
O aprendizado de máquina fundamenta-se na ideia de aprendizado por experiência. Nele, uma máquina é treinada a executar uma determinada tarefa por meio de um conjunto de exemplos. Nesse processo, a máquina aproxima-se da função ideal de solução do problema ao acumular aprendizado dos exemplos anteriores. 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 (TAYE, 2023; PACHECO; PEREIRA, 2018).

3.4.2 Redes Neurais Artificiais

As redes neurais artificiais (RNAs) são mecanismos de aprendizado de máquina baseados no funcionamento do cérebro humano. O cérebro processa informações a partir de uma complexa malha de neurônios que opera de forma não-linear e paralela. As RNAs foram desenvolvidas com base no modelo matemático do cérebro, emulando a atividade neuronal. Os componentes principais de um neurônio humano são: os dendritos, que coletam impulsos dos neurônios conectados a eles; o corpo celular, que processa os sinais recebidos; o axônio, responsável por transmitir os sinais até os terminais de sinapse; e as sinapses, que conectam-se a outro neurônio (PACHECO; PEREIRA, 2018; HAYKIN, 2001). Uma ilustração do neurônio humano pode ser vista na Figura 6.

O neurônio artificial representa uma unidade fundamental para o funcionamento de uma RNA (HAYKIN, 2001). Assim como nos neurônios biológicos, os neurônios artificiais necessitam de sinapses para conectar as entradas da rede ao neurônio, bem como conectar um neurônio no outro. Cada sinapse do neurônio artificial tem um peso, pelos quais as entradas são multiplicadas. Em seguida, essas entradas são somadas, juntamente com um valor constante chamado “*bias*”. O *bias* ajuda a controlar o valor de saída do neurônio. Por fim, uma função de ativação é aplicada para regularizar o valor de saída conforme

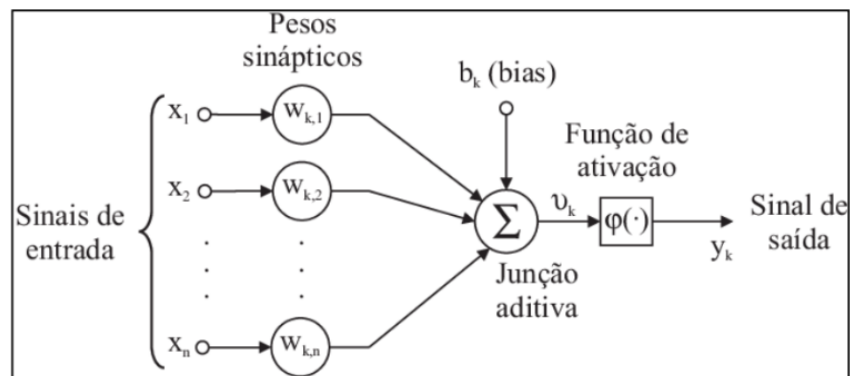
Figura 6 – Neurônio biológico.



Fonte: Bianchini et al. (2004).

algum critério (HAYKIN, 2001; PACHECO; PEREIRA, 2018; TAYE, 2023). O neurônio artificial pode ser observado na Figura 7.

Figura 7 – Neurônio artificial.



Fonte: Bianchini et al. (2004).

O aprendizado de uma RNA é realizado por meio de um processo conhecido como treinamento. Nele são fornecidos exemplos de entrada, que são processados pela rede. Esses exemplos são a base de dados para o aprendizado. Existem duas formas principais de realizar o aprendizado: a supervisionada e a não supervisionada. No aprendizado supervisionado, são utilizados exemplos de saída conhecidos para avaliar a saída do modelo e, dessa forma, ajustar seus parâmetros. Já no aprendizado não supervisionado, o algoritmo procura aprender seja agrupando as semelhanças entre os dados de entrada, seja tentando reconstruir a própria entrada. Após o treinamento, é realizado o teste utilizando dados que a RNA nunca viu, para avaliar seu desempenho (PACHECO; PEREIRA, 2018; TAYE, 2023).

3.4.3 *Perceptron* de Múltiplas Camadas

A Rede Neural *Perceptron* de Múltiplas Camadas (MLP, do inglês *Multilayer Perceptron*) é um tipo de RNA composta de uma sequência de múltiplas camadas de neurônios interconectados. Os neurônios da primeira camada recebem os dados de entrada da rede, enquanto os posteriores recebem a saída da camada anterior à sua. Cada neurônio processa as informações e transmite-as para a camada seguinte. Na última camada é possível obter o mais alto grau de representação das informações. As camadas entre a camada de entrada e a de saída da MLP são chamadas de camadas ocultas (PACHECO; PEREIRA, 2018; TAYE, 2023; SILVA, 2004; DINIZ et al., 2021). O cálculo realizado por uma camada MLP é demonstrado na Equação 3.2, onde $y^{(l)}$ é o vetor de saída da i -ésima camada oculta, σ é a função de ativação, $W^{(l)}$ é a matriz de pesos entre as camadas $l - 1$ e l , $y^{(l-1)}$ é a saída da camada anterior e $b^{(l)}$ é o *bias* da camada.

$$y^{(l)} = \sigma(W^{(l)}y^{(l-1)} + b^{(l)}) \quad (3.2)$$

3.4.4 *Backpropagation*

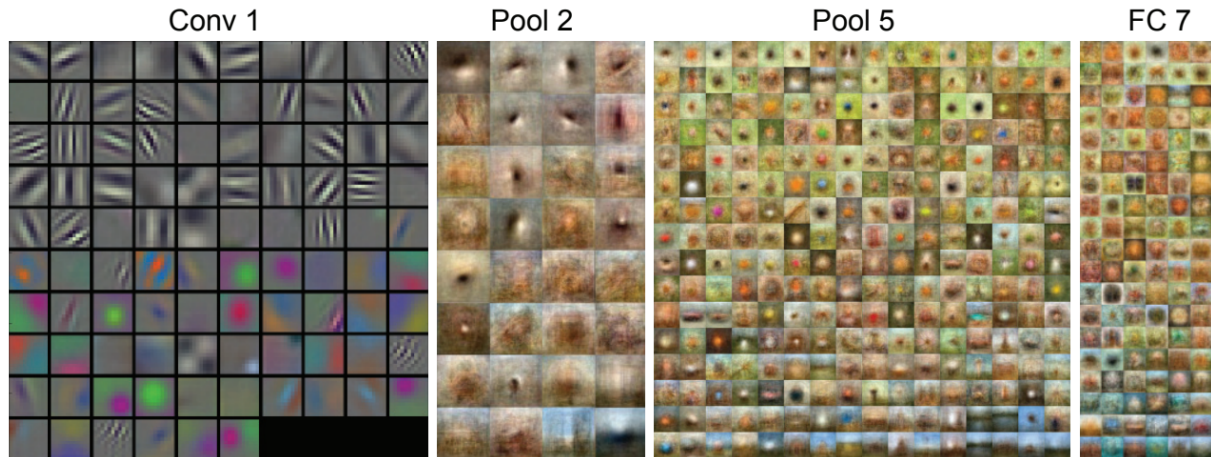
Backpropagation, ou propagação para trás, é um algoritmo usado para calcular os gradientes dos pesos das conexões da rede neural em relação a uma função de custo. Essa função de custo calcula o erro da saída da rede em relação a um alvo. Essa operação permite ajustar os pesos da rede de forma a tentar alcançar o erro mínimo de predição após o treinamento. Antes do *backpropagation* é feita a propagação para frente (*forward propagation*). Durante a propagação para frente, os dados de entrada são introduzidos na rede neural, atravessam cada camada e geram as saídas correspondentes. A função de custo calcula o erro e a propagação para trás é realizada. Ela propaga o erro desde a camada de saída até a camada de entrada, recalculando os pesos de toda a RNA. Esse processo é repetido para cada exemplo de treinamento. Cada passada por todos os exemplos de treinamento é chamada de época e a rede pode ser treinada por várias épocas, até chegar num valor de custo desejado (SILVA, 2004; TAYE, 2023).

3.4.5 Características Profundas

Técnicas de aprendizado de máquina anteriores ao aprendizado profundo são pouco eficazes em processar dados puros, ou seja, sem nenhum tipo de engenharia metódica de características, anterior ao treinamento do modelo (LECUN; BENGIO; HINTON, 2015). No caso deste trabalho, esses dados são as imagens, representadas em um vetor de valores de pixel. Um método de aprendizado profundo aprende a representação dos dados, em múltiplas camadas, cada uma tendo como entrada as características extraídas pela camada anterior. A primeira camada tem como entrada os dados puros. Características configuram-

se, então, como profundas, quando são extraídas por tais métodos. É importante perceber que essas camadas não são desenvolvidas por um engenheiro, mas sim a partir dos dados, através de um processo de aprendizado (LECUN; BENGIO; HINTON, 2015).

Figura 8 – Ilustração das definições aprendidas pelas camadas de uma arquitetura de aprendizado profundo.



Fonte: Zhou et al. (2014).

Uma arquitetura de aprendizado profundo é a composição dessas camadas, que são caracterizadas por módulos não-lineares que transformam a entrada em uma representação mais abstrata. Se essa arquitetura for composta por um número suficiente desses módulos, é capaz de aprender funções extremamente complexas. A primeira camada recebe, por exemplo, a imagem em valores de pixel e usualmente aprende a extrair características que representam a presença ou ausência de bordas em orientações específicas. A camada seguinte então extrai representações a partir de grupos de bordas. A terceira camada vai extrair características referentes a partes de objetos, para que a posterior extraia já os objetos, e assim em diante (LECUN; BENGIO; HINTON, 2015). Uma ilustração dessas camadas pode ser observada na Figura 8. Neste trabalho, são utilizadas características profundas para a detecção da endometriose.

3.4.6 Transferência de Aprendizado

Um grande desafio no treinamento de uma rede profunda é a escassez de dados. Quanto menos diversidade tiver a base de dados utilizada no treinamento, maior o risco de sobreajuste da rede e a ineficácia geral do treinamento. Esse desafio é encontrado principalmente no contexto de imagens médicas, como o deste trabalho. Isto se dá pelo fato de serem escassos os conjuntos de imagens médicas em larga escala e anotados. A transferência de aprendizado é uma forma de amenizar essa dificuldade (SHIN et al., 2016). Ela é a técnica caracterizada pelo treinamento de uma rede neural profunda em um grande

conjunto de dados e, em seguida, é feito um ajuste fino do modelo nos dados que realmente deseja-se representar (YOSINSKI et al., 2014).

Na Seção 3.4.5 foi descrito e ilustrado que as primeiras camadas profundas aprendem padrões de bordas. Esses elementos são comuns a qualquer contexto de processamento de imagens. As camadas finais são responsáveis por aprender características mais abstratas e, portanto, mais específicas ao conjunto de dados alvo. A partir disso, o objetivo da transferência de aprendizado é conseguir maior generalização nas camadas menos específicas, através do treinamento em uma enorme base de imagens gerais que possua milhares de imagens e muitas classes, como a ImageNet. Assim, essa técnica melhora a eficácia dos modelos de aprendizado profundo (DENG et al., 2009; YOSINSKI et al., 2014; CAO et al., 2022). Neste trabalho, utiliza-se a transferência de aprendizado como inicialização da rede profunda explicada na Seção 3.4.14.

3.4.7 Aumento de Dados

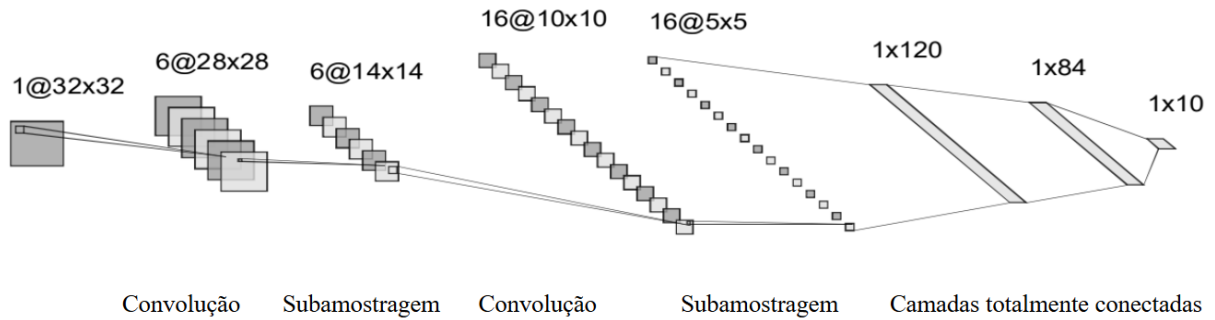
O aumento de dados é outra forma de abordar o problema da escassez de dados ao treinar redes profundas. Nessa técnica, são realizadas transformações artificiais nos componentes da base de dados para gerar novas instâncias. Assim, busca-se obter mais variabilidade nos dados. No contexto de imagens, são utilizadas transformações como giros, rotações, mudança de brilho, zoom e translações. Esse aumento pode ser realizado antes do treinamento, sendo então chamado de aumento *offline*, ou durante o treinamento, sendo chamado de aumento *online*. O treinamento *online* consiste em haver uma chance de realizar uma ou mais transformações na imagem antes desta ser inserida na rede, a cada época. Foi provado que o aumento de dados tem grande contribuição na melhora da eficácia de modelos profundos, inclusive na área de imagens médicas (KUMAR et al., 2024; CAO et al., 2022). Sendo assim, este trabalho aplica o aumento *online* ao treinar a rede utilizada no método.

3.4.8 Redes Neurais Convolucionais

Redes neurais convolucionais (CNNs, do inglês *convolutional neural networks*) são RNAs que utilizam um mecanismo de convolução, como explicado na Seção 3.3.1, para extrair características dos dados de entrada. As CNNs são frequentemente utilizadas em métodos de aprendizado a partir de imagens, para tarefas como a classificação, segmentação e detecção de objetos. A arquitetura padrão de uma CNN é composta de uma camada de entrada, seguida de um ou mais conjuntos de camadas convolucionais e de subamostragem, terminando nas camadas totalmente conectadas (LI et al., 2021; ZHAO et al., 2024). Na Figura 9, é possível ver um exemplo de CNN com os componentes descritos.

A camada de convolução extrai características locais de todas as regiões da imagem,

Figura 9 – Arquitetura da CNN LeNet-5.



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

utilizando filtros quadrados, ou núcleos, que implementam a convolução. Cada filtro configura um neurônio e gera um mapa de características, que irá compor a entrada da camada seguinte. As saídas dessas camadas são funções lineares. Para que seja possível obter vantagem da utilização de camadas profundas e aproximar funções não lineares, é utilizada uma função de ativação após final de cada camada de convolução. Essa função seleciona as informações que devem ser transmitidas de cada neurônio ao neurônio seguinte ([LI et al., 2021](#); [ZHAO et al., 2024](#)).

O tamanho dos mapas de características pode ser controlado pelas configurações de deslocamento dos filtros e do tratamento das bordas da imagem. Porém, comumente utiliza-se uma camada de subamostragem logo após cada convolucional para reduzir a resolução dos mapas. Essa camada divide a entrada em regiões e de cada região vai ser extraído um único valor que representa a região, por exemplo, pegando a média ou o valor máximo da região. Isso faz com que a rede seja menos afetada por pequenas diferenças morfológicas locais, além de diminuir a complexidade computacional tornando necessários menos neurônios nas camadas seguintes ([LI et al., 2021](#); [ZHAO et al., 2024](#)).

Os mapas de características são convertidos em vetores unidimensionais para serem passados às camadas completamente conectadas. Essa parte da CNN é idêntica ao *perceptron* apresentado na Seção 3.4.3. Essas camadas fazem o processamento global das características já abstraídas e compactadas pelas camadas explicadas anteriormente ([ZHAO et al., 2024](#)). O entendimento das CNNs é útil para entender o funcionamento da arquitetura U-Net, descrita na Seção 3.4.9. Além disso, essas informações são necessárias para se observar as vantagens do uso das *transformers*, descritas na Seção 3.4.11.

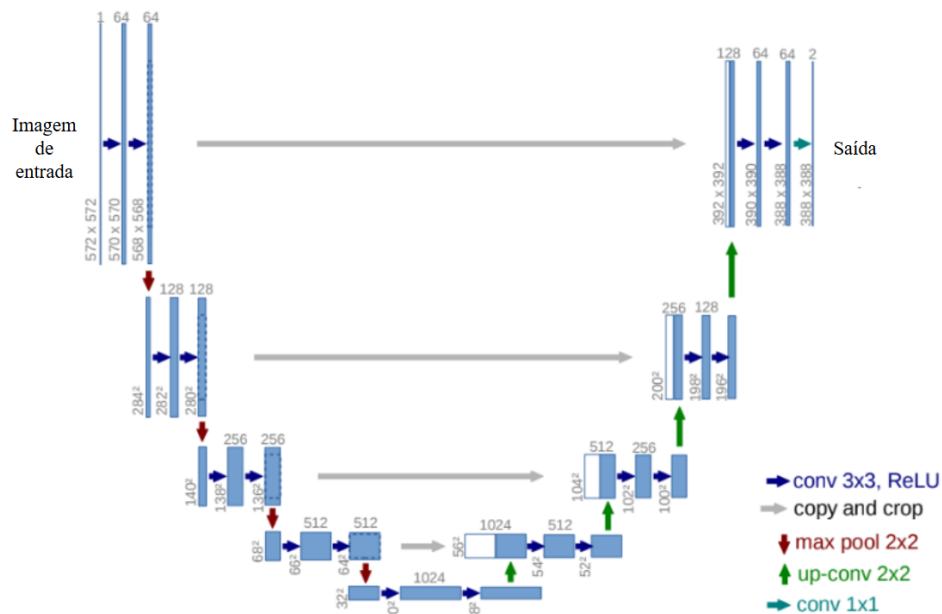
3.4.9 U-Net

A U-Net é uma arquitetura de CNN proposta por [Ronneberger, Fischer e Brox \(2015\)](#) para a segmentação de imagens. As CNNs como apresentadas na Seção 3.4.8 podem ser entendidas como codificadores. Isso significa que extraem características das imagens

de forma cada vez mais compacta e abstrata, criando uma eficaz representação semântica dos dados. Isso é útil principalmente em tarefas de classificação, onde a imagem como um todo deve ser atribuída a uma ou mais classes. Porém, a tarefa de segmentação exige que a classificação seja a nível de pixel, gerando uma máscara que represente com precisão a localização dos objetos de interesse na imagem. Sendo assim, a U-Net é uma arquitetura composta por um codificador e um decodificador, conectados por um gargalo.

O codificador é composto de camadas convolucionais com subamostragens a cada duas convolucionais, aumentando a quantidade de mapas de características e diminuindo a resolução dos mesmos. O gargalo extrai, então, a representação mais compacta das informações. Em seguida, o decodificador faz o caminho inverso. Utiliza-se camadas de convolução transposta para aumentar a resolução dos mapas de características, enquanto a quantidade de mapas diminui. Cada convolução transposta é seguida de duas camadas convolucionais convencionais (RONNEBERGER; FISCHER; BROX, 2015). A arquitetura da U-Net pode ser observada na Figura 10, tendo o mesmo número de etapas de codificação e decodificação. Como é possível observar, antes de cada subamostragem do codificador o mapa de características é copiado e cortado para ser concatenado à etapa correspondente do decodificador.

Figura 10 – Arquitetura da U-Net.



Fonte: Ronneberger, Fischer e Brox (2015).

Ao final, uma convolução com a quantidade de filtros referente à quantidade de classes desejadas na segmentação é aplicada para obter as máscaras como saída. A U-Net vem sendo utilizada com grande sucesso para a segmentação de imagens, inclusive imagens médicas. É possível substituir o codificador da U-Net pelo de outras arquiteturas mais modernas de CNN. O decodificador pode ser então substituído por uma versão de expansão

de características da arquitetura escolhida (CAO et al., 2022). Neste trabalho foi utilizada uma rede baseada em U-Net, porém com blocos de Transformer. Essa abordagem será discutida na Seção 3.4.14.

3.4.10 Autoatenção de Múltiplas Cabeças

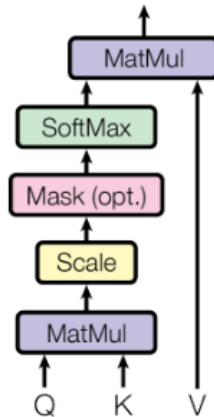
Uma função de atenção pode ser representada como o mapeamento de um vetor consulta e conjuntos de vetores de chave e valor a uma saída. A saída é calculada por uma soma ponderada dos vetores de valor, sendo que os pesos são determinados através de uma função de compatibilidade do vetor de consulta com o de chave correspondente a cada valor. A projeção que resulta nos vetores de consulta, chave e valor são aprendidos pela rede. A autoatenção é caracterizada pela aplicação desse mecanismo para relacionar cada elemento de uma sequência para obter uma representação global da sequência. As primeiras e principais utilizações bem-sucedidas da autoatenção foram em métodos de processamento de linguagem natural. Nessa tarefa as sequências são compostas por palavras, ou por sílabas, ou frases, dentre outros tipos de componentes (VASWANI et al., 2017; DOSOVITSKIY et al., 2020).

Vaswani et al. (2017) propuseram a atenção por produto escalar escalonado. Nela é calculado o produto escalar da consulta com todas as chaves e cada resultado é dividido por $\sqrt{d_k}$, sendo d_k a dimensão das chaves. A atenção é calculada simultaneamente nas consultas agrupando-as em uma matriz Q , bem como são definidas matrizes K e V para chaves e valores, respectivamente. Assim, a saída é descrita na Equação 3.3. K^T é a transposta da matriz K . A função *Softmax* comprime um vetor de tamanho L em um vetor de tamanho L com valores entre 0 e 1 e os seus elementos tendo soma 1. Na Figura 11 é possível ver a estrutura da camada de autoatenção. A operação “*mask*” é opcional, quando deseja-se evitar conexões indesejadas. A operação “*scale*” é onde o produto escalar é escalado por $\frac{1}{\sqrt{d_k}}$ para evitar que dimensões muito grandes de chave prejudiquem o aprendizado (ZHAO et al., 2024; VASWANI et al., 2017).

$$Atenção(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.3)$$

Em seguida, Vaswani et al. (2017) propõem a camada de autoatenção de múltiplas cabeças. Na prática, é um número h de camadas de autoatenção por produto escalar escalonado, dispostas em paralelo. Cada uma recebe como entrada projeções de consulta, chave e valor diferentes das demais. Dessa forma, o modelo pode processar informações de diferentes espaços representativos. É possível ver na Figura 12 a configuração dessa camada. A autoatenção de múltiplas cabeças permite à rede aprender relações de componentes que estejam distantes entre si nas sequências. Uma camada convolucional, como foi exposto na Seção 3.4.8, tem o fator limitante de apenas capturar correlações locais. A camada de

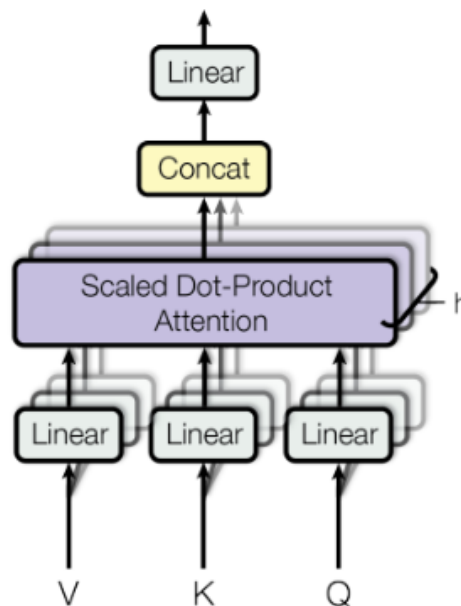
Figura 11 – Atenção por produto escalar escalonado.



Fonte: Vaswani et al. (2017).

autoatenção é o principal componente utilizado nas redes profundas *transformers*, que serão apresentadas na Seção 3.4.11.

Figura 12 – Autoatenção de múltiplas cabeças



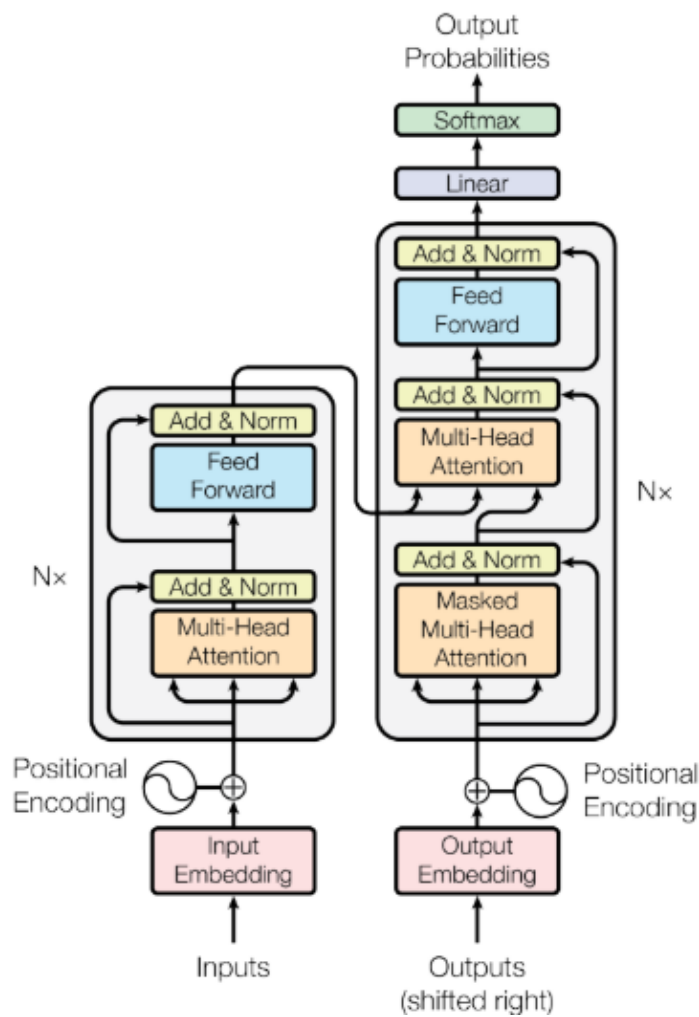
Fonte: Vaswani et al. (2017).

3.4.11 Transformer

A Transformer é uma rede neural profunda proposta por Vaswani et al. (2017), composta principalmente por camadas de autoatenção e camadas completamente conectadas. O codificador proposto pelos autores possui N camadas idênticas. Cada camada possui duas subcamadas, sendo uma a autoatenção de múltiplas cabeças e a outra sendo uma camada

completamente conectada, como o *perceptron* descrito na Seção 3.4.3. O decodificador também é composto por N camadas semelhantes às do codificador. Porém, possui mais uma camada de autoatenção, que processa a saída do codificador. Na Figura 13 é apresentada a arquitetura da Transformer.

Figura 13 – Arquitetura da Transformer.



Fonte: Vaswani et al. (2017).

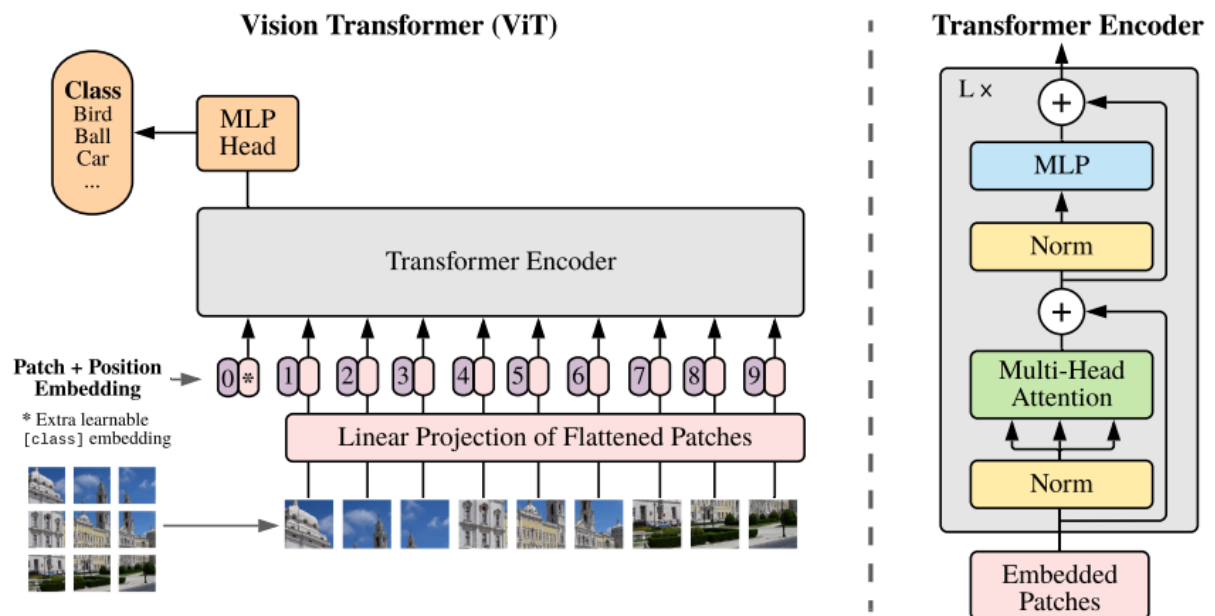
A camada de *embedding* na entrada de cada rede é responsável por transformar os dados de entrada em vetores de tamanho padronizado. Normalmente isso é realizado por uma projeção linear aprendida durante o treinamento. Já a codificação posicional é um componente necessário para o modelo aprender informações relevantes de acordo com a posição dos elementos na sequência de entrada. Essas codificações são vetores cujos valores podem ser definidos através de uma função fixa, ou serem aprendidos pelo modelo. Ambas as abordagens eficazes. Porém, codificações posicionais aprendidas pelo modelo podem limitar sua eficácia ao tamanho das sequências utilizadas no treinamento. Os vetores são somados aos *embeddings* de entrada (VASWANI et al., 2017).

A Transformer é uma arquitetura muito estudada. Ela e suas evoluções têm obtido excelente desempenho, principalmente em tarefas de processamento de dados sequenciais. Esse sucesso estimulou pesquisas com essa arquitetura para tarefas de processamento de imagens (ISLAM et al., 2024; DOSOVITSKIY et al., 2020). Na seção 3.4.12 é discorrido como a Transformer foi adaptada para extrair características de imagens. Neste trabalho foi utilizada uma rede Transformer para a tarefa de segmentação da endometriose.

3.4.12 Vision Transformer

A Vision Transformer (ViT) é a arquitetura proposta por Dosovitskiy et al. (2020) que aplica a Transformer no contexto de imagens. 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 14.

Figura 14 – Arquitetura da ViT e sua camada de codificação.



Fonte: Dosovitskiy et al. (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 et al., 2020; HENDRYCKS; GIMPEL, 2016).

Os autores observaram que a ViT tinha melhor desempenho em bases pequenas ao usar transferência de aprendizado, pretreinando-a em uma base maior. Eles concluíram que a ViT se equiparou ou superou muitas arquiteturas do estado-da-arte. Além disso, realizaram um estudo que indica que a ViT dá mais atenção aos objetos de interesse da imagem, principalmente quanto mais profunda for a camada. As cabeças de atenção passam a convergir, apresentando campos receptivos cada vez mais semelhantes. No campo das imagens médicas, regiões-alvo espalhadas e diferenças de formato dos objetos de interesse são um desafio para as redes profundas. A ViT foi recebida com muito entusiasmo nessa área, pois aliviava esse problema. Devido à sua grande eficácia, muitos estudos ainda são realizados para aperfeiçoar a ViT em tarefas de classificação, detecção e segmentação com imagens médicas (PU et al., 2024).

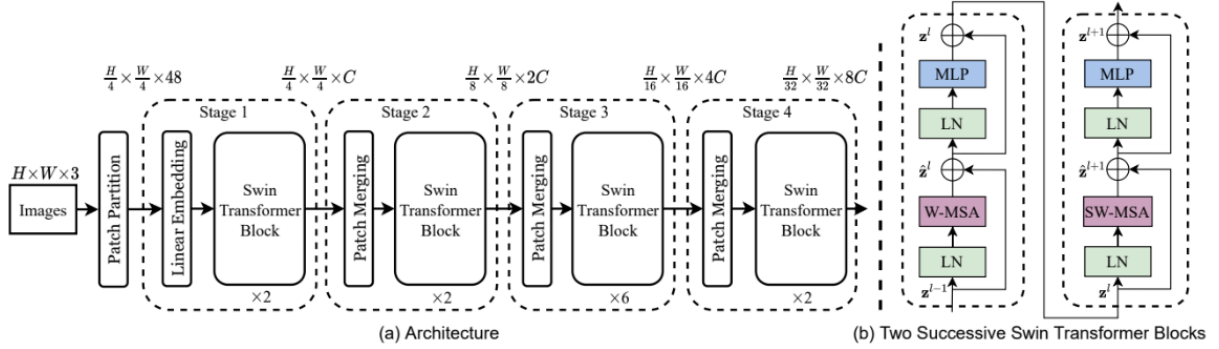
3.4.13 Swin Transformer

Para abordar o desafio da alta carga computacional necessária para executar as *transformers*, Liu et al. (2021) propuseram a *Swin Transformer*. Sua arquitetura modifica o funcionamento da autoatenção nas camadas da Transformer. A utilização da autoatenção é custosa quanto maior a resolução da imagem de entrada. Os autores, então, propuseram as camadas de autoatenção de múltiplas cabeças com janelas (W-MSA, do inglês *window-based multi-head self-attention*) e com janelas deslizantes (SW-MSA, do inglês *shifted window-based multi-head self-attention*). Essas camadas dividem os recortes da imagem em janelas. Toda janela terá um mesmo número M de recortes. Na W-MSA auto-atenção é realizada apenas entre os recortes de cada janela. Para não perder muito da capacidade de processamento global de informações, o resultado da W-MSA passa pela SW-MSA, que desloca as janelas na direção da diagonal inferior direita. Os recortes que ficam fora das janelas, após o deslizamento, são reorganizados nos novos espaços vazios definidos. Esse processo permite que recortes vizinhos, antes separados pelas janelas, participem da autoatenção juntos.

A técnica original de autoatenção, como proposta por Dosovitskiy et al. (2020), tem complexidade quadrática, uma vez que todo recorte participa da autoatenção de todo recorte. A Swin Transformer tem complexidade linear, em suas camadas. Uma vez fixado o valor de M , aumentar a resolução da imagem não afeta significativamente o custo computacional. A arquitetura da Swin é semelhante à da ViT, como é apresentado na Figura 15. Cada camada de Swin Transformer é composta por duas subcamadas, semelhantes à camada da Transformer. A primeira utiliza W-MSA e a segunda utiliza

SW-MSA. Além disso, a resolução dos mapas de características é reduzida quanto mais fundo na rede. Isto é feito empilhando os recortes dois a dois, concatenando-os, reduzindo a resolução do mapa de característica e aumentando seu volume (LIU et al., 2021).

Figura 15 – Arquitetura da Swin Transformer e sua camada de codificação.



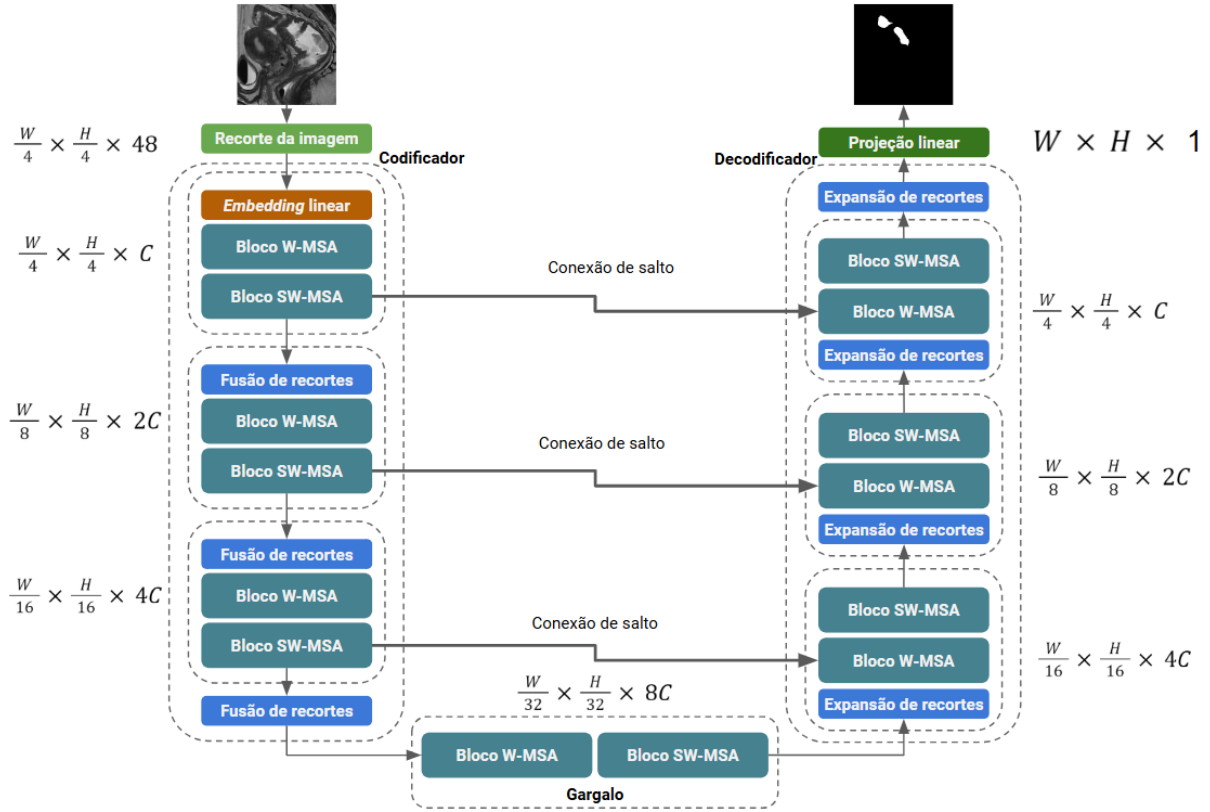
Fonte: Liu et al. (2021).

3.4.14 Swin-Unet

A Swin-Unet foi proposta por Cao et al. (2022) para converter a Swin Transformer em uma rede de segmentação. Ela assume o formato característico da U-Net, porém utiliza camadas de Swin Transformer ao invés das convolucionais. No codificador é utilizada a técnica de concatenação dos recortes das janelas, explicada na Seção 3.4.13, para diminuir a resolução dos mapas de características. No codificador, é aplicada a operação inversa, desconcatenando os recortes e retornando-os à sua disposição original, para aumentar a resolução em cada camada. No final do decodificador há uma camada convolucional para transformar os mapas de características nas máscaras correspondentes às classes segmentadas. A arquitetura da Swin-Unet pode ser observada na Figura 16.

Abordagens anteriores de segmentação com Swin Transformer comumente utilizam as camadas de autoatenção apenas no codificador, sendo o decodificador convolucional. A Swin-Unet é uma das primeiras a utilizar apenas camadas Swin. Os autores alcançaram um desempenho excelente com seu modelo, ultrapassando várias arquiteturas do estado-da-arte. As bases de dados nas quais a Swin-Unet foi testada são compostas de imagens médicas. Uma base de tomografia e a outra de RM. Percebe-se, então, como promissora a utilização dessa rede para a segmentação da endometriose em imagens de RM. Cao et al. (2022) também verificaram que a maior eficácia da rede foi alcançada utilizando transferência de aprendizado, inicializando os pesos da rede com um pretreinamento na base ImageNet (DENG et al., 2009).

Figura 16 – Arquitetura da Swin-Unet.



Fonte: Elaborada pelo autor.

3.5 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 da endometriose.

A comparação das previsões 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 endometriose. FP são os pixels que não contêm doença, mas que foram segmentados como endometriose. VN pixels que não apresentam a doença e que não foram segmentados como se a apresentassem. Finalmente, FN são os pixels que contêm endometriose mas não foram segmentados pelo método (MÜLLER; SOTO-REY; KRAMER, 2022).

A precisão (Equação 3.4) verifica, dentre as segmentações de positivos realizadas

pelo modelo, quantas estão corretas. Ela é penalizada quando a rede segmenta muitos falsos positivos. A sensibilidade (Equação 3.5) é a proporção de verdadeiros positivos, ou seja, o quanto a rede é capaz de detectar a doença, independentemente dos falsos positivos. Ela é extremamente importante, pois diminui caso o modelo não esteja conseguindo detectar a doença, caso mais perigoso no contexto médico. O *F1-Score*, ou coeficiente de similaridade Dice (Equação 3.6), é a média harmônica entre precisão e sensibilidade. Essa métrica possibilita balancear a segmentação eficaz dos positivos, sem que sejam gerados muitos falsos positivos. Dessa forma, é uma das métricas mais úteis para avaliar os resultados. Por fim, a intersecção sobre a união, ou coeficiente de similaridade Jaccard (Equação 3.7), indica a similaridade em posição e formato das segmentações (MÜLLER; SOTO-REY; KRAMER, 2022; ZULUAGA-GOMEZ et al., 2020).

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (3.4)$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (3.5)$$

$$\text{Dice} = \frac{2 * VP}{2 * VP + FP + FN} \quad (3.6)$$

$$\text{Jaccard} = \frac{VP}{VP + FP + FN} \quad (3.7)$$

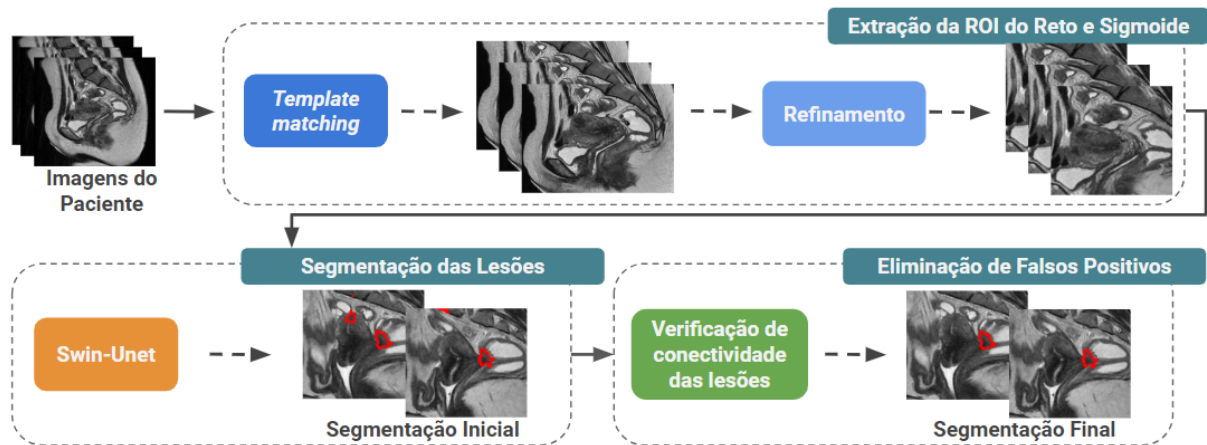
3.6 Considerações Finais

Este capítulo apresentou a fundamentação teórica do trabalho desenvolvido. Foram percorridos todos os conceitos necessários para a compreensão do problema abordado, do método desenvolvido e da base de imagens utilizada. No próximo capítulo será descrito o método proposto.

4 Método Proposto

Este capítulo apresenta o método proposto para a segmentação das lesões de endometriose. O método é ilustrado na Figura 17 e consiste em três etapas. Primeiro, é utilizado processamento de imagens para a extração da região de interesse (ROI) das imagens, para diminuir a área a ser processada pela rede profunda. Em seguida, utiliza-se a arquitetura Swin-Unet para segmentar as lesões. Então, é realizada uma etapa de eliminação de falsos positivos da segmentação, através de uma técnica baseada na conectividade entre as lesões de cada paciente.

Figura 17 – Esquema do método proposto.



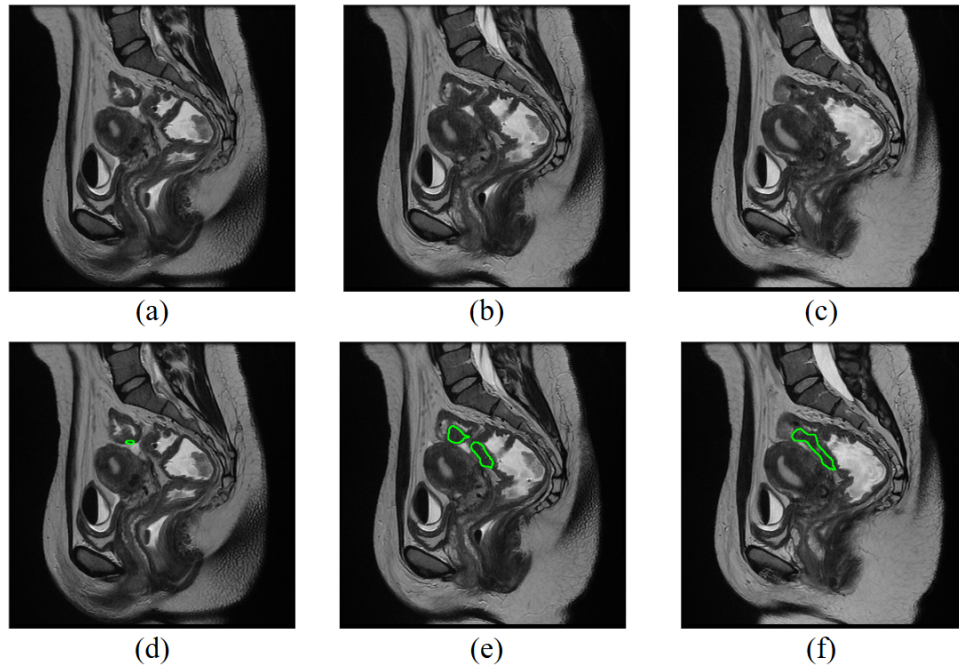
Fonte: Elaborado pelo autor.

4.1 Extração da Região de Interesse

Na Figura 18 é possível ver exemplos de imagens de RM da região pélvica. São três imagens do exame de uma mesma paciente com endometriose. As imagens (a), (b) e (c) correspondem às imagens (d), (e) e (f), respectivamente. As imagens (d), (e) e (f) estão com a área afetada por endometriose marcada em verde. Nesse exemplo é possível perceber a diferença em tamanho, formato e conectividade das lesões, mesmo em imagens consecutivas do exame. Além disso, percebe-se a dificuldade de identificar a doença a partir das imagens, pois apresentam cor e textura semelhantes a partes de outros órgãos em volta.

As imagens possuem muitas áreas que não agregam informação à detecção de endometriose, como a parede abdominal, a área externa à paciente e os órgãos acima do reto e sigmoide. Manter essas áreas torna a tarefa mais custosa, computacionalmente, sem benefícios. Além disso, regiões irrelevantes da imagem dificultam a capacidade de

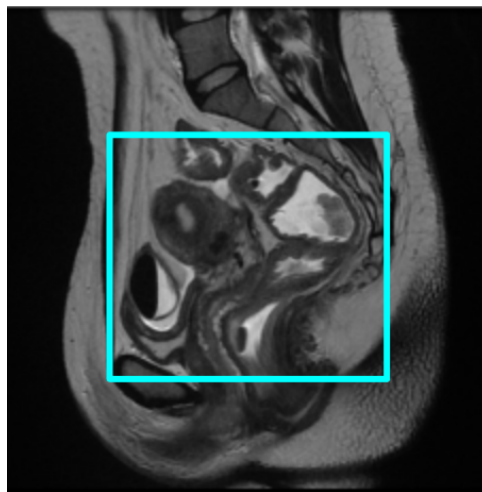
Figura 18 – Fatias consecutivas de uma paciente, que apresentam lesão. As marcações das lesões pelo especialista estão em verde.



Fonte: Elaborado pelo autor.

generalização da rede, que pode atentar-se a detalhes de pouca importância durante o treinamento. Portanto, definiu-se uma etapa de pré-processamento para recorte das imagens da base, de forma a manter apenas a região do reto e sigmoide. Essa região, chamada de região de interesse (ROI), é indicada pelo quadrado azul na Figura 19. O método utilizado foi proposto por [Figueredo et al. \(2023\)](#) e é baseado em processamento de imagens.

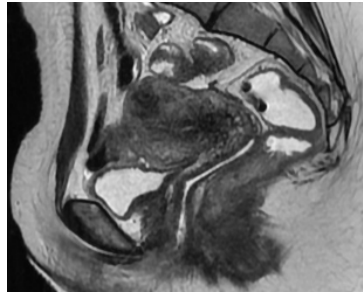
Figura 19 – Imagem da base com a região de interesse marcada em azul.



Fonte: Elaborado pelo autor.

Primeiro realiza-se um *template matching* (Seção 3.3.2). Para ser a imagem de *template* foi selecionada, empiricamente, uma imagem central de um dos volumes de RM, onde a região do reto e sigmoide estava bem visível. A ROI dessa imagem foi manualmente demarcada. Diferenças na posição e tamanho dos órgãos, bem como diferenças na escala de cinza entre as imagens diminuem a eficácia do *template matching*. Assim, foi necessário incluir no *template* a parede abdominal, fornecendo mais informações de posicionamento. O *template* utilizado é apresentado na Figura 20. Os autores desse método definiram empiricamente que a variação de tamanho do *template* na execução deveria ser de aumentar em 2% deste tamanho e também diminuir 2%. Esses valores englobam com eficácia a variação de tamanho e posicionamento das ROIs.

Figura 20 – *Template* utilizado na etapa de *template matching*.



Fonte: Elaborado pelo autor.

Para cada exame, o *template matching* é realizado em todas as suas imagens, até encontrar a região mais semelhante entre todas. Uma vez encontrada esta ROI primária, é necessário refiná-la, ou seja, remover a parte externa ao abdômen, além da parede abdominal. Primeiro binariza-se a imagem utilizando a técnica de Otsu 3.3.3, que separa áreas escuras das claras. A Figura 21 apresenta o resultado da binarização em uma ROI primária definida pelo *template matching* aplicado no exame ao qual pertence a imagem da Figura 19. A seta vermelha aponta a parede abdominal. A parte externa do abdômen e o músculo abdominal viram áreas totalmente escuras, enquanto a parede abdominal e os órgãos logo após o músculo abdominal viram áreas totalmente claras.

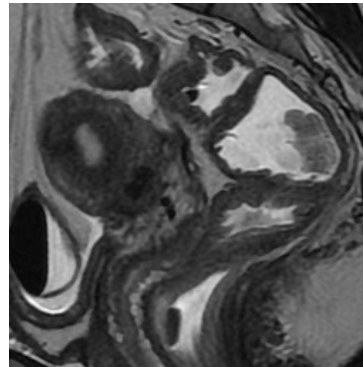
Figura 21 – Exemplo de ROI primária binarizada.



Fonte: Elaborado pelo autor.

Em seguida realiza-se um caminharmento pixel a pixel no eixo x , partindo do canto superior esquerdo da ROI primária, até encontrar a parede abdominal, delimitada pelo primeiro pixel branco encontrado. Em seguida prossegue-se o caminharmento, buscando o músculo abdominal, encontrado ao chegar no próximo pixel preto. Novamente, segue-se caminhando para encontrar o limite do músculo, que é o último pixel preto encontrado. Enfim chega-se no limite esquerdo da ROI final desejada. Então, define-se uma área de 256×256 , partindo do ponto no qual o caminharmento parou. Esse tamanho de ROI reduz o tamanho da imagem pela metade, enquanto tamanhos menores resultam em perdas de área de lesão. Obtém-se, a partir disso, a ROI refinada, exibida na Figura 22. Para todas as demais imagens do exame a ROI é extraída na mesma posição e com o mesmo tamanho da definida na imagem que foi mais similar ao *template*.

Figura 22 – Exemplo de ROI refinada.



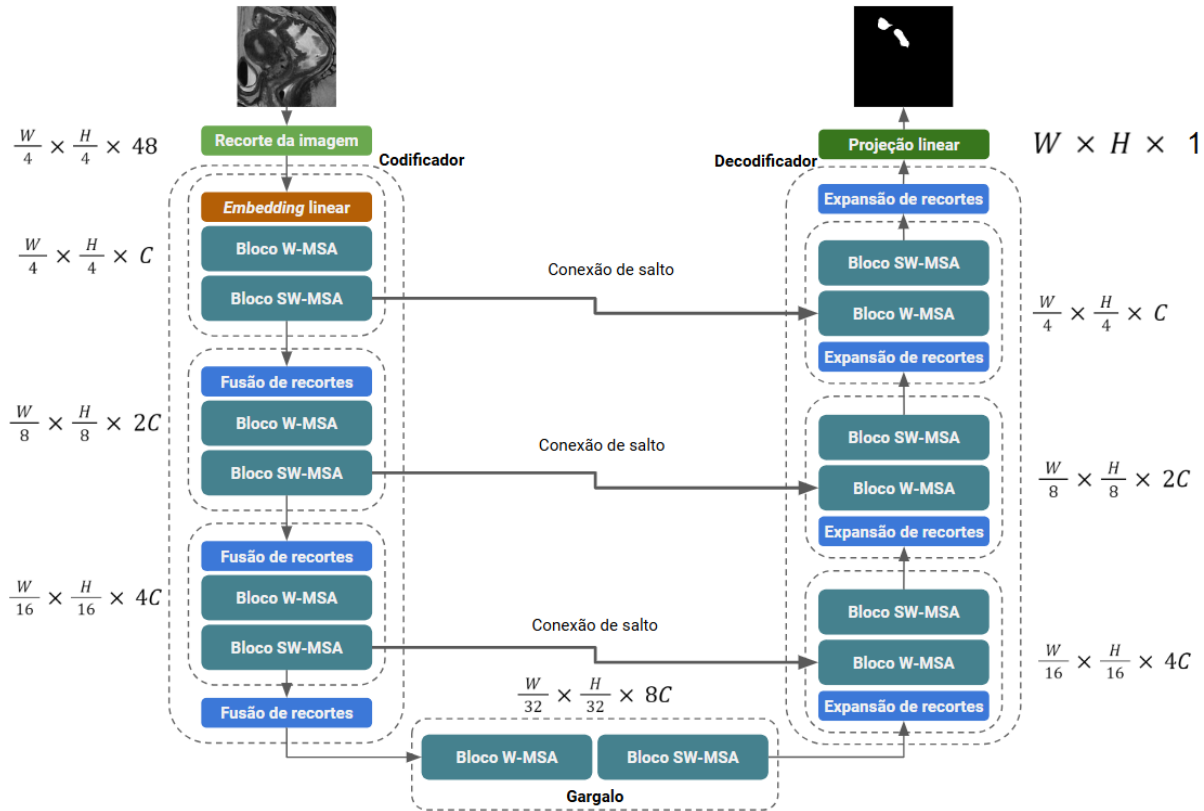
Fonte: Elaborado pelo autor.

4.2 Segmentação das Lesões

Nesta etapa é realizada a segmentação das lesões de endometriose pela Swin-Unet Tiny (CAO et al., 2022; LIU et al., 2021) treinada no conjunto de imagens com lesão da base. Essa rede é semelhante às redes completamente convolucionais, possuindo um codificador, gargalo e decodificador. Como a U-Net (RONNEBERGER; FISCHER; BROX, 2015), possui conexões de salto entre cada camada do codificador e sua respectiva camada de decodificador. Porém, ao invés de camadas convolucionais, a Swin-Unet é totalmente composta de camadas de Swin Transformer (Seção 3.4.13). A arquitetura dessa rede é apresentada na Figura 23.

Antes de passar pelas camadas da rede, a imagem de entrada é primeiro dividida em recortes quadrados de mesmo tamanho. Ao entrarem na rede, sequencialmente, esses recortes são dispostos no formato de vetores 1D, através do processo de *embedding* proposto pelos autores da rede. Na Figura 23 é mostrado que, para o tamanho de recorte de 4×4 , cada vetor entra com tamanho $4 \times 4 \times 3 = 48$, sendo que o valor 3 é a quantidade de

Figura 23 – Arquitetura da Swin-UNet.



Fonte: Elaborado pelo autor.

canais da imagem. Os vetores passam por uma camada de *embedding*, que é uma camada convolucional, treinável, cuja saída é o conjunto de vetores, ou *embeddings*, de um mesmo tamanho C que pode ser definido arbitrariamente.

Os *embeddings* são inseridos na primeira camada de Swin Transformer (explicada na Seção 3.4.13), onde passam primeiro pela subcamada de W-MSA e em seguida pela SW-MSA. Esse processo permite que características globais e locais da imagem sejam processadas com pesos relevantes, mas sem precisar de uma autoatenção extremamente custosa recorte a recorte. Ainda assim, obtém-se o poder da autoatenção em atribuir pesos maiores a regiões mais relevantes dos mapas de características, como explicado na Seção 3.4.10. Dessa forma busca-se identificar características tanto das áreas de interesse menores quanto das maiores, necessário ao tentar identificar lesões de tamanhos e formatos muito variados. Além disso, espera-se relacionar melhor a posição das lesões no reto e sigmoide, tendo a visão global da imagem, característica permitida pela arquitetura *transformer*.

A última camada da rede é também uma camada convolucional, por onde cada vetor de características do decodificador passa sequencialmente. Esses vetores são organizados de volta para duas dimensões e passam por uma função sigmoide, que gera as probabilidades de haver lesão em cada pixel, a partir das características. Utiliza-se, então, o limiar de

pixels com 50% ou mais de probabilidade de lesão, para ser definida uma máscara binária da segmentação, onde os pixels positivos apresentam a doença.

A Swin-Unet foi selecionada para o método após serem realizados testes com várias arquiteturas. A Swin foi a rede mais eficaz. Isso tem conformidade com a literatura, que demonstra que redes baseadas em Swin Transformer obtém resultados de estado-da-arte em suas aplicações (LIU et al., 2021; CAO et al., 2022). A Swin-Unet tem melhor desempenho quando seus pesos são inicializados através de transferência de aprendizado (Seção 3.4.6), como demonstrado por Cao et al. (2022). Essa característica é comum às *transformers*, principalmente quando são treinadas com bases de dados pequenas, como é o caso da base deste trabalho. Juntamente disso, utilizar a variação da arquitetura Tiny de Swin permite não só ter menor custo computacional ao treinar a rede, como também permite utilizar seus pesos pré-treinados na ImageNet, disponibilizados pelos autores da Swin-Unet.

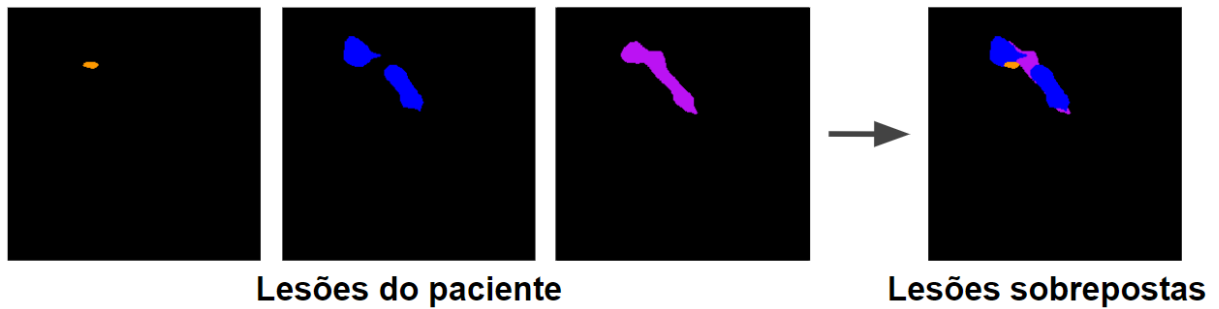
4.3 Eliminação de Falsos Positivos

Observa-se que, para o especialista responsável por identificar as lesões de endometriose, é importante obter alta precisão ao demarcar as áreas de lesão. A demarcação precisa dessas áreas oferece melhores referências no momento de realizar tratamentos cirúrgicos da doença. Porém, a segmentação das lesões pode gerar falsos positivos, que são áreas cujos pixels foram marcados pela rede neural como apresentando lesão, quando na verdade não apresentam. Isso diminui a eficácia do modelo de segmentação como auxílio ao especialista, ao exibir muitas áreas sem lesão que este precisará analisar. Áreas de textura e cor semelhante às de lesão podem estar presentes em outras partes da imagem, levando o modelo ao erro. Além disso as lesões muito pequenas, desconsideradas no protocolo de marcação das imagens (Seção 5.1.1), podem estar influenciando o modelo. Assim, faz-se necessário uma etapa de eliminação de falsos positivos. Com essa etapa espera-se aumentar a precisão da segmentação, enquanto preservam-se os positivos verdadeiros.

A característica das lesões é que elas podem se estender pelo volume da RM de uma paciente em múltiplas imagens consecutivas, como foi possível observar na Figura 18. Assim, a área de lesão em uma imagem apresenta uma conectividade de seus pixels com áreas de lesão de imagens anteriores e posteriores. É possível observar isso na Figura 24, que apresenta as máscaras binárias das lesões de imagens consecutivas de uma mesma paciente, sobrepostas. É visível que a sobreposição das máscaras demonstra a conectividade entre as lesões. Dessa forma, entende-se que segmentações realizadas pelo modelo em um mesmo paciente devam apresentar essa conectividade para serem consideradas corretas. Existem lesões muito pequenas que podem ser segmentadas, mas que não foram consideradas no momento da marcação. Se estiverem em regiões minimamente distantes das que possuem lesões marcadas, serão menores as chances do modelo segmentá-las erroneamente em mais

de uma imagem, portanto não apresentarão conectividade.

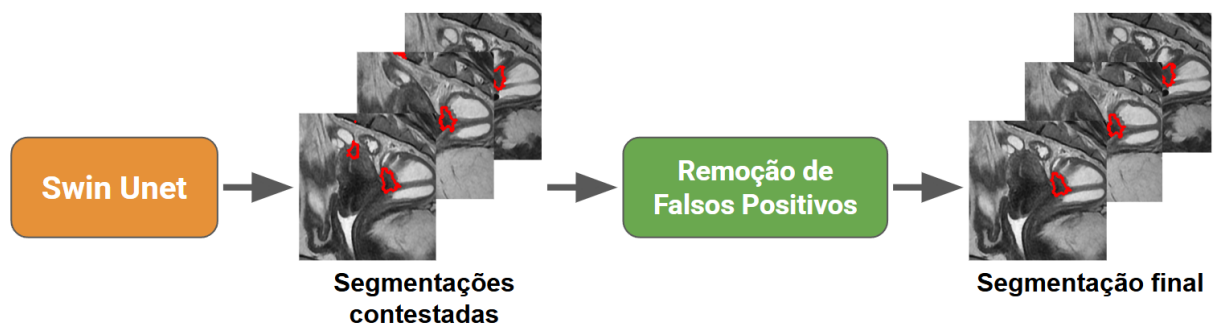
Figura 24 – Máscaras binárias das imagens com lesão de um paciente e sua sobreposição.



Fonte: Elaborado pelo autor.

Na Figura 25 é apresentado o efeito da remoção dos falsos positivos no fluxo do método. Na etapa descrita na seção anterior a rede neural segmenta, em vermelho, objetos menores e espalhados pelas imagens. Após a etapa de remoção de falsos positivos, são mantidas apenas as lesões mais relevantes, conexas entre si. Entretanto, deve-se ter cautela para não perder segmentações corretas de lesões das pacientes. Um paciente que tenha muitas imagens apresentando áreas de lesão pode ter, por exemplo, uma dessas áreas na sua imagem inicial conexa com a imagem seguinte, mas desconexa com as últimas. A Figura 24 é um exemplo disso, onde a máscara azul possui uma área de lesão que é desconexa da lesão da máscara laranja. Além disso, o modelo treinado para a etapa de segmentação (Seção 4.2) considera apenas as informações da imagem que está processando, sem contexto tridimensional, portanto pode segmentar regiões de lesão de forma pouco conexa entre as imagens da paciente.

Figura 25 – Remoção dos falsos positivos.



Fonte: Elaborado pelo autor.

Com base nessa análise, é considerada a conectividade entre segmentações de duas imagens diferentes quando há interseção de ao menos um pixel entre elas, quando sobrepostas. As segmentações iniciais obtidas são, a partir daqui, chamadas de segmentações contestadas. A partir dessas definições foi estabelecido um critério de conectividade entre

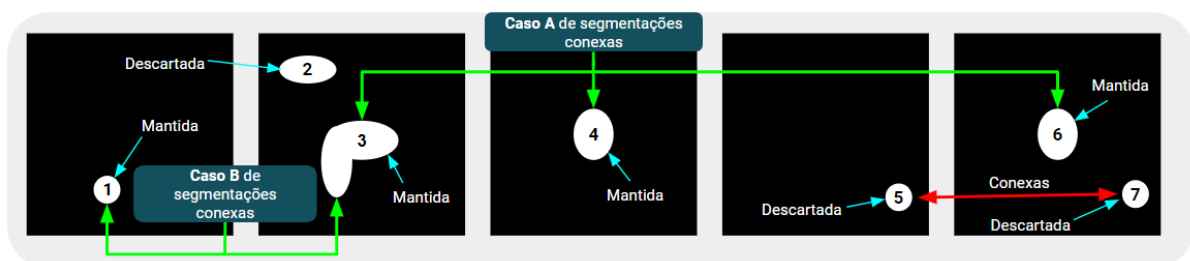
as segmentações para decidir quais delas manter e quais descartar. Esse critério é composto de duas regras, onde a concordância de uma segmentação contestada a qualquer uma delas a valida como correta:

1. Uma segmentação contestada, para que seja considerada correta, deve apresentar conectividade com outras segmentações através de, no mínimo, 50% das imagens com lesão da paciente, contando com a imagem da segmentação sendo analisada. Essas imagens não precisam ser consecutivas;
2. Toda a segmentação contestada que seja conexa com ao menos uma segmentação que já adequou-se à primeira regra e seja pertencente a uma imagem adjacente à contestada, também é preservada.

As segmentações contestadas que não adequam-se a nenhuma das regras do critério são descartadas. Essas regras são verificadas uma após a outra, na ordem numérica da lista. O limite mínimo de 50% de imagens conexas na primeira regra faz com que pacientes que possuam apenas uma ou duas imagens com lesão tenham todas as suas segmentações mantidas. Nesses casos não há informação suficiente para determinar erros na segmentação. A segunda regra leva em consideração a possibilidade de lesões relevantes não estarem conexas com as de imagens além da sua adjacente. Se a adjacente adequa-se à primeira regra, maiores são as chances de que a contestada em questão seja verdadeira.

A Figura 26 simula o comportamento da eliminação de falsos positivos em um paciente que possui 5 imagens com lesão. Os 7 objetos brancos numerados são exemplos das segmentações geradas pela rede profunda. O caso A de segmentações conexas exemplifica segmentações que adedquam-se à primeira regra do critério de conectividade. Ocorre conectividade entre as segmentações 3, 4 e 6. Cada uma delas possui conectividade com outras ao longo de no mínimo 50% das imagens da paciente (3 imagens de 5), portanto são mantidas. Deve-se perceber que a imagem que contém a segmentação 6 não é adjacente às demais, porém isso não fere a primeira regra do critério.

Figura 26 – Simulação do comportamento da eliminação de falsos positivos.



Fonte: Elaborado pelo autor.

O caso *B* exemplifica uma segmentação que não adequa-se à primeira regra, mas adequa-se à segunda. Nele, a conectividade ocorre entre as segmentações 1 e 3. A 1 seria descartada por não passar na primeira regra do critério de conectividade. Porém, é conexa com a segmentação 3, na imagem adjacente à sua, portanto é mantida por adequar-se à segunda regra do critério, afinal a segmentação 3 passou na primeira regra. As segmentações 5 e 7 são conexas com segmentações contestadas em menos de 50% das imagens com lesão (2 de 5), portanto são descartadas. A segmentação 2 não é conexa com nenhuma outra, então também não será mantida.

5 Resultados

Este capítulo apresenta a base de imagens utilizada no trabalho, as configurações dos experimentos realizados e os resultados obtidos. Esses resultados são discutidos ao longo do capítulo, primeiramente analisando o desempenho da segmentação e, em seguida, da remoção de falsos positivos. Casos de uso são apresentados para corroborar a análise.

5.1 Base de Imagens

A base de imagens utilizada neste trabalho é a mesma citada por [Figueredo et al. \(2023\)](#), mas com novos volumes de pacientes com endometriose. É composta de imagens de RM, ponderadas em T2, da pelve feminina, no plano sagital. Essa base é privada e foi desenvolvida em conjunto pela Clínica Fonte Imagem Medicina Diagnóstica e o Departamento de Ginecologia do Hospital Universitário Pedro Ernesto, localizados no Rio de Janeiro, Brasil. A clínica é coordenada pela Dra. Alice Brandão, uma das especialistas responsáveis pelas marcações das lesões. As imagens foram adquiridas utilizando um aparelho de RM SIGNA™ Artist 1.5T da GE Healthcare. Os exames desta base estão anonimizados, garantindo a proteção da privacidade das pacientes.

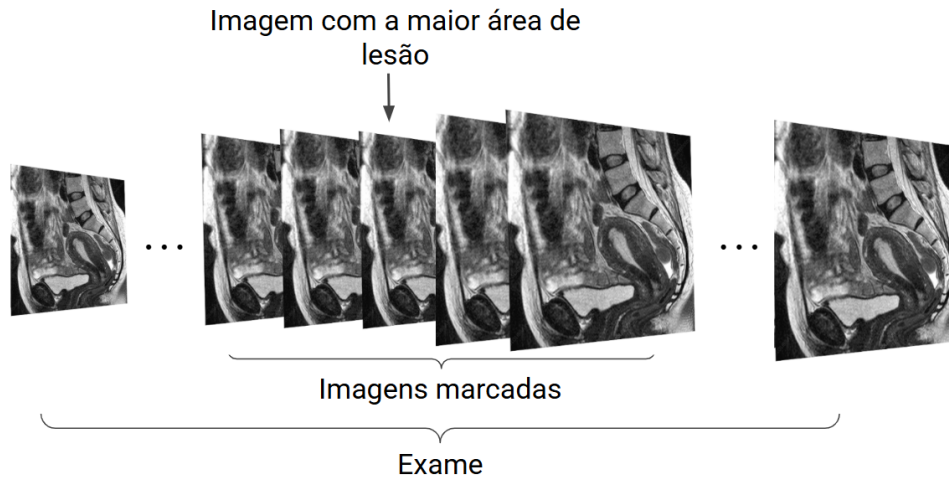
A base de imagens inclui 179 exames de RM, sendo 107 de pacientes com lesões de endometriose e 72 de pacientes consideradas saudáveis. Os exames são volumes em escala de cinza, ou seja, cada pixel de cada fatia contém um valor de 0 a 255. Os exames contêm em média 32 fatias (imagens), com cada uma medindo 512×512 pixels. O total de imagens é 5.730, sendo que 451 têm lesões de endometriose marcadas. Todas as lesões são na área do reto e sigmoide. O menor tamanho de lesão marcada é dois centímetros, mas podem haver lesões menores sem marcação. Cada paciente com endometriose possui de 1 a 9 imagens apresentando lesão.

5.1.1 Protocolo de Marcação das Imagens

O protocolo de marcação foi definido para equilibrar o esforço e o tempo necessários para anotações precisas, considerando a dificuldade em distinguir essas lesões devido à sua semelhança com outras estruturas anatômicas nas regiões do retossigmoide. Primeiramente, a imagem onde a lesão é mais visível é anotada. Em seguida, as duas imagens anteriores e posteriores a essa imagem primária também são marcadas, com base na avaliação do especialista sobre a dificuldade da marcação e a visibilidade da lesão. A Figura 27 apresenta o protocolo. O trabalho de marcação das lesões foi realizado por três especialistas, sendo um com mais de 25 anos de experiência no diagnóstico e tratamento de endometriose, e os

outros dois com aproximadamente três anos de experiência. Cada marcação feita por um especialista foi revisada e avaliada por outro.

Figura 27 – Protocolo de marcação da base de imagens.



Fonte: Elaborado pelo autor.

5.2 Descrição dos Experimentos

Para avaliar a robustez da etapa de segmentação das lesões, utilizou-se a técnica *k-fold cross-validation*, validada por Wong (2015) como essencial para verificar a confiabilidade de um método. As pacientes da base de imagens foram divididos em conjuntos de treinamento e teste, ou seja, cada paciente teve todas as suas imagens destinadas ao conjunto do qual ele participa. O conjunto de treinamento foi composto por 87 pacientes (369 imagens) e o de teste por 18 (82 imagens). As pacientes de treinamento foram divididos em 5 *folds*. Foram realizados 5 treinamentos da Swin-Unet, cada vez com uma *fold* sendo utilizada para validação e as demais para treinamento. Cada treinamento teve limite máximo de 100 épocas e o melhor modelo de cada *fold* foi selecionado pela menor perda de validação. O conjunto de teste foi usado para testar cada melhor modelo.

Os hiperparâmetros de treinamento foram tamanho de lote 32 e otimizador SGD com taxa de aprendizado de 0,05. O tamanho de recorte da imagem foi de 4×4 pixels, utilizando janelas de 7×7 recortes. A função de perda utilizada foi uma combinação de Perda Cross-Entropy e Perda Dice, com 40% e 60% de peso, respectivamente, proposta por Cao et al. (2022). As imagens foram redimensionadas para 224×224 pixels, o tamanho de entrada necessário para utilização dos hiperparâmetros de tamanho de recorte e janela da Swin-Unet.

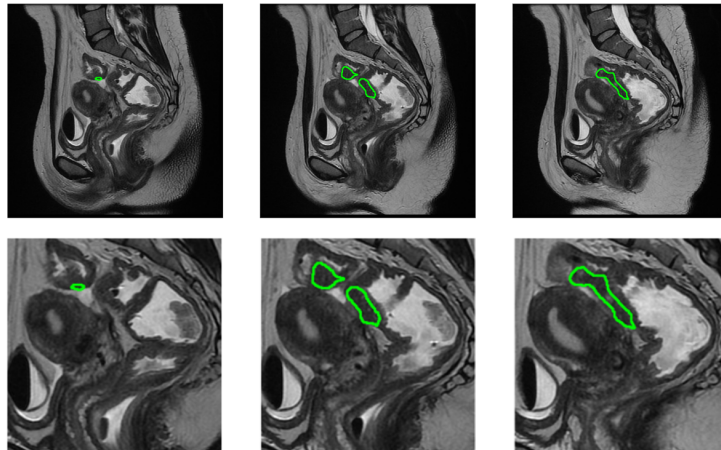
Foi utilizado aumento de dados *online*, ou seja, cada lote de imagens sofreu transformações aleatórias antes de ser inserido na rede. Isso é feito para aumentar a

variabilidade do conjunto de dados, algo essencial para *transformers* (DOSOVITSKIY et al., 2020). As transformações utilizadas foram rotação de até 20° , nos dois sentidos, chance de giro horizontal e vertical, até 20% de aumento ou diminuição das dimensões da imagem, bem como até 20% de aumento ou diminuição de brilho. O treinamento foi realizado em uma GPU NVIDIA GeForce RTX 3060, no sistema operacional Ubuntu 20.04, utilizando o *framework* PyTorch ¹.

5.3 Resultados da Extração da Região de Interesse

A etapa de extração da ROI foi bem-sucedida em manter uma área menor de 256×256 pixels para toda paciente, sem perder qualquer área das lesões. Na Figura 28 podemos observar exemplos do resultado, onde cada imagem na primeira linha corresponde à imagem diretamente abaixo. As lesões de endometriose, demarcadas pelo especialista, estão apresentadas em verde. Percebe-se que muita informação desnecessária foi eliminada das imagens, preservando apenas a região do reto e sigmoide.

Figura 28 – Imagens e suas respectivas ROIs delimitadas pelo método.



Fonte: Elaborado pelo autor.

5.4 Resultados da Segmentação das Lesões

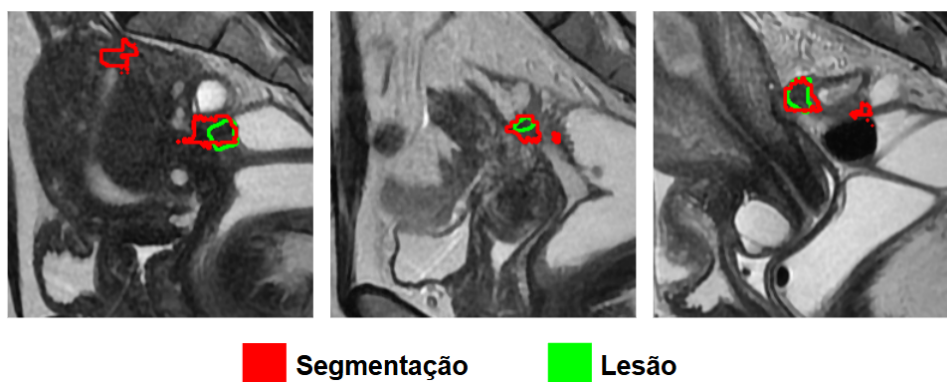
Primeiramente analisou-se o desempenho dos modelos treinados na perspectiva da base de imagens inteira, aqui chamada de análise "por imagem". Ou seja, para cada imagem de teste segmentada, foram aferidas as métricas (Seção 3.5) comparando cada máscara de segmentação com sua respectiva máscara real definida pelos especialistas. A análise por imagem é realizada ao calcular a média das métricas entre todas as imagens do conjunto de teste. Para o resultado médio da *k-fold cross-validation* foram calculadas,

¹ Disponível em: <<https://pytorch.org/>>

entre as 5 *folds*, as médias das suas métricas obtidas em suas respectivas análises por imagem. Obteve-se precisão de $46,6\% \pm 4,0\%$, sensibilidade de $57,0\% \pm 3,2\%$, Dice de $46,0\% \pm 1,9\%$ e Jaccard de $35,2\% \pm 1,8\%$.

Esses resultados demonstram certo sucesso do modelo em segmentar as lesões da base de imagens, independentemente de quais imagens são utilizadas para treinamento e validação. Percebe-se que foi alcançada sensibilidade média bem maior que a precisão. Como apresentado na Figura 29, a Swin-Unet acaba gerando segmentações desconexas e distantes da real área de lesão. Isso ocorre frequentemente em regiões da imagem em que a textura e cor são muito semelhantes às lesões. Além disso, a doença tem a característica de poder ocorrer em várias regiões diferentes, aumentando a variabilidade do problema. A seguir será utilizado o melhor modelo de Swin-Unet obtido no *k-fold cross-validation* para as próximas etapas de análise do método. Esse modelo alcançou o maior Dice no teste, de 47,7%. Sua precisão foi 45,6%, sensibilidade 61,9% e Jaccard 36,2%.

Figura 29 – Exemplos da segmentação inicial das lesões.



Fonte: Elaborado pelo autor.

A análise por imagem generaliza os casos de teste. Isso pode ser um problema pois pacientes diferentes possuem lesões de tamanhos e formatos muito diferentes, conferindo uma característica de singularidade a cada caso. Assim, é possível que grandes erros em pacientes muito difíceis prejudiquem o desempenho geral analisado por essa abordagem. Dessa forma, torna-se interessante uma análise dos resultados por paciente. Ou seja, para cada paciente obtém-se as médias das métricas de suas imagens. Em seguida, é calculada a média das métricas entre as pacientes. A comparação dos resultados dessa análise com a análise por imagem são apresentados na Tabela 2. Percebe-se pouca diferença em comparação com os resultados por imagem. Isso mostra que os erros e acertos do modelo estão bem distribuídos entre as pacientes. Ou seja, toda paciente possui lesões mais difíceis e mais fáceis de segmentar.

Na Seção 5.1.1 foi explicado que o protocolo de marcação parte da imagem com a lesão mais visível do exame e, em seguida, marca as lesões nas imagens adjacentes. O

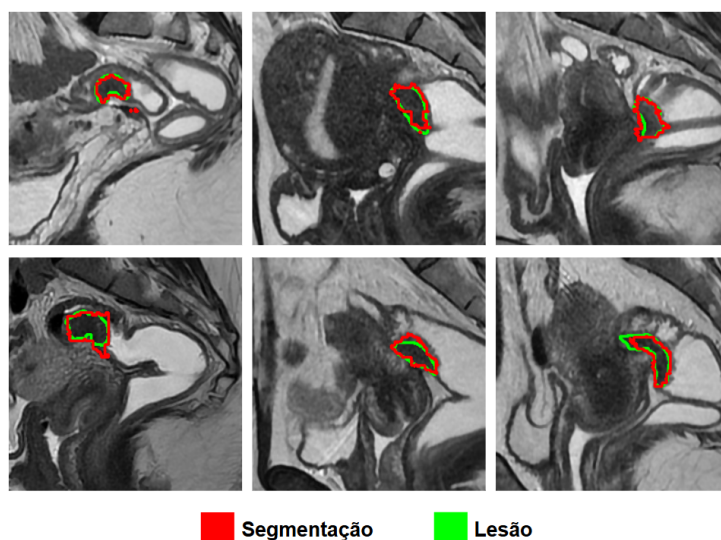
Tabela 2 – Resultado médio da predição do modelo sobre o conjunto de teste.

Análise	Precisão	Sensibilidade	Dice	Jaccard
Por Imagem	45,6%	61,9%	47,7%	36,2%
Por Paciente	44,3%	62,7%	48,0%	36,6%

primeiro desafio dos especialistas é encontrar essa primeira lesão em exames com uma média de 32 imagens. Assim, entende-se que se o modelo for capaz de segmentar com grande eficácia ao menos uma imagem do exame de MR da paciente, já tornará mais eficiente o trabalho do profissional, pois este poderá encontrar a área total de lesão a partir dessa fatia segmentada. Ou seja, o especialista não precisará buscar a lesão no volume inteiro, e poderá focar em áreas específicas das imagens indicadas pelas segmentações.

Sendo assim, é feita também a análise das métricas de melhor segmentação de cada paciente. Esta análise envolve selecionar para toda paciente uma imagem cuja segmentação obteve o maior valor de Dice. Em seguida, é calculada a média do Dice entre as pacientes apenas levando em consideração essas imagens selecionadas. Essa abordagem traz um Dice médio de 72,6%. Como a sensibilidade é uma métrica extremamente importante para métodos da área médica, foi calculada a média da sensibilidade desses melhores casos, obtendo a sensibilidade média de 84,5%. Essa análise estabelece uma boa eficácia do modelo em realizar uma marcação de grande qualidade em ao menos uma das imagens de cada paciente. Na Figura 30 são apresentadas as melhores segmentações de lesão de seis pacientes da base.

Figura 30 – Melhores segmentações de diferentes pacientes da base.

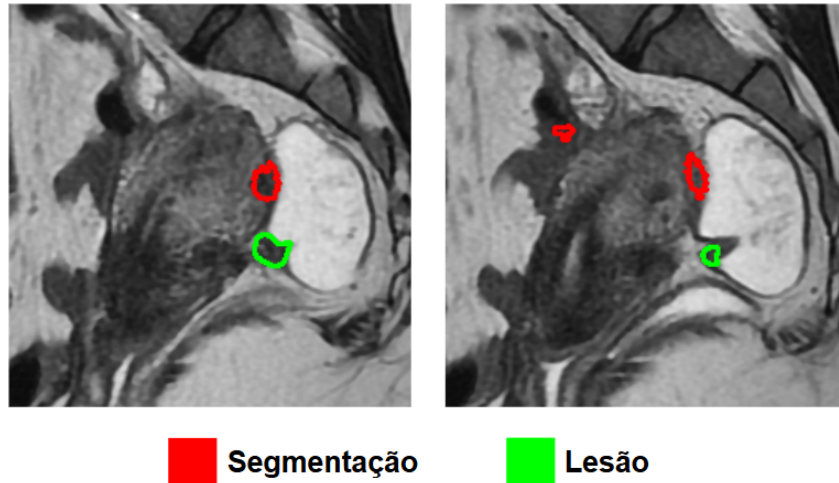


Fonte: Elaborado pelo autor.

É importante destacar que o modelo foi capaz de segmentar ao menos uma lesão em 17 das 18 pacientes de teste. Isso significa que para a maioria dos casos o protocolo de marcação das lesões pode ser aplicado partindo de uma indicação da localização das

mesmas. Na Figura 31 é apresentado o Estudo de Caso 1, a única paciente no qual o modelo falhou em segmentar qualquer lesão. Como é possível observar, a paciente possui uma lesão bem pequena, visível em apenas duas imagens de seu exame. Esse caso é extremamente desafiador para a segmentação, além de não poder ser melhorado pela etapa seguinte do método. Ainda assim, a Swin-Unet tentou segmentar lesões, demarcando áreas escuras da imagem que se assemelham muito à endometriose.

Figura 31 – Estudo de Caso 1



Fonte: Elaborado pelo autor.

5.4.1 Comparação com Outras Arquiteturas

Para comprovar a vantagem da Swin-Unet em comparação a outras arquiteturas para a etapa de segmentação das lesões, foram treinadas também U-Nets utilizando diferentes *backbones* comumente utilizados para segmentação, além de novas arquiteturas de *autoencoders* baseadas em Mamba. Todas as redes de comparação foram treinadas com os mesmos hiperparâmetros da Swin-Unet descritos na Seção 5.2, além de também utilizarem as ROIs extraídas pela etapa anterior do método. A única exceção é a resolução das imagens, que foi mantida em 256×256 , pois essas arquiteturas contam com camadas iniciais específicas para redimensionar a imagem de entrada para o tamanho necessário. Todas as redes foram pré-treinadas na ImageNet (DENG et al., 2009). Além disso, utilizou-se aumento de dados em todos os treinamentos. Para cada rede neural, foram aferidas a média e desvio padrão do resultado do da inferência do melhor modelo treinado de cada *fold* sobre o conjunto de teste. Foi utilizada apenas a análise por imagem.

As arquiteturas testadas como *backbone* de U-Net foram a VGG-16 (SIMONYAN; ZISSERMAN, 2015), ResNet50, ResNet152 (HE et al., 2016), Efficient Net B7 (TAN; LE, 2019), DenseNet121 (HUANG et al., 2017). Nessas redes, o codificador da U-Net é substituído pelo codificador da rede cujo nome é citado. As arquiteturas baseadas em

Mamba testadas foram a RM-Unet (TANG et al., 2024), VM-Unet (RUAN; LI; XIANG, 2024) e VM-UNET-V2 (ZHANG et al., 2024).

A comparação dos resultados é apresentada na Tabela 3. Percebe-se que todos os modelos foram muito menos capazes em aprender as características necessárias para a segmentação da endometriose. Ela reforça a dificuldade do problema e demonstra que a Swin-Unet pré-treinada tem maior eficácia na tarefa, tanto comparada a uma CNN convencional, quanto comparada com outras abordagens modernas de aprendizado profundo. Os modelos que zeraram as métricas sequer foram capazes de aprender, mesmo utilizando transferência de aprendizado.

Uma provável razão para a falha de aprendizado das outras arquiteturas, além da dificuldade do problema, é o tamanho reduzido da base de imagens, insuficiente para um treinamento eficaz. Já o sucesso da Swin-Unet possivelmente dá-se pela capacidade dessa arquitetura processar o contexto local e global das imagens. Ou seja, a rede processa, através da autoatenção limitada aos recortes por janela, o contexto local. Em imagens, frequentemente elementos que estão relacionados encontram-se próximos uns dos outros. Porém também há muita informação relevante na visão global da imagem, a qual a Swin consegue processar após deslizar as janelas. Assim, o contexto da Swin-Unet pode conter tanto informações detalhadas sobre as lesões em relação a seu entorno, como também em relação à sua localização nos órgãos da ROI.

Tabela 3 – Resultado médio do *k-fold cross-validation* de cada modelo sobre o conjunto de teste.

Arquitetura	Precisão	Sensibilidade	Dice	Jaccard
Swin-Unet	46,6% ± 4,0%	57,0% ± 3,2%	46,0% ± 1,9%	35,2% ± 1,8%
VGG-16	14,6% ± 13,1%	16,1% ± 22,8%	10,5% ± 11,2%	6,9% ± 7,4%
Efficient Net B7	10,6% ± 15,0%	4,6% ± 6,5%	5,2% ± 7,3%	3,6% ± 5,1%
ResNet50	1,3% ± 2,2%	0,8% ± 1,2%	0,7% ± 1,0%	0,5% ± 0,7%
ResNet152	2,9% ± 3,9%	3,0% ± 5,6%	1,9% ± 2,8%	1,2% ± 1,8%
DenseNet121	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%
RM-Unet	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%
VM-Unet	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%
VM-UNET-V2	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%	0,0% ± 0,0%

5.5 Resultados da Eliminação de Falsos Positivos

Ao aplicar a etapa de eliminação de falsos positivos nas segmentações obtidas, foram alcançados os resultados que podem ser encontrados na Tabela 4. A coluna “FP Eliminados” indica se foi realizada a eliminação de falsos positivos antes de aferir as métricas. Comparando os resultados com os da etapa de segmentação das lesões, é possível ver que há uma pequena melhora nas métricas, menos na sensibilidade que sofre uma

leve piora. Porém, a melhora das demais métricas é de maior magnitude que a piora da sensibilidade.

Tabela 4 – Comparação de resultados da etapa de segmentação das lesões com a de eliminação de falsos positivos

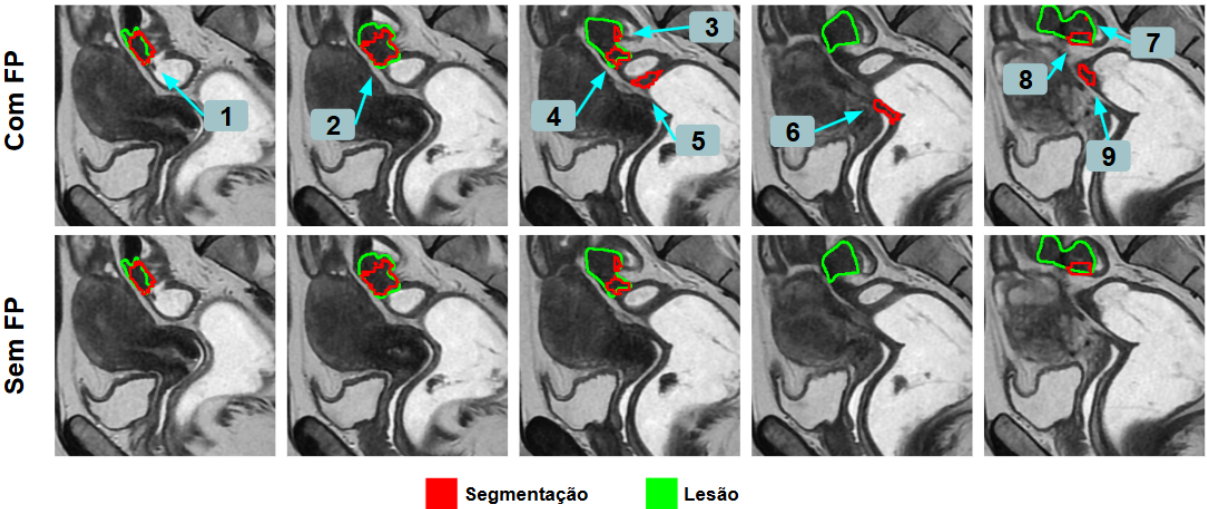
Análise	FP Eliminados	Precisão	Sensibilidade	Dice	Jaccard
Por Imagem	Não	45,6%	61,9%	47,7%	36,2%
	Sim	54,1%	61,7%	48,7%	37,2%
Por Paciente	Não	44,3%	62,7%	48,0%	36,6%
	Sim	47,9%	62,5%	48,9%%	37,6%

A média do Dice das imagens com as melhores segmentações de cada paciente aumentou 1%, de 72,6% para 73,6%. A média da sensibilidade das melhores segmentações continuou sendo 84,5%. Dessa forma, há o indício de que a remoção de falsos positivos afeta positivamente os resultados da segmentação, eliminando erros sem perder muitas lesões verdadeiras. Foi realizada, então, uma análise das melhores segmentações obtidas pelo método. Observando quais lesões foram as melhores segmentadas, percebeu-se que, das 17 pacientes que tiveram alguma lesão segmentada, em 9 delas (isto é, 53% das pacientes) a melhor segmentação ocorreu na maior lesão da paciente. Em 6 pacientes (35%) a segunda maior lesão foi a melhor segmentada, sendo que em 3 das 6 a maior lesão foi a segunda melhor segmentada. Apenas 2 pacientes (12%) não apresentaram nenhum padrão interessante nas melhores segmentações. Isso mostra que o método está agindo de forma semelhante ao protocolo do especialista (Seção 5.1.1). Este fato torna o método promissor para um sistema de CAD ou CADx, pois a partir da melhor lesão segmentada automaticamente o especialista pode determinar as demais áreas de lesão da paciente.

O comportamento da técnica de eliminação de falsos positivos (Seção 4.3) pode ser observado de forma bem clara no Estudo de Caso 2 (Figura 32). Esse Estudo de Caso apresenta as segmentações em cinco imagens consecutivas de uma paciente. Na linha “Com FP” estão os resultados da etapa da segmentação de lesões e na linha “Sem FP” estão as segmentações mantidas pela etapa de eliminação de falsos positivos. As segmentações estão numeradas, seus números indicados pelas setas azuis. As segmentações 1, 2, 3 e 8 são conexas entre si, adequando-se à primeira regra do critério de conectividade explicada na Seção 4.3. A segmentação 4 é conexa apenas com a 2, então não adequa-se à primeira regra do critério de conectividade. Porém, a segmentação 2 já adequou-se à primeira regra, além de pertencer a uma imagem adjacente à imagem da 4. Dessa forma, a segmentação 4 adequa-se à segunda regra. As demais segmentações são consideradas falsos positivos e eliminadas, por não se enquadrarem em nenhum caso como os anteriores.

Analisando quantitativamente os resultados da eliminação de falsos positivos em comparação com as segmentações iniciais das lesões, percebe-se que ainda é suave a eliminação de segmentações incorretas. Propõe-se então uma análise dos erros e acertos

Figura 32 – Estudo de Caso 2.



Fonte: Elaborado pelo autor.

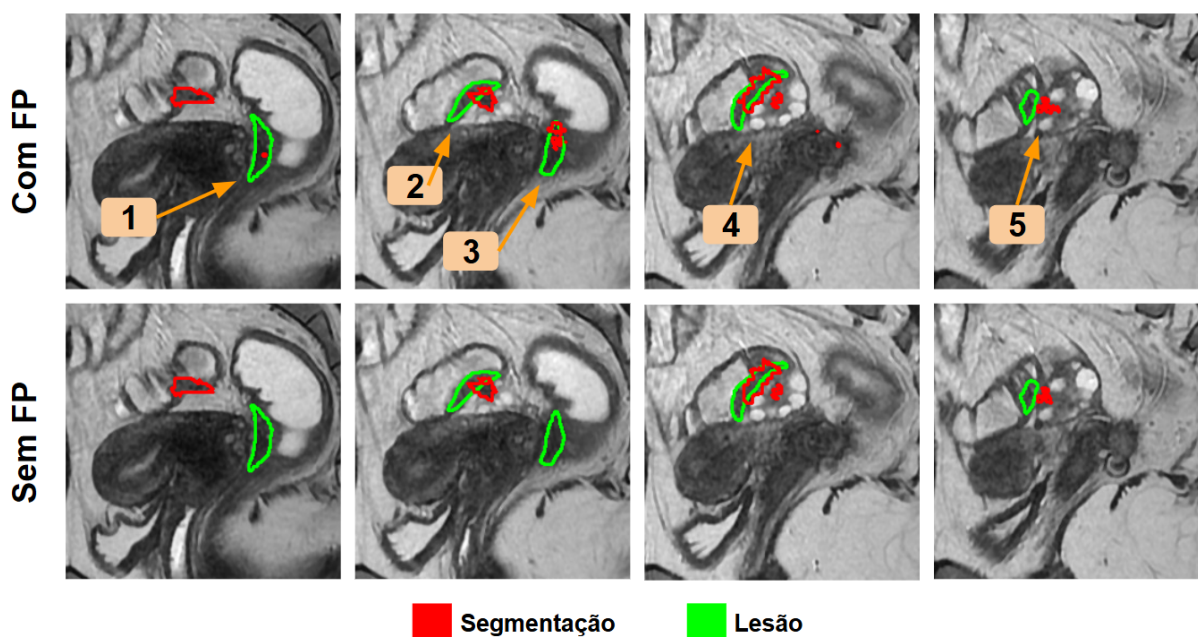
por imagem, tomando como abordagem verificar se as áreas segmentadas pelo modelo têm alguma interseção com as áreas de lesão. Essa análise demonstra a quantidade de segmentações desconexas foram removidas, para que seja possível ter uma primeira noção do impacto qualitativo do método. Os resultados dessa análise podem ser encontrados na Tabela 5. Nessa tabela, as lesões que possuem interseção com as segmentações e as que não possuem são apresentadas nas colunas “Lesões Atingidas” e “Lesões Erradas”, respectivamente. Além disso, na coluna “Falsos Positivos” é apresentado o número de segmentações sem nenhuma interseção com qualquer lesão, ou seja, falsos positivos. Essa análise oferece uma perspectiva adicional do efeito da etapa de eliminação de falsos positivos.

Tabela 5 – Resultados de erro e acerto por imagem.

FP Eliminados	Total de Lesões	Lesões Atingidas	Lesões Erradas	Falsos Positivos
Não	85	72	13	120
Sim	85	70	15	38

É notável a pequena piora na sensibilidade, representada pela diminuição da quantidade de lesões atingidas. Apesar disso, em grande maioria as segmentações corretas foram preservadas. As duas lesões perdidas pertencem a apenas uma das pacientes, analisado no Estudo de Caso 3, apresentado na Figura 33. Nele, as lesões reais estão numeradas e indicadas pelas setas em laranja. A linha “Com FP” mostra as segmentações antes da eliminação de falsos positivos e a linha “Sem FP” exibe-as após esta etapa. É possível observar que foram perdidas as segmentações das lesões 1 e 3, devido à eliminação de falsos positivos. As segmentações geradas pelo modelo para essas lesões são muito pequenas, então não foram conexas nem entre si nem com as das outras imagens. Entretanto, outras lesões da paciente foram segmentadas com maior sucesso.

Figura 33 – Estudo de Caso 3.



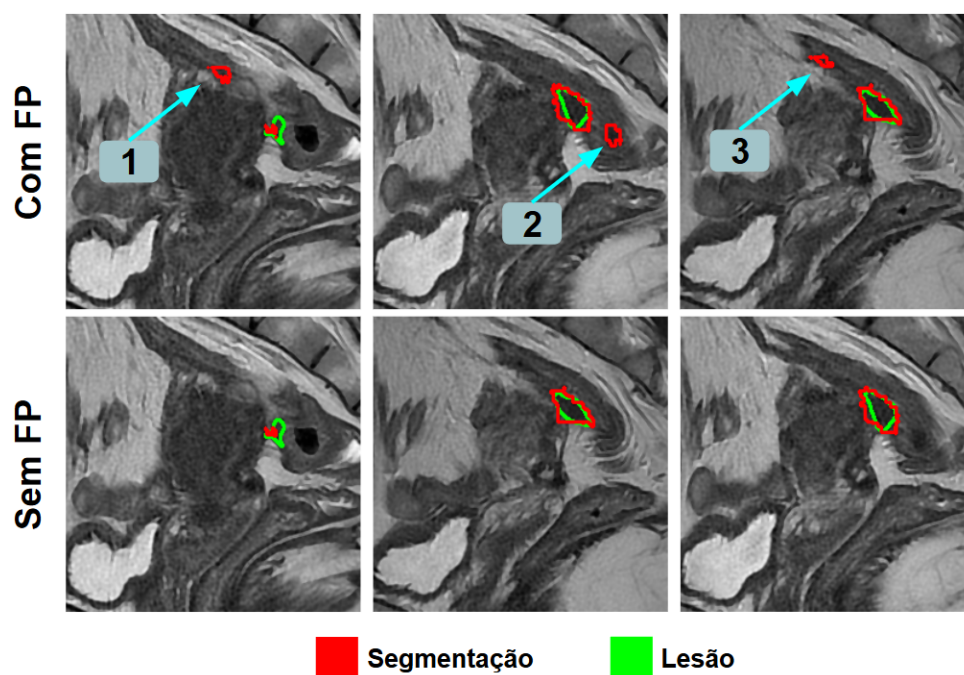
Fonte: Elaborado pelo autor.

Seguindo a análise qualitativa dos resultados, o Estudo de Caso 4, apresentado na Figura 34, apresenta um caso de grande sucesso do método. O paciente possui lesão em três de suas imagens, em uma área do exame onde o útero e o reto e sigmoide já são menos visíveis. As segmentações 1, 2 e 3, indicadas pelas setas azuis, são incorretas. Pelas imagens percebe-se a sutileza da diferença entre regiões de lesão e outras regiões escuras, o que causa os erros do modelo. Por pouco as segmentações 1 e 3 não foram mantidas, porém a Swin-Unet as segmentou de forma desconexa, então a etapa de eliminação de falsos positivos foi bem-sucedida em eliminá-las. As lesões reais foram, em sua maioria, bem marcadas.

Já no Estudo de Caso 5, apresentado na Figura 35, é vista uma situação em que a remoção de falsos positivos falhou em eliminar segmentações incorretas geradas pela Swin-Unet. Essas segmentações estão apontadas com os números 1 e 2 na imagem, pelas setas azuis. Essa paciente possui três imagens com lesão em seu exame. É possível perceber que a rede segmentou uma região escura, de textura semelhante a lesões de endometriose, entre o útero e o reto e sigmoide, região comum para lesões. Como a mesma região foi segmentada em duas das três imagens, adequou-se à primeira regra do critério de conectividade, então foi mantida. Pode ser que a região marcada seja uma lesão muito pequena, ignorada pelos especialistas. No fim, com apenas três imagens não há informações suficientes para validar a conectividade dessas segmentações erradas, pelo método proposto. Ainda assim, a lesão real foi segmentada com grande sucesso pelo modelo.

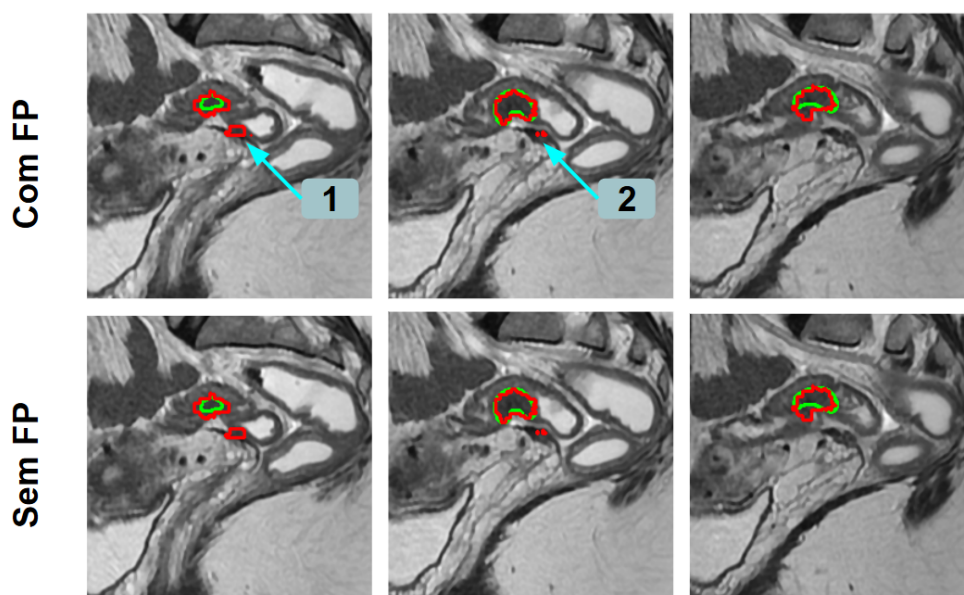
Concluindo a discussão dos resultados, viu-se que, na etapa de eliminação de

Figura 34 – Estudo de Caso 4.



Fonte: Elaborado pelo autor.

Figura 35 – Estudo de Caso 5.



Fonte: Elaborado pelo autor.

falsos positivos, muitas segmentações incorretas foram removidas. Isso deu-se ao custo de perder a segmentação de duas lesões de uma paciente que teve outras lesões corretamente segmentadas. Assim, esse método de redução de falsos positivos mostra-se promissor como pós-processamento das segmentações do modelo. O resultado final do método proposto oferece mais precisão da área que deve ser avaliada pelo profissional em busca das lesões.

Porém, se a etapa de segmentação das lesões gerar falhas nas segmentações, ou gerar segmentações muito pequenas em comparação às lesões reais, estas poderão ser perdidas pela eliminação de falsos positivos. Assim como se segmentações incorretas forem repetidas com um mínimo de conectividade ao longo das imagens, principalmente em pacientes com poucas imagens de lesão, estas não sofrerão influência da eliminação de falsos positivos.

A Tabela 6 compara o resultado por imagem do método proposto com o obtido pelo método de [Figueredo et al. \(2024\)](#), apresentado na Seção 2.1. Este é o método da literatura apresentada que pode ser mais adequadamente comparado com o proposto. É possível ver na tabela que os autores obtiveram resultados superiores ao método proposto. Porém, é um método mais denso, composto de mais etapas de processamento do que o método apresentado neste trabalho. Assim, observa-se que o método proposto alcançou valores não muito inferiores, mesmo sendo menos complexo. O método proposto utiliza apenas uma etapa baseada em aprendizado profundo, complementado por etapas de processamento de imagens, tornando-o mais leve. Além disso o método proposto abre novas possibilidades de abordagem do problema, apresentando a análise utilizada para o desenvolvimento da etapa de remoção de falsos positivos (Seção 4.3).

Tabela 6 – Comparação quantitativa do método proposto com o método de [Figueredo et al. \(2024\)](#).

Método	Precisão	Sensibilidade	Dice	Jaccard
Figueredo et al. (2024)	66,1%	72,5%	65,4%	36,2%
Proposto	54,1%	61,7%	48,7%	37,2%

6 Conclusão

A endometriose afeta dezenas de milhões de mulheres de várias faixas etárias ao redor do mundo. Ela afeta a qualidade de vida das pacientes e pode apresentar muitos sintomas diferentes, além de haver a possibilidade de evoluir para problemas mais sérios. Dessa forma, o diagnóstico eficaz da doença é essencial para iniciar o tratamento. O padrão-ouro atual para diagnóstico é a laparoscopia, um método invasivo. Através da RM, especialistas podem avaliar a presença da doença de forma não invasiva, porém a um alto custo de tempo. Além disso, as lesões apresentam diversos formatos, tamanhos e regiões de ocorrência, tornando a tarefa extremamente desafiadora até para radiologistas experientes. Neste contexto, métodos computacionais podem auxiliar os especialistas no diagnóstico, apresentando uma indicação da possível ocorrência das lesões.

Uma das regiões de mais comum incidência da endometriose é o reto e sigmoide. Neste contexto, este trabalho aplicou a Swin-Unet, uma rede de aprendizado profundo baseada em *transformer*, para a segmentação da endometriose profunda no reto e sigmoide em uma base de imagens de RM que apresentam a doença nessa região. A segmentação da doença por aprendizado profundo é pouco explorada na literatura, principalmente com RM. Dessa forma, foi analisado o desempenho da rede em comparação a outras arquiteturas, com a Swin-Unet mostrando-se mais eficaz. Ainda assim a rede cometeu vários erros, tanto ao falhar em segmentar áreas de lesão quanto em segmentar outras regiões muito semelhantes às lesões. Esses erros ocorrem devido às características de tamanho, textura, cor e posição das lesões, que podem ser confundidas com outras regiões dos exames. Uma segmentação eficaz da doença nesse tipo de imagem é um grande desafio.

Para obter uma segmentação mais eficaz, foi proposta uma etapa de refinamento das segmentações, utilizando a característica de conectividade das lesões entre as imagens do exame. Essa conectividade foi por várias vezes bem-sucedida em validar corretamente as segmentações obtidas, aumentando a precisão do método, a baixo custo de sensibilidade. Assim foram segmentadas com sucesso lesões de 17 das 18 pacientes utilizadas no teste do método. Ao menos uma lesão em cada paciente foi segmentada com grande eficácia, o que já é promissor para um sistema de auxílio ao diagnóstico, no contexto da endometriose em imagens de RM. Isto se dá principalmente pelo fato do método comportar-se de forma semelhante ao protocolo já utilizado pelo especialista, podendo então tornar mais eficiente a marcação das lesões. Dessa forma, ao segmentar com sucesso uma lesão de cada paciente e eliminar a maioria das regiões de erro, o método proposto mostra-se eficaz em indicar a área de lesão para o especialista.

O método desenvolvido alcançou resultados promissores, que indicam o bom

desempenho da arquitetura baseada em *transformer*, pré-treinada, em tarefas mais difíceis para uma CNN convencional. A etapa de eliminação de falsos positivos promoveu uma pequena melhora deste resultado. O método completo foi capaz de segmentar com grande sucesso ao menos uma lesão em 17 dos 18 pacientes. Tendo em vista a dificuldade da segmentação manual das lesões, esses resultados indicam o potencial para um sistema de auxílio ao diagnóstico. Isso é promissor principalmente ao analisar quão desafiadora é a tarefa.

6.1 Limitações e Trabalhos Futuros

Esse problema é pouco explorado na literatura, então algumas tarefas podem ser desempenhadas para contribuir com o método proposto e complementar o estudo realizado. A principal limitação do método é sua incapacidade de relacionar informações tridimensionais dos exames de RM durante o aprendizado da rede. Abordagens nesse contexto foram experimentadas, porém não obtiveram bons resultados. O método de remoção de falsos positivos sofre diretamente com essa falta de uma análise tridimensional no momento da segmentação. Isto se dá pelo fato do critério de conectividade não ser capaz de preservar segmentações muito pequenas, realizadas em lesões grandes, que acabam sendo desconexas. Assim, a seguir serão apresentadas algumas sugestões de trabalhos futuros.

Primeiro, para abordar a principal limitação do trabalho, propõe-se utilizar informações tridimensionais no momento do treinamento do modelo. Acredita-se que, extraindo características das imagens adjacentes à analisada, será possível obter uma segmentação mais conexa. Redes 3D podem ter baixo desempenho devido ao tamanho da base, então é interessante analisar abordagens 2.5D, onde apenas algumas imagens adjacentes à analisada são concatenadas à mesma antes de inserí-la na rede.

Outra proposta é utilizar aprendizado profundo para determinar a validade das segmentações contestadas, para a etapa de remoção de falsos positivos. Ao invés de utilizar as regras fixas propostas do critério de conectividade, regras aprendidas por uma rede profunda podem validar de forma mais eficaz os erros e acertos da segmentação inicial.

Além disso, propõe-se seguir avaliando etapas complementares ao método, para refinar as segmentações com outros modelos de aprendizado profundo, mesmo que sejam da mesma arquitetura. Aponta-se o desenvolvimento de uma etapa de redução da ROI, após a remoção de falsos positivos, a partir da posição das regiões segmentadas, para treinar um novo modelo na ROI reduzida e aumentar a qualidade da segmentação. A redução da ROI gera um espaço de busca reduzido, onde uma rede pode operar de forma mais precisa.

Também é proposto avaliar alterações arquiteturais da Swin-Unet, como a utilização

de blocos Kolmogorov-Arnold como substitutos às MLPs (LIU et al., 2025), além da combinação com blocos de CNN, aproveitando as qualidades dos dois tipos de arquitetura. Aproveitar as diferentes características de blocos de aprendizado profundo é promissor em melhorar o resultado. Porém isso exige muita experimentação para encontrar a combinação mais eficaz.

Outra sugestão é a avaliação de técnicas de melhoramento de imagem que sejam capazes de realçar áreas de lesão. Até o presente momento toda técnica experimentada neste trabalho acabou por piorar o desempenho da Swin-Unet. Ainda assim, a área de processamento de imagens é extensa, com muitas outras combinações de métodos e hiperparâmetros sendo possíveis. A execução dessa proposta pode definir uma etapa do método que diminua ruídos e aumenta o contraste das imagens, aumentando a visibilidade das lesões.

6.2 Produção Científica

Este trabalho gerou o artigo Segmentação Automática de Endometriose Profunda em Imagens de Ressonância Magnética Baseada em Swin-Unet, aceito e apresentado no 24º Simpósio Brasileiro de Computação Aplicada à Saúde, SBCAS 2024, com Qualis A4.

Referências

- AGARWAL, N.; SUBRAMANIAN, A. Endometriosis-morphology, clinical presentations and molecular pathology. *Journal of laboratory physicians*, Thieme Medical and Scientific Publishers Private Ltd., v. 2, n. 01, p. 001–009, 2010. Citado 2 vezes nas páginas 17 e 29.
- ANDREUCCI, C. G. L.; SOUZA, L. M.; LOMBARDI, L.; ANDREUCCI, H. T. C. Principais achados radiologicos de endometriose em ressonancia magnetica. *Atas de Ciências da Saúde (ISSN 2448-3753)*, v. 2, n. 3, 2014. Citado 3 vezes nas páginas 17, 28 e 29.
- BAZOT, M.; DARAĬ, E. Diagnosis of deep endometriosis: clinical examination, ultrasonography, magnetic resonance imaging, and other techniques. *Fertility and sterility*, Elsevier, v. 108, n. 6, p. 886–894, 2017. Citado 2 vezes nas páginas 21 e 27.
- BAZOT, M.; DARAI, E.; HOURANI, R.; THOMASSIN, I.; CORTEZ, A.; UZAN, S.; BUY, J.-N. Deep pelvic endometriosis: Mr imaging for diagnosis and prediction of extension of disease. *Radiology*, Radiological Society of North America, v. 232, n. 2, p. 379–389, 2004. Citado 2 vezes nas páginas 22 e 27.
- BELLELIS, P.; GIACOMETTI, C. F. Utilização da cannabis medicinal no tratamento da endometriose. *BrJP*, SciELO Brasil, v. 6, p. 95–96, 2023. Citado na página 17.
- BENDIFALLAH, S.; PUCHAR, A.; SUISSE, S.; DELBOS, L.; POILBLANC, M.; DESCAMPS, P.; GOLFIER, F.; TOUBOUL, C.; DABI, Y.; DARAĬ, E. Machine learning algorithms as new screening approach for patients with endometriosis. *Scientific Reports*, Nature Publishing Group UK London, v. 12, n. 1, p. 639, 2022. Citado 2 vezes nas páginas 26 e 27.
- BERBEL, B. T.; PODGAEC, S.; ABRÃO, M. S. Análise da associação entre o quadro clínico referido pelas pacientes portadoras de endometriose e o local de acometimento da doença. *Revista de Medicina*, v. 87, n. 3, p. 195–200, 2008. Citado 2 vezes nas páginas 17 e 29.
- BIANCHINI, M.; MAGGINI, M.; SARTI, L.; SCARSELLI, F. Recursive neural networks for object detection. In: IEEE. *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*. [S.l.], 2004. v. 3, p. 1911–1915. Citado na página 35.
- BLOCH, F. Nuclear induction. *Physical review*, APS, v. 70, n. 7-8, p. 460, 1946. Citado na página 31.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001. Citado na página 26.
- 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 8 vezes nas páginas 18, 34, 38, 41, 46, 52, 54 e 59.

- CARAÇA, D. B.; PODGAEC, S.; BARACAT, E. C.; ABRÃO, M. S. Mecanismos fisiopatológicos da dor pélvica na endometriose profunda. *Diagn Tratamento*, v. 16, n. 2, p. 57–61, 2011. Citado na página 17.
- CHEN, J.; LU, Y.; YU, Q.; LUO, X.; ADELI, E.; WANG, Y.; LU, L.; YUILLE, A. L.; ZHOU, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. Citado na página 20.
- CHEN, L.-C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*. [S.l.: s.n.], 2018. p. 801–818. Citado na página 24.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794. Citado 2 vezes nas páginas 25 e 26.
- 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 4 vezes nas páginas 23, 38, 46 e 63.
- 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 2 vezes nas páginas 18 e 36.
- DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. Citado 5 vezes nas páginas 18, 41, 44, 45 e 60.
- FERNANDO, S.; SOH, P. Q.; COOPER, M.; EVANS, S.; REID, G.; TSALTAS, J.; ROMBAUTS, L. Reliability of visual diagnosis of endometriosis. *Journal of minimally invasive gynecology*, Elsevier, v. 20, n. 6, p. 783–789, 2013. Citado 2 vezes nas páginas 23 e 27.
- FERRERO, S.; BARRA, F.; STABILINI, C.; VELLONE, V. G.; MAGGIORE, U. L. R.; SCALA, C. Does bowel preparation improve the performance of rectal water contrast transvaginal ultrasonography in diagnosing rectosigmoid endometriosis? *Journal of Ultrasound in Medicine*, Wiley Online Library, v. 38, n. 4, p. 1017–1025, 2019. Citado 2 vezes nas páginas 25 e 27.
- 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. Citado 5 vezes nas páginas 20, 21, 27, 50 e 58.
- FIGUEREDO, W. K. R.; SILVA, A. C.; PAIVA, A. C. de; DINIZ, J. O. B.; BRANDÃO, A.; OLIVEIRA, M. A. P. Automatic segmentation of deep endometriosis in the

- rectosigmoid using deep learning. *Image and Vision Computing*, Elsevier, v. 151, p. 105261, 2024. Citado 4 vezes nas páginas 12, 20, 27 e 69.
- FILHO, O. M.; NETO, H. V. *Processamento digital de imagens*. [S.l.]: Brasport, 1999. Citado na página 32.
- FREUND, Y.; SCHAPIRE, R.; ABE, N. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, JAPANESE SOC ARTIFICIAL INTELL, v. 14, n. 771-780, p. 1612, 1999. Citado na página 26.
- GIUSTI, S.; FORASASSI, F.; BASTIANI, L.; CELA, V.; PLUCHINO, N.; FERRARI, V.; FRUZZETTI, E.; CAMELLA, D.; BARTOLOZZI, C. Anatomical localization of deep infiltrating endometriosis: 3d mri reconstructions. *Abdominal imaging*, Springer, v. 37, p. 1110–1121, 2012. Citado 3 vezes nas páginas 17, 22 e 27.
- GOLDSTEIN, A.; COHEN, S. Self-report symptom-based endometriosis prediction using machine learning. *Scientific Reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 5499, 2023. Citado 2 vezes nas páginas 25 e 27.
- GONCALVES, M. O.; NETO, J. S.; ANDRES, M. P.; SIUFI, D.; MATTOS, L. A. de; ABRAO, M. S. Systematic evaluation of endometriosis by transvaginal ultrasound can accurately replace diagnostic laparoscopy, mainly for deep and ovarian endometriosis. *Human Reproduction*, Oxford University Press, v. 36, n. 6, p. 1492–1500, 2021. Citado 2 vezes nas páginas 25 e 27.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento digital de imagens. 3. ed.* [S.l.]: São Paulo: Pearson Prentice Hall, 2010. Citado 5 vezes nas páginas 23, 24, 25, 32 e 33.
- HAGE, M. C. F. N. S.; IWASAKI, M. Magnetic resonance imaging–basics/imagem por ressonancia magnetica: principios basicos. *Ciência rural*, Universidade Federal de Santa Maria, v. 39, n. 4, p. 1287–1296, 2009. Citado 2 vezes nas páginas 31 e 32.
- HARKINS, J. M.; SAJJAD, H. Anatomy, abdomen and pelvis, sigmoid colon. In: *StatPearls [Internet]*. StatPearls Publishing, 2023. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK549824/>>. Citado na página 29.
- HAYKIN, S. Kalman filters. *Kalman filtering and neural networks*, Wiley Online Library, p. 1–21, 2001. Citado 2 vezes nas páginas 34 e 35.
- HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2961–2969. Citado na página 23.
- 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 (CVPR)*. [S.l.: s.n.], 2016. Citado 3 vezes nas páginas 22, 23 e 63.
- HENDRYCKS, D.; GIMPEL, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. Citado na página 45.
- HOWARD, A. G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. Citado na página 23.

- HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 4700–4708. Citado na página 63.
- ISLAM, S.; ELMEEKKI, H.; ELSEBAI, A.; BENTAHAR, J.; DRAWEL, N.; RJOUB, G.; PEDRYCZ, W. A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications*, Elsevier, v. 241, p. 122666, 2024. Citado na página 44.
- JUNIOR, A. C. C.; LIMA, C. M. A. d. O.; COUTINHO, E. P. D.; RIBEIRO, É. B.; AIDAR, M. N.; GASPARETTO, E. L. Ressonância magnética na endometriose pélvica profunda: ensaio iconográfico. *Radiologia Brasileira*, SciELO Brasil, v. 41, p. 129–134, 2008. Citado na página 29.
- JÚNIOR, E. A.; YAMASHITA, H. Aspectos básicos de tomografia computadorizada e ressonância magnética. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 23, p. 2–3, 2001. Citado na página 32.
- KIMORI, Y. Mathematical morphology-based approach to the enhancement of morphological features in medical images. *Journal of clinical bioinformatics*, Springer, v. 1, p. 1–10, 2011. Citado na página 18.
- KRÜGER, K.; BEHRENDT, K.; NIEDOBITEK-KREUTER, G.; KOLTERMANN, K.; EBERT, A. D. Location-dependent value of pelvic mri in the preoperative diagnosis of endometriosis. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, Elsevier, v. 169, n. 1, p. 93–98, 2013. Citado 2 vezes nas páginas 21 e 27.
- KUMAR, T.; BRENNAN, R.; MILEO, A.; BENDECHACHE, M. Image data augmentation approaches: A comprehensive survey and future directions. *IEEE Access*, IEEE, 2024. Citado na página 38.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado 2 vezes nas páginas 36 e 37.
- LEIBETSEDER, A.; KLETZ, S.; SCHOEFFMANN, K.; KECKSTEIN, S.; KECKSTEIN, J. GLEND: gynecologic laparoscopy endometriosis dataset. In: *MultiMedia Modeling - 26th International Conference, MMM 2020, Daejeon, South Korea, January 5-8, 2020, Proceedings, Part II*. [S.l.]: Springer, 2020. (Lecture Notes in Computer Science, v. 11962), p. 439–450. Citado na página 23.
- LEIBETSEDER, A.; SCHOEFFMANN, K.; KECKSTEIN, J.; KECKSTEIN, S. Endometriosis detection and localization in laparoscopic gynecology. *Multimedia Tools and Applications*, Springer, v. 81, n. 5, p. 6191–6215, 2022. Citado na página 18.
- LEIBETSEDER, A.; SCHOEFFMANN, K.; KECKSTEIN, J.; KECKSTEIN, S. Endometriosis detection and localization in laparoscopic gynecology. *Multim. Tools Appl.*, v. 81, n. 5, p. 6191–6215, 2022. Disponível em: <<https://doi.org/10.1007/s11042-021-11730-1>>. Citado 3 vezes nas páginas 22, 23 e 27.
- LEVINE, B. *Endometriosis and painful sex: How to cope*. 2022. Acesso em: 12 de Dezembro de 2024. Disponível em: <<https://www.everydayhealth.com/endometriosis/severe-sex-pain-treatments/>>. Citado na página 28.

- LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, IEEE, v. 33, n. 12, p. 6999–7019, 2021. Citado 2 vezes nas páginas 38 e 39.
- LIU, Y.; ZHOU, Q.; PENG, B.; JIANG, J.; FANG, L.; WENG, W.; WANG, W.; WANG, S.; ZHU, X. Automatic measurement of endometrial thickness from transvaginal ultrasound images. *Frontiers in bioengineering and biotechnology*, Frontiers Media SA, v. 10, p. 853845, 2022. Citado 2 vezes nas páginas 24 e 27.
- 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 5 vezes nas páginas 18, 45, 46, 52 e 54.
- LIU, Z.; WANG, Y.; VAIDYA, S.; RUEHLE, F.; HALVERSON, J.; SOLJAČIĆ, M.; HOU, T. Y.; TEGMARK, M. KAN: Kolmogorov-Arnold Networks. 2025. Disponível em: <<https://arxiv.org/abs/2404.19756>>. Citado na página 72.
- LUNDERVOLD, A. S.; LUNDERVOLD, A. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, Elsevier, v. 29, n. 2, p. 102–127, 2019. Citado na página 18.
- MANGANARO, L.; FIERRO, F.; TOMEI, A.; IRIMIA, D.; LODISE, P.; SERGI, M.; VINCI, V.; SOLLAZZO, P.; PORPORA, M.; DELFINI, R. et al. Feasibility of 3.0 t pelvic mr imaging in the evaluation of endometriosis. *European journal of radiology*, Elsevier, v. 81, n. 6, p. 1381–1387, 2012. Citado 2 vezes nas páginas 17 e 29.
- MARCHIORI, E.; ZANETTI, G.; RODRIGUES, R. S.; SOUZA, L. S.; JR, A. S. S.; FRANCISCO, F. A. F.; HOCHHEGGER, B. Pleural endometriosis: findings on magnetic resonance imaging. *Jornal Brasileiro de Pneumologia*, SciELO Brasil, v. 38, p. 797–802, 2012. Citado 2 vezes nas páginas 17 e 28.
- METTLER, L.; SCHOLLMEYER, T.; LEHMANN-WILLENBROCK, E.; SCHÜPPLER, U.; SCHMUTZLER, A.; SHUKLA, D.; ZAVALA, A.; LEWIN, A. Accuracy of laparoscopic diagnosis of endometriosis. *JSLs: Journal of the Society of Laparoendoscopic Surgeons*, Society of Laparoscopic & Robotic Surgeons, v. 7, n. 1, p. 15, 2003. Citado 2 vezes nas páginas 24 e 27.
- MIRANDA, A. R.; GÓMEZ, M. B.; ARRONES, S.; SALVADOR, Z.; GOOSMAN, C. A. *What Is Endometriosis? - Causes, Symptoms and Treatment*. 2022. Acesso em: 12 de Dezembro de 2024. Disponível em: <<https://www.invitro.com/en/endometriosis/retrograde-menstruation/>>. Citado na página 29.
- MOURA, A. P. C.; RIBEIRO, H. S. A. A.; BERNARDO, W. M.; SIMÕES, R.; TORRES, U. S.; D'IPPOLITO, G.; BAZOT, M.; RIBEIRO, P. A. A. G. Accuracy of transvaginal sonography versus magnetic resonance imaging in the diagnosis of rectosigmoid endometriosis: Systematic review and meta-analysis. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 14, n. 4, p. e0214842, 2019. Citado 3 vezes nas páginas 17, 21 e 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 2 vezes nas páginas 47 e 48.
- NOGUEIRA, A. C. R.; SANTIAGO, M. T.; BAHIA, C. P.; SOARES, H. H. P. Tratamento da endometriose pélvica: uma revisão sistemática. 2018. Citado 2 vezes nas páginas 17 e 29.
- OTSU, N. et al. A threshold selection method from gray-level histograms. *Automatica*, v. 11, n. 285-296, p. 23-27, 1975. Citado na página 33.
- PACHECO, C. A. R.; PEREIRA, N. S. Deep learning conceitos e utilização nas diversas áreas do conhecimento. *Revista Ada Lovelace*, 2018. Citado 3 vezes nas páginas 34, 35 e 36.
- PODDA, A. S.; BALIA, R.; BARRA, S.; CARTA, S.; NERI, M.; GUERRIERO, S.; PIANO, L. Multi-scale deep learning ensemble for segmentation of endometriotic lesions. *Neural Computing and Applications*, Springer, p. 1-14, 2024. Citado 2 vezes nas páginas 24 e 27.
- PONTES, I. F.; CLAUDINO, E. L. Dor pélvica e achados indiretos da endometriose na ecografia pélvica: Uma correlação estatística. *Research, Society and Development*, v. 10, n. 8, p. e49210817709-e49210817709, 2021. Citado 2 vezes nas páginas 17 e 29.
- PU, Q.; XI, Z.; YIN, S.; ZHAO, Z.; ZHAO, L. Advantages of transformer and its application for medical image segmentation: a survey. *BioMedical Engineering OnLine*, Springer, v. 23, n. 1, p. 14, 2024. Citado na página 45.
- PURCELL, E. M.; TORREY, H. C.; POUND, R. V. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, APS, v. 69, n. 1-2, p. 37, 1946. Citado na página 31.
- REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 39, n. 6, p. 1137-1149, 2016. Citado na página 23.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 234-241. ISBN 978-3-319-24574-4. Citado 4 vezes nas páginas 22, 39, 40 e 52.
- RUAN, J.; LI, J.; XIANG, S. Vm-unet: Vision mamba unet for medical image segmentation. *arXiv preprint arXiv:2402.02491*, 2024. Citado na página 64.
- SAADATI, S.; AMIRMAZLAGHANI, M. Revolutionizing endometriosis treatment: automated surgical operation through artificial intelligence and robotic vision. *Journal of Robotic Surgery*, Springer, v. 18, n. 1, p. 383, 2024. Citado 2 vezes nas páginas 22 e 27.
- SACCARDI, C.; COSMI, E.; BORGHERO, A.; TREGNAGHI, A.; DESSOLE, S.; LITTA, P. Comparison between transvaginal sonography, saline contrast sonovaginography and magnetic resonance imaging in the diagnosis of posterior deep infiltrating endometriosis. *Ultrasound in obstetrics & gynecology*, Wiley Online Library, v. 40, n. 4, p. 464-469, 2012. Citado 2 vezes nas páginas 22 e 27.

- SARMA, D.; IYENGAR, P.; MAROTTA, T. R.; TERBRUGGE, K. G.; GENTILI, F.; HALLIDAY, W. Cerebellar endometriosis. *American journal of roentgenology*, Am Roentgen Ray Soc, v. 182, n. 6, p. 1543–1546, 2004. Citado 2 vezes nas páginas 17 e 28.
- SCARDAPANE, A.; LORUSSO, F.; BETTOCCHI, S.; MOSCHETTA, M.; FIUME, M.; VIMERCATI, A.; PEPE, M.; ANGELELLI, G.; IANORA, A. A. S. Deep pelvic endometriosis: accuracy of pelvic mri completed by mr colonography. *La radiologia medica*, Springer Milan, p. 1–16, 2013. Citado 2 vezes nas páginas 21 e 27.
- SCHNEIDER, C.; OEHMKE, F.; TINNEBERG, H.-R.; KROMBACH, G. Mri technique for the preoperative evaluation of deep infiltrating endometriosis: current status and protocol recommendation. *Clinical Radiology*, Elsevier, v. 71, n. 3, p. 179–194, 2016. Citado 2 vezes nas páginas 17 e 18.
- 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 37.
- SILVA, A. *Algoritmos para diagnóstico assistido de nódulos pulmonares solitários em imagens de tomografia computadorizada*. 2004. 140f. Tese (Doutorado) — Tese (Doutorado)-Departamento de Informática, Pontifícia Universidade ... , 2004. Citado na página 36.
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. Disponível em: <<https://arxiv.org/abs/1409.1556>>. Citado 3 vezes nas páginas 20, 23 e 63.
- STRATTON, P.; WINKEL, C.; PREMKUMAR, A.; CHOW, C.; WILSON, J.; HEARNS-STOKES, R.; HEO, S.; MERINO, M.; NIEMAN, L. K. Diagnostic accuracy of laparoscopy, magnetic resonance imaging, and histopathologic examination for the detection of endometriosis. *Fertility and sterility*, Elsevier, v. 79, n. 5, p. 1078–1085, 2003. Citado 2 vezes nas páginas 22 e 27.
- SU, R.; ZHANG, D.; LIU, J.; CHENG, C. Msu-net: Multi-scale u-net for 2d medical image segmentation. *Frontiers in Genetics*, v. 12, 2021. ISSN 1664-8021. Citado na página 24.
- SZABÓ, G.; HUDELIST, G.; MADÁR, I.; DOBÓ, N.; FINTHA, A.; LIPTÁK, L.; KALOVICS, E.; FANCISOVITS, V.; BOKOR, A. et al. Diagnostic accuracy of the idea protocol for non invasive diagnosis of rectosigmoid de—a prospective cohort study. *Ultraschall in der Medizin-European Journal of Ultrasound*, Georg Thieme Verlag KG, v. 35, 2023. Citado 2 vezes nas páginas 24 e 27.
- SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 2818–2826. Citado na página 23.
- TAN, M.; LE, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: CHAUDHURI, K.; SALAKHUTDINOV, R. (Ed.). *Proceedings of the 36th International Conference on Machine Learning*. [S.l.]: PMLR, 2019. (Proceedings of Machine Learning Research, v. 97), p. 6105–6114. Citado 2 vezes nas páginas 23 e 63.

TANG, H.; HUANG, G.; CHENG, L.; YUAN, X.; TAO, Q.; CHEN, X.; ZHONG, G.; YANG, X. Rm-unet: Unet-like mamba with rotational ssm module for medical image segmentation. *Signal, Image and Video Processing*, Springer, v. 18, n. 11, p. 8427–8443, 2024. Citado na página 64.

TAYE, M. M. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers*, MDPI, v. 12, n. 5, p. 91, 2023. Citado 3 vezes nas páginas 34, 35 e 36.

TAYLOR, H. S. Endometriosis: a complex systemic disease with multiple manifestations. *Fertility and sterility*, Elsevier, v. 112, n. 2, p. 235–236, 2019. Citado 2 vezes nas páginas 17 e 29.

TORE, U.; ABILGAZYM, A.; BARCO, A. Asunsolo-del; TERZIC, M.; YEMENKHAN, Y.; ZOLLANVARI, A.; SARRIA-SANTAMERA, A. Diagnosis of endometriosis based on comorbidities: a machine learning approach. *Biomedicines*, MDPI, v. 11, n. 11, p. 3015, 2023. Citado 2 vezes nas páginas 25 e 27.

TRAN, D.; BOURDEV, L.; FERGUS, R.; TORRESANI, L.; PALURI, M. Learning spatiotemporal features with 3d convolutional networks. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2015. p. 4489–4497. Citado na página 21.

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*. [S.l.]: Curran Associates, Inc., 2017. v. 30. Citado 3 vezes nas páginas 41, 42 e 43.

VISALAXI, S.; MUTHU, T. Automated prediction of endometriosis using deep learning. *International Journal of Nonlinear Analysis and Applications*, Semnan University, v. 12, n. 2, p. 2403–2416, 2021. ISSN 2008-6822. Citado 2 vezes nas páginas 23 e 27.

VISALAXI, S.; PUNNOOSE, D.; MUTHU, T. Lesion extraction of endometriotic images using open computer vision. In: *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. [S.l.: s.n.], 2021. p. 747–751. Citado 2 vezes nas páginas 23 e 27.

VISALAXI, S.; SUDALAIMUTHU, T. Automated segmentation of endometriosis using transfer learning technique. *F1000Research*, F1000 Research Limited London, UK, v. 11, p. 360, 2022. Citado 2 vezes nas páginas 23 e 27.

WANG, X.; BAO, N.; XIN, X.; TAN, J.; LI, H.; ZHOU, S.; LIU, H. Automatic evaluation of endometrial receptivity in three-dimensional transvaginal ultrasound images based on 3d u-net segmentation. *Quantitative Imaging in Medicine and Surgery*, v. 12, n. 8, 2022. ISSN 2223-4306. Citado 2 vezes nas páginas 24 e 27.

WONG, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern recognition*, Elsevier, v. 48, n. 9, p. 2839–2846, 2015. Citado na página 59.

- YANG, M.; LIU, M.; CHEN, Y.; HE, S.; LIN, Y. Diagnostic efficacy of ultrasound combined with magnetic resonance imaging in diagnosis of deep pelvic endometriosis under deep learning. *The Journal of Supercomputing*, Springer, v. 77, n. 7, p. 7598–7619, 2021. Citado 2 vezes nas páginas 21 e 27.
- YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks? *Advances in neural information processing systems*, v. 27, 2014. Citado na página 38.
- YUSHKEVICH, P. A.; PIVEN, J.; HAZLETT, H. C.; SMITH, R. G.; HO, S.; GEE, J. C.; GERIG, G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, v. 31, n. 3, p. 1116–1128, 2006. Citado na página 22.
- ZHANG, M.; YU, Y.; JIN, S.; GU, L.; LING, T.; TAO, X. Vm-unet-v2: rethinking vision mamba unet for medical image segmentation. In: SPRINGER. *International Symposium on Bioinformatics Research and Applications*. [S.l.], 2024. p. 335–346. Citado na página 64.
- ZHAO, X.; WANG, L.; ZHANG, Y.; HAN, X.; DEVECI, M.; PARMAR, M. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, Springer, v. 57, n. 4, p. 99, 2024. Citado 3 vezes nas páginas 38, 39 e 41.
- ZHOU, B.; LAPEDRIZA, A.; XIAO, J.; TORRALBA, A.; OLIVA, A. Learning deep features for scene recognition using places database. In: GHAHRAMANI, Z.; WELLING, M.; CORTES, C.; LAWRENCE, N.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. [S.l.]: Curran Associates, Inc., 2014. v. 27. Citado na página 37.
- ZULUAGA-GOMEZ, J.; MASRY, Z. A.; BENAGGOUNE, K.; MERAGHNI, S.; ZERHOUNI, N. A cnn-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, p. 1–15, 2020. Citado na página 48.
- ZWAS, F. R.; LYON, D. T. Endometriosis: an important condition in clinical gastroenterology. *Digestive diseases and sciences*, Springer, v. 36, p. 353–364, 1991. Citado na página 29.