



UNIVERSIDADE FEDERAL DO MARANHÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
Doutorado em Ciência da Computação Associação
UFMA/UFPI

Wesley Kelson Ribeiro Figueredo

**Abordagem Computacional Baseada em *Deep Learning*
para a Classificação e Segmentação de Endometriose
Profunda no Retossigmoide através de Imagens de
Ressonância Magnética**

Orientador: Prof. Dr. Aristófanés Corrêa Silva
Co-orientador: Prof. Dr. Marco Aurelio Pinho de Oliveira

São Luís - MA
Janeiro, 2025

Wesley Kelson Ribeiro Figueredo

**Abordagem Computacional Baseada em *Deep Learning*
para a Classificação e Segmentação de Endometriose
Profunda no Retossigmoide através de Imagens de
Ressonância Magnética**

TESE DE DOUTORADO

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

Orientador: Prof. Dr. Aristófanés Corrêa Silva
Co-orientador: Prof. Dr. Marco Aurelio Pinho de Oliveira

São Luís - MA
Janeiro, 2025

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

Figueredo, Wesley Kelson Ribeiro.

Abordagem Computacional Baseada em Deep Learning para a Classificação e Segmentação de Endometriose Profunda no Retossigmoide através de Imagens de Ressonância Magnética / Wesley Kelson Ribeiro Figueredo. - 2025.

96 f.

Coorientador(a) 1: Marco Aurelio Pinho de Oliveira.

Orientador(a): Aristófanês Corrêa Silva.

Tese (Doutorado) - Programa de Pós-graduação Doutorado em Ciência da Computação - Associação UFMA/UFPI, Universidade Federal do Maranhão, São Luís - Ma, 2025.

1. Endometriose. 2. Aprendizado Profundo. 3. Classificação. 4. Segmentação. 5. Ressonância Magnética. I. Oliveira, Marco Aurelio Pinho de. II. Silva, Aristófanês Corrêa. III. Título.

Wesley Kelson Ribeiro Figueredo

Abordagem Computacional Baseada em *Deep Learning* para a Classificação e Segmentação de Endometriose Profunda no Retossigmoide através de Imagens de Ressonância Magnética

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

Prof. Dr. Aristófanês Corrêa Silva

Orientador

Universidade Federal do Maranhão

Prof. Dr. Marco Aurelio Pinho de Oliveira

Co-orientador

Universidade Federal do Maranhão

Prof. Dr. Anselmo Cardoso de Paiva

Examinador Interno

Universidade Federal do Maranhão

Prof. Dr. Antonio Oseas de Carvalho Filho

Examinador Interno

Universidade Federal do Piauí

Prof. Dr. António Manuel Trigueiros da Silva Cunha

Examinador Externo

A Universidade de Trás-os-Montes e Alto Douro

Prof. Dr. Aura Conci

Examinadora externa

Universidade Federal Fluminense

Prof. Dr. João Dallyson Sousa de Almeida

Examinador Interno

Universidade Federal do Maranhão

Certificamos que esta é a versão original e final da Tese de Doutorado que foi julgada adequa

Doutor em Ciência da Computação.



Documento assinado digitalmente

ARISTOFANES CORREA SILVA

Data: 13/02/2025 11:08:20-0300

Verifique em <https://validar.iti.gov.br>

Prof. Dr. Aristófanês Corrêa Silva

Orientador

Documento assinado digitalmente



ANSELMO CARDOSO DE PAIVA

Data: 17/02/2025 10:41:26-0300

Verifique em <https://validar.iti.gov.br>

Prof. Dr. Anselmo Cardoso de Paiva

Coordenador

São Luís - MA, 20 de Janeiro de 2025

À minha família e aos verdadeiros amigos

Agradecimentos

A Deus, primeiramente. Até aqui, sempre me ajudou o Senhor.

À minha família. Em especial à minha mãe, pelo incentivo e suporte na minha jornada acadêmica.

À minha namorada, Mayara, pelo suporte, apoio e companheirismo em todos os momentos.

Ao meu orientador, Aristófares Corrêa Silva, pela orientação, apoio, paciência e as incontáveis oportunidades oferecidas.

Aos especialistas Dra. Alice Brandão, Dr. Marco Oliveira e à Clínica Fonte de Imagem pelo suporte e disponibilização das imagens.

Aos amigos que fiz pelo caminho, pela amizade, companheirismo, ensinamentos, conselhos, momentos de descontração e todo apoio dado ao longo desta jornada.

Ao Núcleo de Computação Aplicada (NCA), em especial aos professores Anselmo, Simara, Geraldo e Dallyson, pelo suporte durante todo esse período.

Ao Programa de Pós-Graduação do Doutorado em Ciência da Computação em Associação UFMA e UFPI (DCCMAPI) pelo apoio técnico.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Maranhão (FAPEMA) pelo apoio financeiro.

Agradeço a todos que de alguma forma contribuíram para minha formação acadêmica.

*“A vitalidade é demonstrada não apenas pela persistência,
mas pela capacidade de começar de novo.”*

F. Scott Fitzgerald.

Resumo

A endometriose é uma doença que afeta principalmente a região pélvica, impactando significativamente a qualidade de vida das mulheres, especialmente em idade fértil. Este trabalho propõe um método automatizado para a classificação de pacientes com endometriose, localização e segmentação das lesões em imagens de ressonância magnética do retossigmoide, com o objetivo de auxiliar no diagnóstico, reduzir a necessidade de métodos invasivos e mapear a extensão das lesões. O método combina técnicas de processamento de imagens e aprendizado profundo, com três objetivos principais: classificação de pacientes, localização e segmentação de lesões. Para a classificação, foi alcançado um F1-score de 94,74%, utilizando um *ensemble* de redes neurais convolucionais, incluindo uma modificação proposta da VGG-16. A localização das lesões atingiu uma sensibilidade de 98,70%, empregando uma etapa inicial de segmentação e extração da região de interesse com uma combinação de processamento de imagens e a rede neural artificial TransUNet. Para a segmentação das lesões, foi obtido um índice Dice de 66,55%, também utilizando a TransUNet. Além disso, propõe-se uma nova aplicação de active learning para a seleção de imagens no treinamento, aprimorando o desempenho do modelo. Os resultados demonstram o potencial do método para auxiliar no manejo clínico da endometriose, especialmente no diagnóstico.

Palavras-chave: Endometriose, Aprendizado Profundo, Classificação, Segmentação, Ressonância Magnética

Abstract

Endometriosis is a condition that primarily affects the pelvic region, significantly impacting the quality of life of women, especially those of reproductive age. This study proposes an automated method for classifying patients with endometriosis, localizing, and segmenting lesions in magnetic resonance images of the rectosigmoid region. The method aims to assist in diagnosis, reduce the need for invasive procedures, and map the extent of the lesions. It combines image processing techniques and deep learning, focusing on three main objectives: patient classification, lesion localization, and lesion segmentation. For patient classification, an F1-score of 94.74% was achieved using an ensemble of convolutional neural networks, including a proposed modification of the VGG-16 architecture. Lesion localization reached a sensitivity of 98.70% through an initial segmentation step and a region of interest extraction, employing a combination of image processing and the TransUNet neural network model. Lesion segmentation achieved a Dice score of 66.55%, also using TransUNet. Additionally, the study introduces a novel application of active learning for selecting training images to enhance model performance. These results demonstrate the potential of the proposed method to support the clinical management of endometriosis, particularly in diagnosis.

Keywords: Endometriosis, Deep Learning, Classification, Segmentation, Magnetic Resonance Images

Lista de ilustrações

Figura 1 – Equipamento de RM.	26
Figura 2 – Exemplos de imagens de RM da pelve	26
Figura 3 – Representação da endometriose na pelve	27
Figura 4 – Representação do fluxo da menstruação retrógrada	28
Figura 5 – Principais estruturas pélvicas	28
Figura 6 – Exemplos de lesão de endometriose	29
Figura 7 – Neurônio biológico.	32
Figura 8 – Estrutura básica de rede neural.	33
Figura 9 – Camada convolucional recorrente.	36
Figura 10 – Exibição dos campos receptivos dos filtros em diferentes camadas para duas arquiteturas de redes diferentes.	37
Figura 11 – Arquiteturas da VGG-16 e VGG-19.	39
Figura 12 – Arquitetura da Xception	40
Figura 13 – Arquiteturas DenseNet	41
Figura 14 – Arquitetura da rede Transfomer.	42
Figura 15 – Fluxograma do método proposto	49
Figura 16 – Etapa de Identificação da ROI	50
Figura 17 – Imagem de referência utilizada na especificação do histograma	50
Figura 18 – <i>Template</i> utilizado no <i>template matching</i>	51
Figura 19 – Exemplo de imagem antes e depois da extração da ROI do retossigmoide	53
Figura 20 – Etapa Classificação	54
Figura 21 – VGG-16 modificada	55
Figura 22 – Etapa de segmentação inicial	56
Figura 23 – Ilustração da segmentação inicial	57
Figura 24 – Representação da etapa de extração da ROI da lesão com exemplo real.	58
Figura 25 – Etapa de segmentação da endometriose	59
Figura 26 – Exemplo de imagens da base	61
Figura 27 – Representação do protocolo de marcação das imagens.	62
Figura 28 – Exemplos de resultados da extração da ROI do retossigmoide	63
Figura 29 – Exemplo do processo de construção do conjunto de treinamento na classificação de imagens da lesão.	65
Figura 30 – Exemplo da segmentação inicial da lesão de endometriose para a paciente 22	70
Figura 31 – Exemplo de erro segmentação inicial da lesão de endometriose para a paciente 22	71

Figura 32 – Exemplo da segmentação inicial da lesão de endometriose para a paciente 80	71
Figura 33 – Exemplo da extração da ROI da lesão para a paciente 22	72
Figura 34 – Comparação da segmentação inicial da lesão com a extração da ROI da lesão para a paciente 22	73
Figura 35 – Exemplo da segmentação final da lesão de endometriose para a paciente 22	75
Figura 36 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 22	76
Figura 37 – Exemplos de resultados da extração da ROI do retossigmoide . . .	77
Figura 38 – Monitoramento do F1-score da ViT	78
Figura 39 – Exemplos de resultados da segmentação inicial da lesão em diferentes pacientes e seus respectivos mapas de calor	79
Figura 40 – Comparação da segmentação inicial com e sem extração da ROI do retossigmoide	80
Figura 41 – Comparação entre exemplos de resultados da detecção da lesão com a YOLOv9 e da Segmentação inicial da lesão	81
Figura 42 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 75	82
Figura 43 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 99	82
Figura 44 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 23	83
Figura 45 – Exemplos de resultados da segmentação final da lesão em diferentes pacientes e seus respectivos mapas de calor	84

Lista de tabelas

Tabela 1 – Visão geral dos trabalhos relacionados.	24
Tabela 2 – Resultados da classificação de imagens com e sem <i>active learning</i> (AL)	63
Tabela 3 – Resultados da classificação de pacientes com e sem <i>active learning</i> (AL)	64
Tabela 4 – Distribuição de imagens nos conjunto de treino, validação e teste .	65
Tabela 5 – Parâmetros de função de perda, tamanho do <i>Batch</i> , taxa de aprendi- zado e otimizador avaliados na classificação.	65
Tabela 6 – Parâmetros de função de perda, tamanho do <i>Batch</i> , taxa de aprendi- zado e otimizador selecionados por meio de <i>Grid Search</i> para cada rede na classificação.	65
Tabela 7 – Parâmetros do modelo XGBoost.	66
Tabela 8 – Parâmetros de aumento de dados para a classificação	66
Tabela 9 – Métricas para a classificação por imagem.	67
Tabela 10 – Métricas para a classificação por paciente.	67
Tabela 11 – Matriz de confusão da classificação por imagem para o <i>Ensemble</i> . .	67
Tabela 12 – Matriz de confusão da classificação por paciente para o <i>Ensemble</i> .	67
Tabela 13 – Matriz de confusão da classificação por imagem final para o <i>Ensemble</i> .	68
Tabela 14 – Distribuição de imagens nos conjunto de treino, validação e teste .	68
Tabela 15 – Operações de aumento de dados para a segmentação inicial da lesão	69
Tabela 16 – Parâmetros de função de perda, tamanho do <i>Batch</i> , taxa de aprendi- zado e otimizador para segmentação	69
Tabela 17 – Métricas de desempenho para diferentes modelos de segmentação	69
Tabela 18 – Métricas de desempenho para o TransUNet com diferentes tamanhos de imagem	70
Tabela 19 – Métricas de desempenho para o TransUNet para diferentes modos de avaliação	71
Tabela 20 – Distribuição de imagens nos conjunto de treino, validação e teste .	73
Tabela 21 – Operações de aumento de dados para a segmentação final da lesão	74
Tabela 22 – Parâmetros de função de <i>loss</i> , tamanho do <i>Batch</i> , taxa de aprendizado e otimizador para segmentação	74
Tabela 23 – Métricas de desempenho do TransUNet para segmentação da lesão da endometriose	74
Tabela 24 – Artigos publicados sobre endometriose.	87

Lista de abreviaturas e siglas

ACC	Acurácia
AL	<i>Active Learning</i>
CNN	<i>Convolutional Neural Network</i> (Redes Neurais Convolucionais)
ESP	Especificidade
F1	F1-Score
GLEND	<i>Gynecologic Laparoscopy Endometriosis Dataset</i>
IMC	Índice de Massa Corporal
IoU	<i>Intersection over Union</i> (Interseção Sobre a União)
MLP	<i>Multilayer Perceptron</i>
QDA	Análise Discriminante Quadrática
RM	Ressonância Magnética
PREC	Precisão
RCL	<i>Recurrent Convolutional Layer</i> (Camada Convolucional Recorrente)
ReLU	Função de Ativação Retificada Linear
RNA	Redes Neurais Artificiais
ROI	<i>Region of Interest</i> (Região de Interesse)
SEN	Sensibilidade
TVUS	<i>Transvaginal Ultrasound</i> (Ultrassom Transvanginal)
ViT	<i>Visual Transformer</i>

Sumário

1	INTRODUÇÃO	16
1.1	Objetivo Geral	18
1.1.1	Objetivos específicos	18
1.2	Contribuições	18
1.3	Organização do Trabalho	19
2	TRABALHOS RELACIONADOS	20
2.1	Diagnóstico em Imagens de Ressonância Magnética	20
2.2	Diagnóstico de Endometriose em Laparoscopia	21
2.3	Diagnóstico de Endometriose em TVUS	22
2.4	Outros Trabalhos	23
2.5	Considerações Finais	23
3	FUNDAMENTAÇÃO TEÓRICA	25
3.1	Ressonância Magnética	25
3.2	Endometriose	26
3.2.1	Reto e sigmoide	28
3.3	Processamento de Imagens Digitais	29
3.3.1	Especificação de Histograma	30
3.3.2	Template Matching	31
3.4	Aprendizagem de Máquina	31
3.4.1	Redes Neurais Artificiais	32
3.4.2	Redes Neurais Perceptrons de Múltiplas Camadas	33
3.4.3	<i>Backpropagation</i>	34
3.4.4	Redes Neurais Convolucionais	34
3.4.4.1	Camada Recorrente Convolucional	35
3.4.5	Grad-CAM	35
3.4.6	Deep Features	37
3.4.7	Transfer Learning	37
3.4.8	Arquitetura VGG	38
3.4.9	Xception	38
3.4.10	DenseNet-121	39
3.4.11	U-Net	40
3.4.12	Transformers	41
3.4.12.1	Arquitetura	41
3.4.12.2	Codificação posicional	42

3.4.12.3	Autoatenção	43
3.4.12.4	Atenção de múltiplas cabeças	43
3.4.13	ViT	44
3.4.14	TransUnet	44
3.4.15	Árvores de decisão	45
3.4.16	XGBoost	45
3.4.16.1	Parâmetros	45
3.4.17	Active Learning	46
3.4.18	Data Augmentation	46
3.5	Métricas de Avaliação	47
3.6	Considerações Finais	48
4	MÉTODO PROPOSTO	49
4.1	Extração da ROI do Retossigmoide	49
4.2	Active learning	53
4.3	Classificação das Imagens com Lesão	54
4.4	Segmentação Inicial da Lesão	56
4.5	Extração da ROI da Lesão	57
4.6	Segmentação Final da Lesão	58
4.7	Considerações Finais	59
5	RESULTADOS E DISCUSSÕES	60
5.1	Base de Imagens	60
5.1.1	Protocolo de Marcação das Imagens	61
5.1.2	Preparação da Base de imagens	62
5.2	Extração da ROI do Retossigmoide	62
5.3	Active Learning	62
5.4	Classificação das Imagens com Lesão	64
5.5	Segmentação Inicial da Lesão	67
5.6	Extração da ROI da Lesão	71
5.7	Segmentação Final da Lesão	72
5.8	Discussão	75
5.8.1	Extração da ROI do Retossigmoide	76
5.8.2	<i>Active Learning</i>	77
5.8.3	Classificação das Imagens com Lesão	77
5.8.4	Segmentação da Lesão	78
5.8.5	Limitações	82
5.8.6	Considerações Finais	85
6	CONCLUSÃO	86

6.1	Trabalhos Futuros	86
6.2	Produção científica	87
	REFERÊNCIAS	88

1 Introdução

A endometriose é uma condição inflamatória caracterizada pela presença de células do revestimento uterino (endométrio) fora da cavidade uterina, que pode apresentar sintomas diversos que afetam consideravelmente a qualidade de vida das mulheres. A origem da endometriose ainda é discutida. A teoria mais reconhecida e aceita é da menstruação retrógrada. Segundo esta, quando o tecido não é eliminado durante o período menstrual, move-se na direção oposta, percorrendo as tubas uterinas e alcançando os ovários ou a cavidade abdominal. Devido ao fluxo retrógrado da menstruação, a endometriose pode manifestar-se em diversos órgãos, apresentando sintomas variados conforme a região afetada. Lesões com mais de 5 milímetros são consideradas profundas, enquanto as menores são classificadas como superficiais ([AGARWAL; SUBRAMANIAN, 2010](#); [CARAÇA et al., 2011](#)).

Somente no Brasil, estima-se que existam cerca de 8 milhões de casos da endometriose ([BELLELIS; GIACOMETTI, 2023](#)). É uma doença de alta incidência, afetando mulheres em diversas faixas etárias, incluindo jovens. É estimado que a prevalência da doença seja em torno de 10% da população mundial, com taxas variando entre 5% e 10% em mulheres em idade fértil e entre 3% a 5% em mulheres no período pós-menopausa ([BELLELIS et al., 2010](#); [PODGAEC, 2014](#); [NOGUEIRA et al., 2018](#)). Além disso, 50% das mulheres com dor pélvica crônica antes dos 20 anos apresentam essa condição ([AGARWAL; SUBRAMANIAN, 2010](#)).

Entre os sintomas mais comuns estão a dor pélvica, dismenorreia (dor durante o período menstrual), dispareunia (dor durante o ato sexual), disquezia (dor ao defecar), disúria (dor ao urinar) e infertilidade ([NOGUEIRA et al., 2018](#); [TAYLOR, 2019](#)). A expressão desses sintomas varia dependendo do órgão afetado pela doença e pode ocorrer em locais como o trato intestinal, septo retovaginal, ligamentos uterossacros, reto e sigmoide ([JUNIOR et al., 2008](#); [BELLELIS et al., 2010](#); [CARAÇA et al., 2011](#)).

Conforme o estudo realizado por [Campagnacci et al. \(2005\)](#), o intestino está envolvido em aproximadamente 37% dos casos de endometriose profunda na região pélvica. O reto e o sigmoide são as partes mais frequentemente afetadas no intestino grosso pela endometriose ([AGARWAL; SUBRAMANIAN, 2010](#); [ZWAS; LYON, 1991](#)), representando cerca de 74% dos casos intestinais ([BAILEY; OTT; HARTENDORP, 1994](#); [MOURA et al., 2019](#)).

A endometriose é uma condição debilitante e tem potencial significativo para afetar a qualidade de vida das mulheres em várias faixas etárias. Existem diversos métodos para o diagnóstico da endometriose, incluindo a laparoscopia, a ressonância magnética

(RM) e a ultrassonografia.

A laparoscopia é considerada o padrão-ouro para o diagnóstico definitivo da endometriose (HSU; KHACHIKYAN; STRATTON, 2010; TAYLOR, 2019). Trata-se de uma cirurgia minimamente invasiva que possibilita ao médico visualizar de perto a região de interesse (HSU; KHACHIKYAN; STRATTON, 2010). O procedimento cirúrgico envolve a inserção de um laparoscópio por meio de uma pequena incisão na região pélvica para avaliar a presença de endometriose (NOGUEIRA et al., 2018; PONTES; CLAUDINO, 2021).

A RM é uma técnica não invasiva amplamente utilizada para identificar a endometriose e avaliar as áreas afetadas por essa condição. Essa abordagem permite a visualização dos tecidos por meio de imagens coronais, sagitais e axiais, dispensando o uso de contraste iodado ou radiação ionizante. Além disso, a RM apresenta uma alta resolução espacial, permitindo uma análise detalhada dos compartimentos anterior e posterior da pelve. Ela também oferece uma visão aprofundada da extensão da infiltração da doença (JUNIOR et al., 2008; ANDREUCCI et al., 2014). A ressonância magnética é uma técnica que permite um diagnóstico com boa acurácia e também proporciona resultados mais específicos quando comparada a outras técnicas de imagem, como a ultrassonografia, demonstrando uma acurácia, sensibilidade e especificidade superiores a 90% para endometriose profunda, quando analisadas por radiologistas experientes (OLIVE; SCHWARTZ, 1993; BALLEYGUIER et al., 2002; WYKES; CLARK; KHAN, 2004; BRANDÃO; SILVA, 2013).

Ainda assim, o diagnóstico da endometriose é desafiador e demorado, podendo se estender por até 12 anos (TAYLOR, 2019). Isso ocorre porque os sintomas e características da endometriose são semelhantes aos de outras condições, o que, por vezes, dificulta a identificação. A doença muitas vezes passa despercebida, tornando-se difícil até mesmo para as próprias pacientes reconhecerem a presença do problema.

Métodos automatizados para auxílio no diagnóstico desempenham um papel importante em diversas áreas da medicina (DINIZ et al., 2021; SILVA et al., 2023). No entanto, observa-se uma escassez de pesquisas relacionadas à visão computacional e endometriose, especialmente no contexto de imagens de ressonância magnética.

Este trabalho visa contribuir para o estado da arte, propondo o desenvolvimento de um novo método baseado em técnicas de processamento de imagem e aprendizado profundo, capazes de classificar a presença de lesões de endometriose e realizar a segmentação dessas lesões em imagens de ressonância magnética.

1.1 Objetivo Geral

Este trabalho tem como objetivo desenvolver um método automático para classificação de pacientes com endometriose profunda, detecção e segmentação da área afetada por lesões de endometriose na região do retossigmoide.

1.1.1 Objetivos específicos

Com o objetivo geral proposto neste trabalho em vista, busca-se realizar os seguintes objetivos específicos:

- Desenvolver um método capaz de realizar a classificação, detectar e segmentar lesões em imagens de ressonância magnética de pacientes saudáveis ou com endometriose, utilizando as técnicas de redes neurais convolucionais e processamento de imagem;
- Desenvolver um método capaz de localizar a área afetada pela lesão de endometriose nas imagens de ressonância magnética;
- Desenvolver um método capaz de segmentar a extensão da lesão;
- Avaliar o método proposto empregando as métricas utilizadas no campo de estudo de imagens médicas.

1.2 Contribuições

No contexto apresentado, o trabalho proposto tem as seguintes contribuições:

- Base de imagens de ressonância magnética contendo pacientes com e sem endometriose — a base possui marcação das lesões da doença — que pode ser usada para classificação, segmentação e/ou detecção;
- Utilização do *Active Learning* para seleção das imagens anotadas mais representativas a serem usadas no conjunto de treinamento;
- Modificação da VGG-16 para a classificação de imagens, substituindo alguns dos blocos densos por blocos recorrentes e utilizando XGBoost para classificar;
- Desenvolvimento de abordagens automáticas para localização e definição de ROIs do retossigmoide e da lesão de endometriose nas imagens de ressonância magnética da pelve, empregando processamento de imagem e visando à redução do escopo de estudo e processamento.

- Desenvolvimento de um método automático que classifica pacientes com endometriose profunda no retossigmoide, detecta e segmenta lesão de endometriose em imagens de ressonância magnética;

1.3 Organização do Trabalho

Este texto está estruturado da seguinte forma: no Capítulo 2, são mostrados os trabalhos relacionados ao tema tratado nesta pesquisa. O Capítulo 3 apresenta os principais fundamentos teóricos para a compreensão do trabalho. O método proposto utilizado é detalhado no Capítulo 4. Os resultados são mostrados e discutidos no Capítulo 5. Por fim, é feita uma conclusão sobre o andamento do estudo e dos resultados obtidos 6.

2 Trabalhos Relacionados

Durante a pesquisa, buscou-se estudos que empregam inteligência artificial para a classificação, detecção ou segmentação da endometriose. Além disso, consideraram-se pesquisas sobre o diagnóstico realizado por especialistas. Embora esses estudos não possam ser comparados ao método proposto, por não estarem diretamente relacionados, eles serviram como referência para o trabalho desenvolvido.

Assim, neste capítulo, são apresentados estudos que avaliaram o diagnóstico realizado por especialistas em imagens de ressonância magnética (RM), além de pesquisas sobre classificação e segmentação automática da endometriose em imagens de RM, laparoscopia e ultrassonografia transvaginal (TVUS). Também são discutidos estudos que exploram a predição da endometriose com base em dados clínicos.

2.1 Diagnóstico em Imagens de Ressonância Magnética

No estudo conduzido por [Bazot et al. \(2004\)](#), foi analisada a acurácia dos diagnósticos de endometriose profunda em imagens de RM, realizados por especialistas. O objetivo do estudo era determinar a capacidade do especialista em identificar lesões da doença em várias regiões, incluindo a região do reto e sigmoide, da pelve antes da cirurgia. Dois radiologistas experientes realizaram os diagnósticos em uma amostra de 195 pacientes com suspeita de endometriose. A confirmação da endometriose foi feita através de procedimento cirúrgico. Os resultados mostraram que, para a região do reto e sigmoide, foi possível obter uma sensibilidade de 88% (53 de 60), especificidade de 97,8% (132 de 135) e acurácia de 94,9% (185 de 195).

[Giusti et al. \(2012\)](#) avaliaram a localização de lesões de endometriose profunda a partir do diagnóstico por especialistas em imagens de RM simples e 3D. As imagens 3D são reconstruídas a partir de imagens de RM, de um método semi-automático de segmentação dos órgãos com uma ferramenta chamada ITK-SNAP e do auxílio de radiologistas. Utilizando uma base de dados com 57 pacientes, obtiveram uma acurácia geral de 64% para RM e 83% para RM 3D. Quando a lesão estava especificamente relacionada à infiltração ao retossigmoide, alcançaram acurácias de 58% e 79% para RM e RM 3D, respectivamente. Assim como nos trabalhos supracitados, o diagnóstico final foi obtido por meio de procedimentos cirúrgicos.

O estudo realizado por [Scardapane et al. \(2013\)](#) teve o mesmo objetivo do trabalho anterior: determinar a acurácia do especialista no diagnóstico de endometriose profunda. A pesquisa contou com uma base de 143 pacientes, sendo que 119 deles tinham endometriose. Em sua análise, no geral, obteve-se sensibilidade, especificidade, acurácia

e precisão de 84%, 94,75%, 93,4%, 91,9%, respectivamente. Os resultados indicaram uma acurácia entre 84% e 100%, sensibilidade entre 67% e 100% e especificidade de 85% para os diferentes locais avaliados. A vagina foi o local onde obteve-se 100% de sensibilidade, porém havia um único caso. No caso do intestino, foram identificados 33 casos, com uma acurácia de 97%, dos quais 20 estavam localizados no reto e oito no sigmoide.

Moura et al. (2019) realizaram um estudo comparativo entre vários estudos sobre o diagnóstico de endometriose no retossigmoide em imagens de RM e TVUS. Segundo o levantamento do estudo de 2019, a sensibilidade obtida para ambas as formas de diagnóstico foi de 90%, com especificidade de 96%.

Os trabalhos citados acima fazem comparação entre o diagnóstico por especialistas em imagens de ressonância magnética e em outras formas de diagnóstico, como a observação feita por meio cirúrgico. As bases de imagens utilizadas não foram encontradas ou disponibilizadas. Além disso, nenhuma menciona marcação da lesão ou anotação das imagens.

O trabalho de Yang et al. (2021) faz a classificação de pacientes com endometriose profunda de TVUS e RM 3D. Além disso, fez um estudo do diagnóstico por especialistas através dessas imagens. A base utilizada é privada e consiste em 118 pacientes com endometriose e 206 com outras doenças ginecológicas. Para a classificação automática dos pacientes em RM 3D, utilizou-se o modelo IC3D, uma modificação da arquitetura C3D (TRAN et al., 2015), na qual as quantidades de filtros das últimas camadas completamente conectadas foram alteradas. Neste modelo, alcançou-se acurácia de 99,2% no diagnóstico.

2.2 Diagnóstico de Endometriose em Laparoscopia

No estudo conduzido por Visalaxi e Muthu (2021), foi utilizada a técnica de aprendizagem para classificar imagens de laparoscopia que apresentam a presença de endometriose. O conjunto de imagens utilizado no estudo foi obtido a partir do Gynecologic Laparoscopy Endometriosis Dataset (GLEND) (LEIBETSEDER et al., 2020), que contém aproximadamente 6 mil imagens extraídas de vídeos de laparoscopia. Para treinamento, foram utilizadas 60% das imagens disponíveis na base de dados, e várias arquiteturas foram aplicadas para avaliar o desempenho do modelo. A arquitetura ResNet50 (HE et al., 2016) utilizando pesos pré-treinados da ImageNet (DENG et al., 2009), obteve o melhor resultado, com uma acurácia de 90%, sensibilidade de 82% e precisão de 83%.

Em outro estudo, realizado por Visalaxi, Punnoose e Muthu (2021), teve-se o objetivo de segmentar áreas afetadas pela endometriose por meio de técnicas de

processamento de imagens, tais como: limiar adaptativo e detecção de contornos. Porém, os resultados obtidos se limitam à identificação da intensidade das áreas de lesão das imagens de base, tendo valor médio de 53%.

O estudo de [Leibetseder et al. \(2022\)](#) utilizou Faster R-CNN e Mask R-CNN com *transfer learning* para segmentação de endometriose em imagens de laparoscopia, alcançando uma precisão média de 32,4% com a Mask R-CNN. Além disso, o trabalho também avaliou o impacto de diferentes operações de *data augmentation*, e concluiu que a combinação de recorte e rotação resultou nos melhores resultados. Este também fez uso da base GLEND A.

[Visalaxi e Sudalaimuthu \(2022\)](#) propõem uma análise de similaridade estrutural da endometriose (SSAE), que avalia o agravamento da endometriose em locais distintos, como útero, ovário, peritônio e reto, a partir de imagens laparoscópicas, da base GLEND A. As imagens foram pré-processadas utilizando aumento de dados, com as operações de rotação, deslocamento horizontal, deslocamento vertical, faixa de cisalhamento e faixa de zoom. Em seguida, os dados foram divididos em 70% para treinamento e 30% para o teste. Foi utilizada a rede U-Net, na qual o tamanho do filtro e os otimizadores também foram analisados. Os melhores resultados obtidos foram de 72% de IOU (*Intersection over Union*, Interseção Sobre a União, em português) ([REZATOFIGHI et al., 2019](#)) e dice de 74%. O texto apresentado indica que esses valores foram alcançados na base de validação.

O trabalho de [Saadati e Amirmazlaghani \(2024\)](#) apresenta um método automatizado voltado para a cirurgia robótica de endometriose superficial e em ovários (ou ambas), utilizando tecnologias avançadas de inteligência artificial e robótica. A metodologia proposta inclui a categorização das lesões de endometriose com base em sua localização anatômica, com o objetivo de melhorar a interpretabilidade do sistema. As localizações consideradas para as lesões são: útero, ovário e peritônio. Foram desenvolvidos três frameworks de ensemble U-Net, um para cada região, projetados para detectar e localizar os tipos mais comuns de lesões de endometriose. Esses frameworks empregaram arquiteturas de redes neurais como ResNet50, ResNet101, VGG19, InceptionV3, MobileNet e EfficientNetB7. Adicionalmente, foram criadas duas U-Nets específicas para localizar útero e ovários, o que reduziu ruídos e aumentou as métricas dos modelos. Os resultados alcançados pelos modelos foram expressivos, com índices de IoU de 97,57% para lesões ovarianas, 96,35% para lesões uterinas e 92,58% para lesões peritoneais.

2.3 Diagnóstico de Endometriose em TVUS

[Podda et al. \(2024\)](#) exploraram o uso de ultrassom como um método não invasivo para detectar lesões de endometriose. Eles propuseram uma abordagem de conjunto para a segmentação automática de lesões de endometriose a partir de TVUS. Esse

método combina a predição de múltiplas U-Net treinadas em diferentes tamanhos de imagem (64×64 , 128×128 , 224×224 , 256×256), aumentando as capacidades de discriminação do sistema. Eles utilizaram 75 imagens de pacientes com a doença divididas em treino, validação e teste, e todas as imagens apresentavam lesões. O método proposto alcançou um coeficiente de Dice de 82%.

2.4 Outros Trabalhos

Há estudos que buscam prever ou identificar casos de endometriose através de dados clínicos. O trabalho de [Ghazi et al. \(2016\)](#) baseia-se no estudo do perfil metabólico para a predição de risco de endometriose. A pesquisa utilizou técnicas de reconhecimento de padrões, incluindo redes neurais artificiais (ANNs) e análise discriminante quadrática (QDA) ([AZEVEDO, 2018](#)). A base de dados consiste em 33 pacientes com endometriose e 15 saudáveis. Os resultados indicaram diferenças significativas nos níveis de metabólitos entre mulheres com endometriose e saudáveis. Além disso, o modelo QDA demonstrou maior eficiência em comparação com ANN, um modelo *multilayer perceptron* (MLP). Além disso, o modelo QDA apresentou valores preditivos positivos e negativos de 71% e 78%.

O estudo conduzido por [Chaichian et al. \(2018\)](#) teve como objetivo identificar os fatores que influenciam a endometriose em mulheres em idade reprodutiva, comparando modelos de regressão logística e redes neurais artificiais (RNA). Este incluiu 250 mulheres diagnosticadas com endometriose por laparoscopia como grupo caso e 250 mulheres sem diagnóstico de endometriose. Para investigar os fatores que afetam a ocorrência da doença, como índice de massa corporal (IMC), duração do ciclo menstrual, idade e ocorrência de manchas pré-menstruais, foram aplicados modelos de RNA e regressão logística. Os resultados mostraram que o modelo baseado em RNA apresentou uma AUC ([BRADLEY, 1997](#)) de 94%, superando a AUC de 72% da regressão logística.

O trabalho de [Guerriero et al. \(2021\)](#) teve como objetivo comparar a precisão de sete modelos clássicos de aprendizado de máquina, incluindo RNAs e regressão logística, na identificação de sinais ultrassonográficos associados ao envolvimento de endometriose no intestino. Os dados utilizados foram extraídos de um estudo prévio realizado com 333 pacientes. Entre as 333 pacientes, 106 apresentaram diagnóstico final de endometriose no retossigmoide. Os resultados mostraram que o modelo baseado em RNA obteve a melhor precisão (73%), com sensibilidade de 72% e especificidade de 73%.

2.5 Considerações Finais

Durante a pesquisa, não foram encontrados trabalhos de métodos automáticos diretamente relacionados aos objetivos propostos para RM, especialmente na região do retossigmoide. Pois, além de classificar pacientes com endometriose, indica-se em quais

fatias possuem a lesão da endometriose. E, diferentemente da maioria dos trabalhos citados que envolvem a segmentação da lesão de endometriose em laparoscopia, neste desenvolve-se um método em imagens de RM, método não invasivo.

Assim, neste capítulo foram percorridos alguns estudos semelhantes ao trabalho desta tese, que de algum modo contribuíram para o desenvolvimento desta pesquisa. Na Tabela 1, tem-se um resumo dos trabalhos apresentados neste capítulo, na qual mostram-se os trabalhos que têm foco na endometriose tanto por métodos manuais quanto automáticos.

Tabela 1 – Visão geral dos trabalhos relacionados.

	Trabalho	Método	Base
RM	Bazot et al. (2004)	Diagnóstico por especialista	Privada
	Giusti et al. (2012)	Diagnóstico por especialista	Privada
	Scardapane et al. (2013)	Diagnóstico por especialista	Privada
	Moura et al. (2019)	Diagnóstico por especialista	Privada
	Yang et al. (2021)	Classificação: IC3D	Privada (3D)
Laparoscopia	Visalaxi e Muthu (2021)	Classificação: ResNet50	GLEND A
	Visalaxi, Punnoose e Muthu (2021)	Segmentação: processamento de imagem	GLEND A
	Leibetseder et al. (2022)	Segmentação: Faster R-CNN e Mask R-CNN	GLEND A
	Visalaxi e Sudalaimuthu (2022)	Segmentação: U-Net	GLEND A
	Saadati e Amirmazlaghani (2024)	Segmentação: Ensemble de U-Net	GLEND A
TVUS	Podda et al. (2024)	Segmentação: U-Net	Privada
Outros	Ghazi et al. (2016)	Classificação: MLP e QDA	Privada
	Chaichian et al. (2018)	Predição: RNA e Regressão logística	Privada
	Guerriero et al. (2021)	Classificação: RNAs	Privada

3 Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica deste trabalho, com a explicação dos conceitos essenciais para a compreensão do método desenvolvido. São explorados temas como ressonância magnética, endometriose, processamento de imagens e aprendizado de máquina.

3.1 Ressonância Magnética

A ressonância magnética (RM) é um método de diagnóstico por imagem avançado e eficaz. Este procedimento é considerado não invasivo, o que significa que não demanda cirurgia ou métodos agressivos ao corpo. Além disso, destaca-se por não utilizar radiação ionizante, ao contrário de alguns outros tipos de exames de imagem, como a tomografia computadorizada. A RM viabiliza a obtenção de imagens claras e detalhadas das estruturas internas do corpo humano (HAGE; IWASAKI, 2009).

A RM é amplamente empregada para a visualização de diversas áreas do corpo, incluindo o cérebro, a coluna vertebral, os músculos e os ossos. 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 1972, o pesquisador Lauterbour, da Universidade de Illinois, obteve as primeiras imagens por meio dessa técnica. Em 1976, em Mansfield, na Universidade de Nottingham, foram realizadas as primeiras imagens de uma parte do corpo humano utilizando a RM (HAGE; IWASAKI, 2009).

A RM é uma técnica de imagem médica que utiliza campos magnéticos e ondas de rádio para gerar imagens do interior do corpo. Este método consiste em posicionar o paciente dentro de um campo magnético intenso. 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 (HAGE; IWASAKI, 2009; JÚNIOR; YAMASHITA, 2001). Na Figura 1, tem-se a representação de um equipamento utilizado em procedimentos de RM.

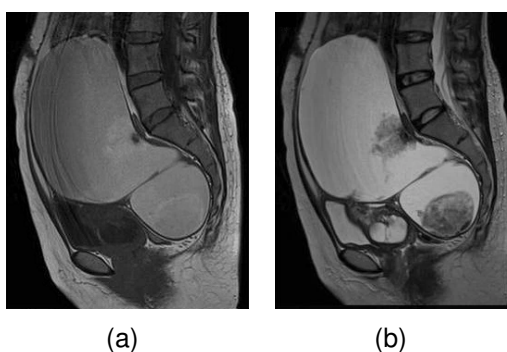
Figura 1 – Equipamento de RM.



Fonte: [Sanar \(2019\)](#).

Existem duas categorias principais de sequências de RM comumente empregadas na prática clínica: as sequências T1 e T2. A sequência de RM T1 é especializada na produção de imagens que revelam a densidade dos tecidos, sendo particularmente eficaz na visualização detalhada da anatomia de tecidos moles e da distribuição de gordura ([JÚNIOR; YAMASHITA, 2001](#)). Por outro lado, as imagens geradas pela sequência T2 são predominantemente sensíveis a líquidos e patologias. Nesse contexto, as regiões líquidas exibem sinais mais intensos, tornando-se visualmente mais claras ([JÚNIOR; YAMASHITA, 2001](#)). A Figura 2 fornece exemplos de imagens de RM da pelve em T1 e T2, respectivamente.

Figura 2 – Exemplos de imagens de RM da pelve. (a) RM em T1 (b) RM em T2.



Fonte: [Yasumoto et al. \(2012\)](#).

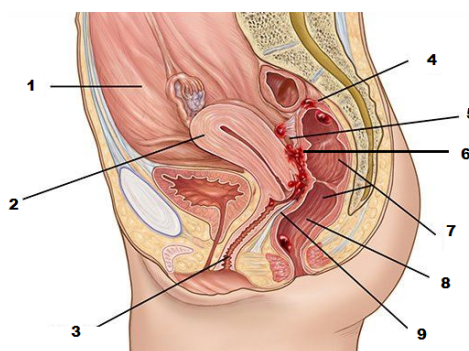
3.2 Endometriose

A endometriose, uma condição inflamatória, caracteriza-se pela presença de células do endométrio fora da cavidade uterina, afetando primariamente órgãos na região pélvica, como o útero, reto e sigmoide (conforme ilustrado na Figura 3). Embora seja mais comum manifestar-se nessas áreas, em casos mais incomuns, a endometriose pode ser identificada em locais atípicos do corpo, como o cérebro e os pulmões ([MARCHIORI et](#)

al., 2012; SARMA et al., 2004).

No estágio inicial da doença, as lesões exibem uma coloração avermelhada. Com a progressão do quadro, essas células são estimuladas e sujeitas a processos inflamatórios, adquirindo tonalidades mais escuras. Ao longo do tempo, a lesão pode apresentar-se mais clara, indicando um processo de fibrose que compromete a funcionalidade do tecido afetado (BERBEL; PODGAEC; ABRÃO, 2008; ANDREUCCI et al., 2014).

Figura 3 – 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.



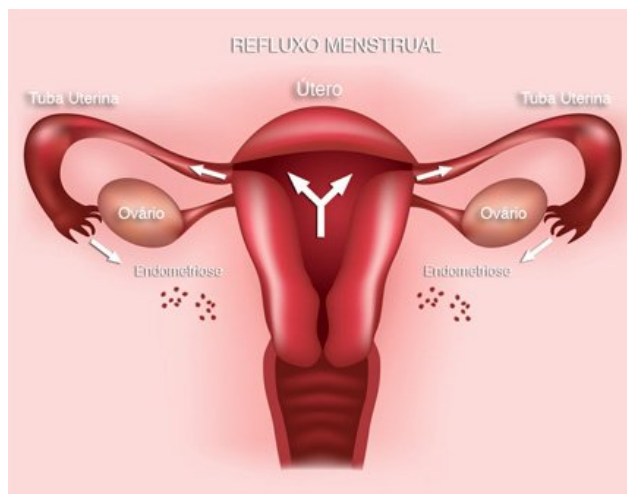
Fonte: Levine (2022).

Diversas teorias foram propostas para explicar a origem da endometriose, sendo a teoria da menstruação retrógrada a mais amplamente aceita. Segundo esta, a menstruação retrógrada ocorre quando o fluxo menstrual se desloca no sentido oposto ao usual, transportando células do endométrio para órgãos além da cavidade uterina, especialmente na região pélvica e abdominal. A Figura 4 ilustra visualmente o curso do fluxo menstrual que resulta na formação das lesões de endometriose.

Outras teorias tentam explicar o surgimento da endometriose em áreas fora da região pélvica. Uma destas teorias está relacionada a mudanças hormonais (AGARWAL; SUBRAMANIAN, 2010). Esta sugere que alterações nos níveis hormonais podem desempenhar um papel na disseminação e implantação de células endometriais em locais não convencionais do corpo.

Quando a endometriose ocorre na região pélvica, a presença da doença pode causar uma série de sintomas, incluindo dor pélvica, sangramento retal, dismenorreia e infertilidade (NOGUEIRA et al., 2018). A gravidade da lesão depende do tamanho da mesma; se for maior que 5 milímetros, é considerada profunda, enquanto, se for menor, é considerada superficial. A lesão profunda é a forma mais grave. Para realizar o diagnóstico da doença, existem vários métodos disponíveis, como RM e laparoscopia. A laparoscopia é considerada o padrão ouro (HSU; KHACHIKYAN; STRATTON, 2010; PONTES; CLAUDINO, 2021; NOGUEIRA et al., 2018), entretanto, este é um procedimento invasivo, diferentemente da RM.

Figura 4 – Representação do fluxo da menstruação retrógrada



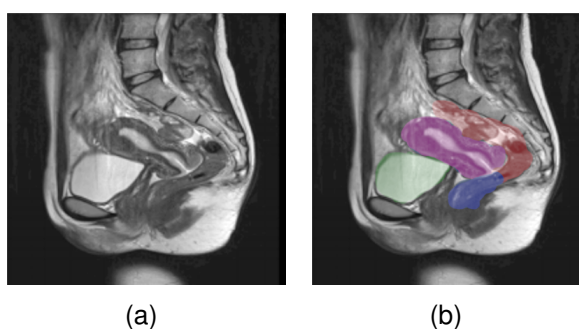
Fonte: [Endometriose \(2012\)](#).

3.2.1 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. Por sua vez, o reto representa 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 ([VARELLA, 2021](#) ; [CARMO, 2022](#)).

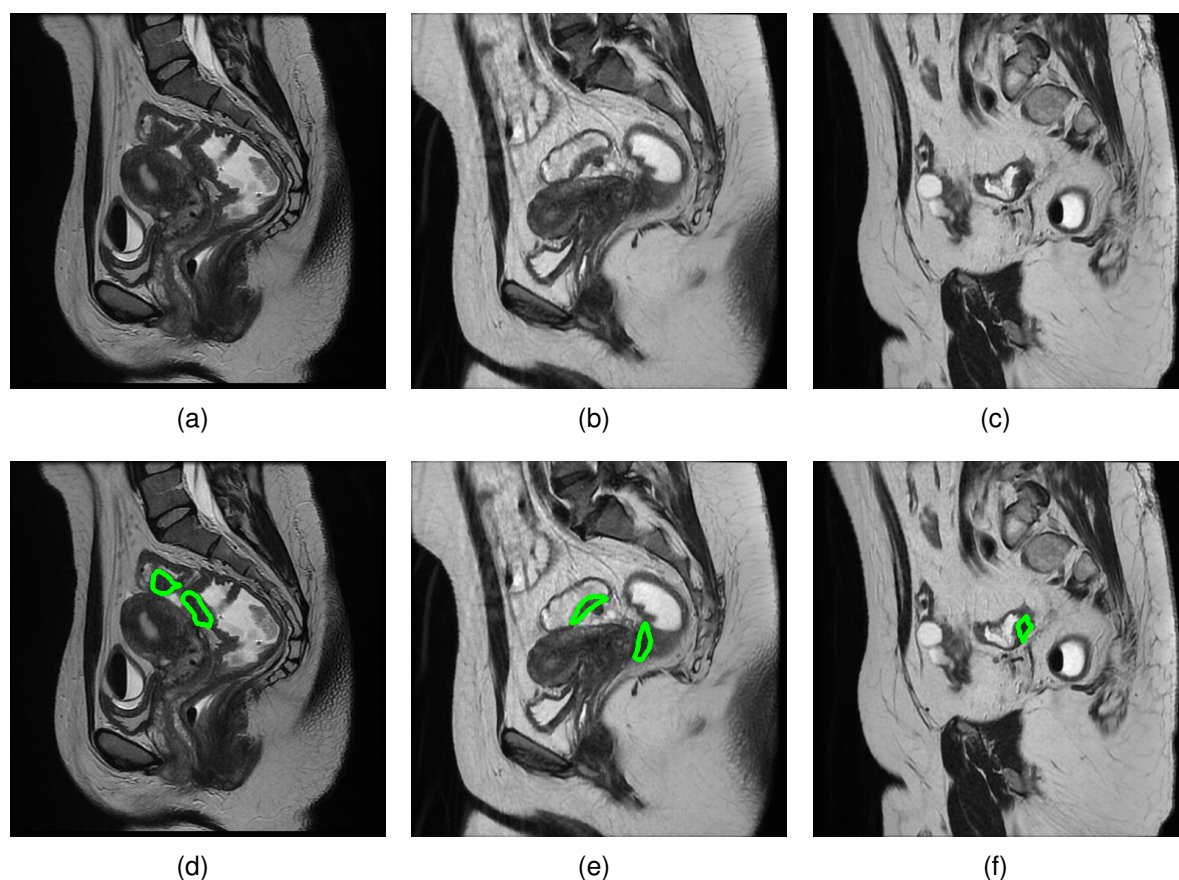
As áreas mais afetadas pela endometriose no intestino grosso são o reto e o sigmoide ([ZWAS; LYON, 1991](#) ; [AGARWAL; SUBRAMANIAN, 2010](#)). A presença dessa condição nessas áreas pode resultar em sintomas como constipação, diarreia, defecação dolorosa, disquezia, hematoquezia, entre outros ([NOGUEIRA et al., 2018](#) ; [TAYLOR, 2019](#)). A Figura 5 ilustra a localização aproximada dessas regiões e de outras estruturas pélvicas femininas importantes em uma RM.

Figura 5 – Principais estruturas pélvicas destacadas: bexiga em verde, útero em rosa, sigmoide em vermelho e reto em azul. (a) Imagem de RM (b) Imagem com estruturas pélvicas destacadas.



Na Figura 6, são apresentados exemplos de lesões de endometriose localizadas na região do retossigmoide. 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 retossigmoide é pouco perceptível.

Figura 6 – Exemplos de lesão de endometriose. (a), (b) e (c) imagens originais, e (d), (e) e (f) suas respectivas lesões marcadas.



3.3 Processamento de Imagens Digitais

Esta seção fornece um detalhamento das técnicas de processamento de imagens digitais empregadas em vários pontos no desenvolvimento do trabalho.

O processamento de imagens digitais refere-se à aplicação de técnicas computacionais capazes de interpretar uma imagem de entrada e transformá-la em informações. Este processo é dividido em três níveis: baixo, médio e alto. No nível baixo, são realizadas operações básicas, como pré-processamento para aprimorar a qualidade da imagem, utilizando técnicas como remoção de ruído ou ajuste de contraste para destacar ou suprimir características específicas da imagem. O nível médio concentra-se na

detecção de atributos significativos, como a identificação de bordas e a extração de áreas relevantes. Por fim, nos processos de alto nível, são executadas operações que geram significado aos atributos extraídos, interpretando os resultados obtidos nas etapas anteriores (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 em partes significativas, a representação seleciona atributos que caracterizam a informação relevante e, por fim, o reconhecimento atribui rótulos e significados aos resultados (FILHO; NETO, 1999). 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, uma metodologia específica pode incluir apenas um subconjunto delas, dependendo do problema a ser solucionado.

3.3.1 Especificação de Histograma

A especificação de histograma é uma técnica empregada para ajustar a distribuição de intensidades de uma imagem de entrada, de forma que ela se aproxime da distribuição de uma imagem de referência (GONZALEZ; WOODS, 2010).

O histograma de uma imagem digital, cujos níveis de intensidade variam entre 0 e $L - 1$, é dado pela expressão $h(r_k) = n_k$, onde r_k representa o k -ésimo nível de intensidade, e n_k é o número de *pixels* na imagem com essa intensidade r_k (GONZALEZ; WOODS, 2010). Esse histograma descreve a distribuição dos níveis de intensidade na imagem.

O processo de especificação de histograma ocorre em duas etapas. Primeiramente, é necessário equalizar o histograma da imagem original $h(x)$ para obter um histograma intermediário $h^*(z)$. Isso é feito utilizando a Equação 3.1, onde n_i representa o número de *pixels* com o nível de intensidade i , e N é o total de *pixels* da imagem, com L níveis de cinza (FERREIRA et al., 2019).

$$y(i) = \frac{L}{N} \sum_{i=0}^{L-1} n_i \quad (3.1)$$

A função $z = f(x)$ é responsável por equalizar o histograma original $h(x)$, gerando o histograma intermediário $h^*(z)$. Em seguida, a transformação $z = g(y)$ equaliza o histograma de referência $h_o(y)$. O mapeamento global dos valores de intensidade de brilho necessários para transformar $h_i(x)$ em $h_o(y)$ é obtido por meio da transformação inversa $y = g^{-1}(f(x))$. Esse procedimento alinha os valores da imagem equalizada com

os valores mais próximos no histograma especificado (ALAN; JIA et al., 2006; FERREIRA et al., 2019).

3.3.2 Template Matching

O *Template Matching* é uma técnica de processamento de imagem que tem o objetivo de identificar uma região de uma imagem baseando-se em uma imagem pré-selecionada, chamada de *Template*. Para a obtenção deste resultado, o *Template* se desloca por toda a imagem em busca da região mais semelhante. Há uma variedade de métodos disponíveis para realizar o cálculo de semelhança entre duas imagens (BRUNELLI, 2009; DINIZ et al., 2021).

O *Template Matching* é uma técnica de processamento de imagem que visa identificar uma região em uma imagem com base em uma imagem predefinida, denominada *Template*. Para alcançar esse resultado, o *Template* percorre toda a imagem em busca da região mais semelhante. Existem diversas abordagens disponíveis para calcular a semelhança entre duas imagens (BRUNELLI, 2009; DINIZ et al., 2021).

Neste trabalho, usa-se a Equação 3.2, que calcula a diferença dos quadrados (R_{sqdif}) entre a imagem I e a imagem usada como *template* T para encontrar a similaridade da combinação (*matching*) de ambas. Os valores da equação comparam T com a sub-imagem de I onde T está. De modo que x e y são as coordenadas das imagens. $R \in [0, \infty]$ como valor resultante, no qual 0 significa perfeita combinação (DINIZ et al., 2021).

$$R_{sqdif}(x, y) = \sum_{x', y'} [T(x', y') - I(x + x', y + y')]^2 \hat{y}(w) = \sigma \left(\sum_{i=1}^N x_i w_i \right) \quad (3.2)$$

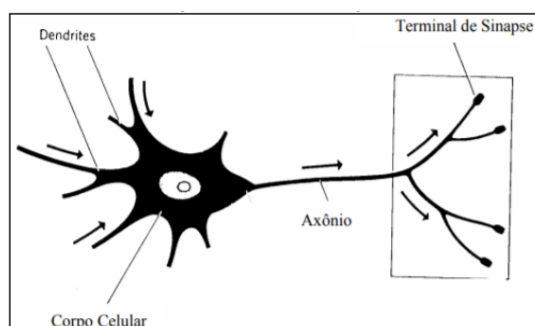
3.4 Aprendizagem de Máquina

A aprendizagem de máquina fundamenta-se na ideia de aprendizado por experiência, em que uma máquina é treinada em relação a um determinado problema por meio de um conjunto de exemplos. Neste processo, a máquina aprende a identificar padrões e classificar informações através da inferência. A área de aprendizagem de máquina é extensivamente pesquisada atualmente devido aos seus resultados notáveis em reconhecimento de padrões. Essa abordagem é amplamente aplicada em diversas áreas, como reconhecimento de fala, diagnósticos médicos, previsões, entre outras (LORENA; CARVALHO, 2007).

3.4.1 Redes Neurais Artificiais

As técnicas de Redes Neurais Artificiais (RNA) empregam mecanismos adaptativos para possibilitar a aprendizagem com base na experiência, aprimorando continuamente o conhecimento adquirido. Inspiradas nos modelos matemáticos de raciocínio do cérebro humano, as RNA buscam emular a atividade dos neurônios. Os componentes fundamentais de um neurônio (Figura 7) incluem os dendritos, que captam os impulsos originados por outros neurônios, o corpo celular, responsável pelo processamento dos sinais recebidos, o axônio, que transmite os sinais, e a sinapse, que conecta o axônio ao dendrito de outro neurônio (PACHECO; PEREIRA, 2018).

Figura 7 – Neurônio biológico.



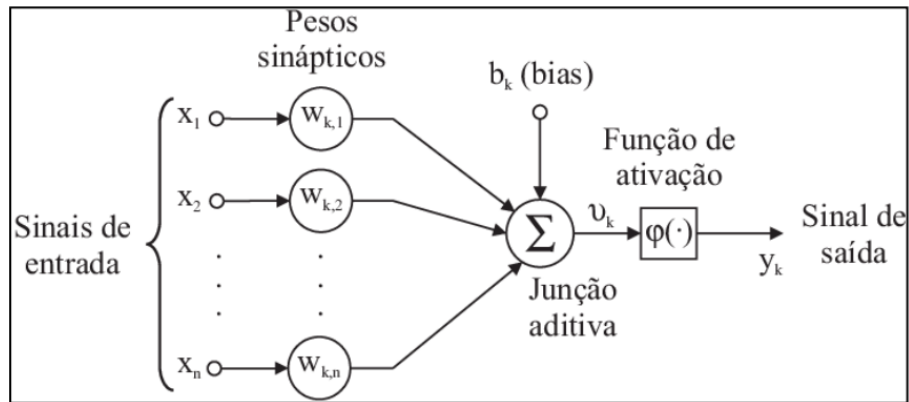
Fonte: Bianchini et al. (2001)

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 ligações (sinapses) para conectar as entradas à rede, sendo essas representadas por pesos sinápticos. As entradas são multiplicadas pelos pesos associados ao neurônio. Adicionalmente, a rede incorpora um termo adicional, conhecido como *bias*, que é somado aos sinais de entrada ponderados pelos pesos. Por fim, uma função de ativação é aplicada para limitar a amplitude do sinal de saída (Figura 8).

A aprendizagem de uma RNA é realizada por meio de um processo conhecido como treinamento, no qual são fornecidos exemplos de entrada e saída, constituindo a base de dados para o aprendizado. Existem duas formas principais de aprendizado: supervisionado e não supervisionado. No aprendizado supervisionado, o algoritmo utiliza exemplos com pares de entrada e saída conhecidos para ajustar seus parâmetros. Em contraste, no aprendizado não supervisionado, o algoritmo procura aprender agrupando as semelhanças entre os dados de entrada, sem a necessidade de exemplos específicos. Após o treinamento, é realizado o teste utilizando dados diferentes para avaliar o desempenho da RNA.

A partir dos estudos das RNAs, originou-se o estudo do aprendizado profundo. O aprendizado profundo se caracteriza pelo uso de múltiplas camadas e, diferentemente

Figura 8 – Estrutura básica de rede neural.



Fonte: Bianchini et al. (2001)

das RNAs tradicionais, pela autonomia na extração de características, sem a necessidade de um processo externo como técnicas de processamento de imagem (HUA et al., 2015). As redes neurais convolucionais são algumas das abordagens de aprendizado profundo empregadas neste trabalho.

3.4.2 Redes Neurais Perceptrons de Múltiplas Camadas

A Rede Neural Perceptrons de Múltiplas Camadas (MLP, do inglês *Multilayer perceptron*) é um tipo de rede neural artificial que consiste em múltiplas camadas de neurônios interconectados. Cada camada é composta por unidades de processamento, conhecidas como neurônios, que recebem entradas, aplicam uma função de ativação e transmitem sinais para a próxima camada.

Uma das características distintivas da MLP é sua arquitetura em camadas, que inclui uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. A camada de entrada recebe os dados de entrada e os transmite para as camadas ocultas, onde ocorre o processamento intermediário. Cada neurônio em uma camada oculta combina as entradas ponderadas com os pesos associados e as passa por uma função de ativação. O cálculo da saída de uma camada MLP é demonstrado na Equação 3.3 (SILVA, 2004; DINIZ et al., 2021).

$$y^{(l)} = \sigma(W^{(l)}y^{(l-1)} + b^{(l)}) \quad (3.3)$$

onde $y^{(l)}$ é o vetor de saída da l -ésima camada oculta, σ é a função de ativação, $W^{(l)}$ é a matriz de peso entre a camada $l - 1$ e l , $y^{(l-1)}$ é a saída da camada anterior e $b^{(l)}$ é o vetor de *bias* da camada.

3.4.3 Backpropagation

Backpropagation é um algoritmo usado para calcular os gradientes dos pesos das conexões da rede neural em relação a uma função de custo, o que permite ajustar esses pesos de forma a minimizar o erro de predição durante o treinamento.

O processo de backpropagation envolve duas etapas principais: a propagação para frente (*forward propagation*) e a propagação para trás (*backward propagation*). Durante a propagação para frente, os dados de entrada são introduzidos na rede neural, atravessam cada camada e geram as saídas correspondentes. Após isso, a propagação para trás é realizada comparando a saída esperada com a saída produzida. Se elas forem diferentes, calcula-se o erro. O erro é propagado desde a camada de saída até a camada de entrada, recalculando os pesos de toda a RNA (SILVA, 2004).

O processo de backpropagation é repetido para múltiplos exemplos do treinamento até que a rede convirja para uma solução que minimize a função de custo.

3.4.4 Redes Neurais Convolucionais

As redes neurais convolucionais (CNN, do inglês *Convolutional Neural Network*) são amplamente utilizadas para detectar, segmentar e reconhecer padrões em imagens e vídeos. A extração de características é realizada através da aplicação de filtros (convoluções) na imagem. A imagem é processada em várias camadas, como a camada de convolução, subamostragem e camada totalmente conectada. Uma CNN pode ter múltiplas camadas de convolução e, comumente, tem camadas de subamostragem entre elas.

Na CNN, a etapa de convolução é realizada aplicando um filtro (núcleo) na imagem de entrada. Este processo serve para destacar características importantes. O resultado é chamado de mapa de características.

Após a etapa de convolução, é feita a subamostragem (*pooling*), que agrupa as características destacadas. O agrupamento pode ser realizado usando técnicas como o *max pooling*, que seleciona o valor máximo em áreas do mapa de características.

A camada completamente conectada é formada por uma rede neural artificial comum (discutida na Seção 3.4.1), que determina a relação existente entre os atributos da camada anterior e cada classe, e ocorre o reconhecimento.

São usadas funções de ativação entre as camadas da CNN. A função de ativação unidade linear retificada (ReLU) (GOODFELLOW; BENGIO; COURVILLE, 2016) é descrita pela Equação 3.4 e consiste em substituir cada valor negativo por 0.

$$f(x) = \max(x, 0) \quad (3.4)$$

Além das camadas mencionadas anteriormente, existem técnicas adicionais que são utilizadas para melhorar o aprendizado e evitar o *overfitting*. O *overfitting* ocorre quando o modelo se ajusta demais aos dados de entrada e apresenta boas performances no treinamento, mas não no teste. Uma das técnicas utilizadas para mitigar esse problema é o *dropout*, que consiste em remover aleatoriamente neurônios e suas conexões de entrada e saída da rede (SRIVASTAVA et al., 2014). Esse desligamento evita o *overfitting*, combinando diferentes formas das arquiteturas da rede.

3.4.4.1 Camada Recorrente Convolutacional

Foram desenvolvidas diversas arquiteturas de RNA com o objetivo de criar modelos mais profundos e aprimorar as capacidades do sistema, incluindo a adição de conexões saltadas. Uma abordagem proposta foi a utilização de uma rede neural convolutacional recorrente (RCNN) (LIANG; HU, 2015), que incorpora a ideia de conexões recorrentes.

Na Figura 9, é possível observar uma camada convolutacional recorrente (RCL, *recurrent convolutional layer*, em inglês), composta por uma convolução simples e duas convoluções recorrentes que compartilham pesos entre si. Os estados de uma RCL evoluem sobre tempo de passo discreto (LIANG; HU, 2015). Sendo x_l uma entrada na layer l , e com pixel na posição (i, j) no mapa de características k , tem-se a saída $(O_{ijk}^l(t))$ para um tempo t definida na Equação 3.5.

$$O_{ijk}^l(t) = (w_k^f)^T x_l^{f(i,j)}(t) + (w_k^r)^T x_l^{r(i,j)}(t-1) + b_k \quad (3.5)$$

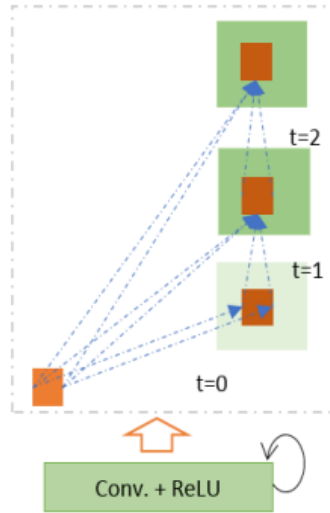
Nesta equação, $x_l^{(i,j)}(t)$ e $x_l^{(i,j)}(t-1)$ representam a entrada direta (f , de *feedforward*) e a da recorrência (r), respectivamente. Assim como $(w_k^f)^T$ e $(w_k^r)^T$ referem-se aos pesos de ambas as entradas citadas anteriormente no mapa de características k . E, b_k é o *bias*. E, por fim, a operação é passada para uma função de ativação ReLU.

Essa arquitetura facilita o fluxo de informação entre as camadas, resultando na redução do número de parâmetros do modelo e possibilitando a construção de uma rede mais profunda. A presença de caminhos mais longos na camada convolutacional recorrente (RCL) permite que o modelo aprenda características complexas da vizinhança (ALOM et al., 2019). Essa camada é utilizada na modificação da VGG-16 (SIMONYAN; ZISSERMAN, 2015) desenvolvida neste trabalho.

3.4.5 Grad-CAM

O método Gradient-weighted Class Activation (SELVARAJU et al., 2017) faz uso dos gradientes provenientes da última camada convolutacional de uma rede neural,

Figura 9 – Camada convolucional recorrente.



Fonte: [Alom et al. \(2019\)](#)

gerando assim um mapa de ativação visual. Esse mapa apresenta as áreas da imagem que o CNN considera relevantes. O Grad-CAM é uma generalização do Class Activation Mapping (CAM). O CAM exige o uso de camada de *global averaging pooling*, limitando o seu uso em diversas CNNs, e não usa as informações do gradiente.

O Grad-CAM pode ser descrito pela Equação 3.6, onde $L_{\text{Grad-CAM}}^c$ é o mapa de ativação da classe c , A_k representa a ativação da última camada convolucional para a k -ésima característica, $\alpha_k^c A_k$ é o peso atribuído à característica k para uma classe c . A função ReLU é aplicada para garantir que as contribuições negativas das ativações sejam descartadas, pois só as ativações positivas influenciam o mapa de ativação ([SELVARAJU et al., 2017](#); [SILVA, 2023](#)).

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A_k \right) \quad (3.6)$$

Os pesos $\alpha_k^c A_k$ são calculados a partir dos gradientes em relação à ativação da última camada convolucional, representado pela Equação 3.7, onde y^c é a saída para classe c e Z é uma constante de normalização.

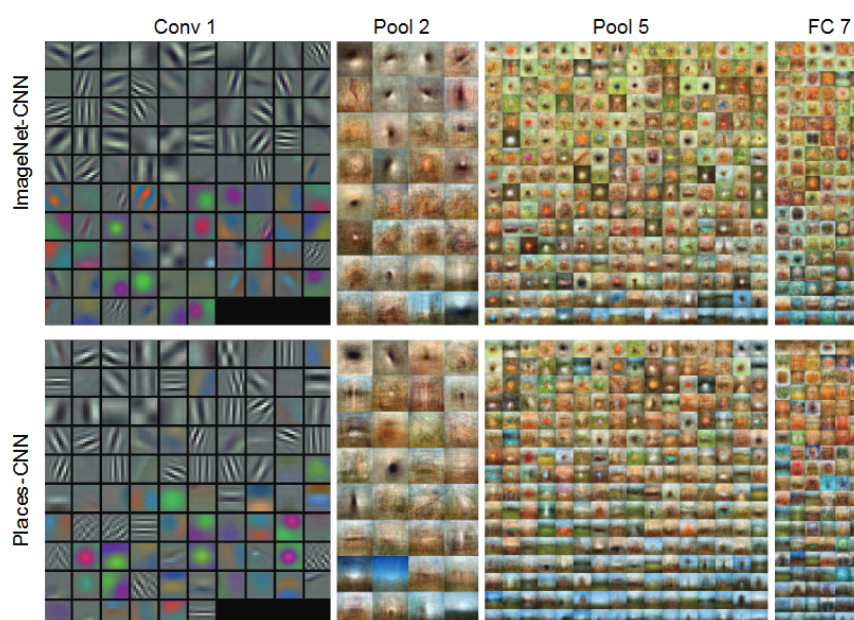
$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \quad (3.7)$$

Deste modo, o mapa de calor obtido pode ser utilizado para visualizar quais regiões de uma imagem influenciam mais a decisão de uma rede neural convolucional.

3.4.6 Deep Features

As *deep features* são representações em camadas mais profundas de uma rede neural, obtidas após o treinamento da rede. Durante o processo, a entrada é alimentada na rede, e as saídas são capturadas a partir de uma camada específica (camada escondida), antes da camada de saída. Essas representações são empregadas para diversos propósitos, como transferência de aprendizado e execução de tarefas relacionadas à visão computacional. A Figura 10 ilustra como as características evoluem de detectores simples de bordas e contornos nas camadas iniciais para características mais complexas, incorporando informações espaciais e de textura nas camadas finais (ZHOU et al., 2014).

Figura 10 – Exibição dos campos receptivos dos filtros em diferentes camadas para duas arquiteturas de redes diferentes.



Fonte: Zhou et al. (2014)

Além disso, *deep features* são frequentemente utilizadas em tarefas de *Transfer learning*, em que uma rede pré-treinada é utilizada como extratora de características para resolver algum outro problema. Essa abordagem é especialmente útil quando há escassez de dados para treinar uma nova rede do zero, e a rede pré-treinada já aprendeu características úteis para o novo problema. No contexto deste trabalho, essas características são aplicadas no método proposto.

3.4.7 Transfer Learning

Para a construção de um modelo de reconhecimento robusto, é geralmente necessário um conjunto considerável de imagens, por vezes exigindo milhares delas

(SHIN et al., 2016). Uma abordagem para contornar a limitação de dados é a utilização de *Transfer Learning*, uma técnica na qual um modelo pré-treinado em um problema é empregado em outro contexto. Pode-se utilizar um modelo pré-treinado para extrair características e treinar um novo modelo a partir dessas características. Alternativamente, é possível inicializar os pesos de uma rede neural com base no pré-treinamento e treiná-la desde o início até a camada de saída (YOSINSKI et al., 2014).

Neste trabalho, o *Transfer Learning* foi usado na inicialização dos pesos nas etapas de classificação e segmentação.

3.4.8 Arquitetura VGG

As arquiteturas de rede neural de aprendizado profundo chamadas de VGG-16 e VGG-19 foram criadas por uma equipe da Universidade de Oxford em 2014. Ambas foram desenvolvidas na base de dados ImageNet, que inclui milhões de imagens categorizadas em 1000 categorias distintas (SIMONYAN; ZISSERMAN, 2015).

A arquitetura VGG se destaca pela sua arquitetura simples. Ela é composta por camadas consecutivas de filtros de convolução e camadas de *max pooling*, que são seguidas por camadas densas, também conhecidas como camadas totalmente conectadas. Essa configuração é repetida diversas vezes, criando uma rede profunda e com uma grande quantidade de parâmetros que podem ser treinados.

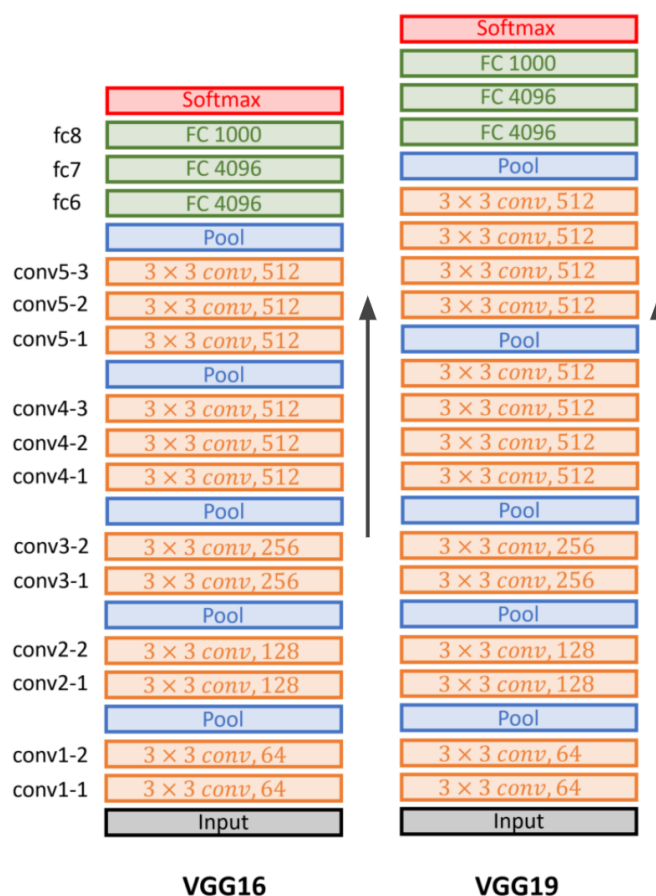
A rede processa a entrada através de camadas consecutivas de convolução combinadas com a função de ativação ReLU. As camadas de filtro de convolução usam filtros 3×3 com diferentes canais de saída, dependendo da camada. Depois de algumas camadas desta sequência, é realizada uma operação de *max pooling* para diminuir a dimensionalidade da representação da imagem. Nas camadas subsequentes às camadas de filtro de convolução, encontram-se três camadas totalmente conectadas, também referidas como camadas densas. Essas camadas têm a função de combinar as informações adquiridas nas camadas anteriores, a fim de realizar uma previsão final quanto à categoria da imagem (SIMONYAN; ZISSERMAN, 2015). Desse modo, esta é usada para classificação das imagens do paciente. A Figura 11 apresenta a arquitetura da VGG-16 e VGG-19 em detalhes.

3.4.9 Xception

A Xception (Extreme Inception) é uma arquitetura de rede neural convolucional (CNN) que foi proposta por Chollet (2017). Ela é considerada uma evolução da arquitetura Inception (SZEGEDY et al., 2015), introduzindo uma abordagem chamada convolução separável em profundidade (*depthwise separable convolution*).

Na Xception, ao invés de aplicar uma convolução tradicional em todos os canais

Figura 11 – Arquiteturas da VGG-16 e VGG-19.



Fonte: [Hacker \(2022\)](#)

de entrada para cada filtro, o processo é dividido em duas etapas. Na primeira, é aplicado um filtro em cada canal da entrada. Em seguida, emprega-se uma convolução 1×1 a fim de combinar a saída dos canais.

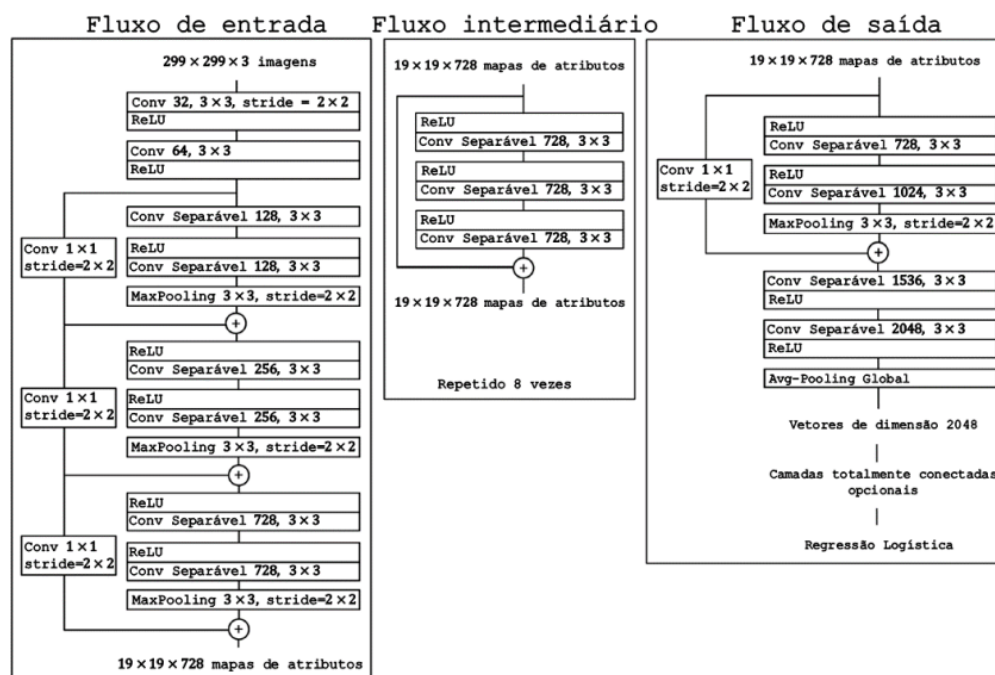
A arquitetura da Xception é formada por 36 camadas de convolução separável divididas em 14 módulos, cada um destes possui conexões residuais, introduzidas por [He et al. \(2016\)](#). Pode ser dividida em 3 partes: fluxo de entrada, fluxo intermediário e fluxo de saída ([CHOLLET, 2017](#)) (Figura 12).

3.4.10 DenseNet-121

A arquitetura DenseNet foi introduzida por ([HUANG et al., 2017](#)), e o número “121” indica a quantidade total de camadas na rede. A principal característica da DenseNet é a conexão densa entre camadas adjacentes.

Em vez de seguir a arquitetura tradicional onde as camadas são conectadas sequencialmente, a DenseNet adota uma abordagem densa, conectando cada camada a todas as camadas subsequentes. Assim, uma camada utiliza todos os mapas de características das camadas anteriores. Tal alteração ajuda na redução do problema de

Figura 12 – Arquitetura da Xception



Fonte: BARBOSA et al. (2022)

fuga do gradiente, no qual a informação se perde devido à profundidade da rede (HUANG et al., 2017).

A arquitetura de uma dense possui dois principais elementos: os blocos densos e as camadas de transição. Um bloco denso consiste em várias camadas com conexões densas. As camadas de transição incluem uma camada de convolução e uma camada de *pooling* e são usadas para reduzir a resolução espacial e o número de canais.

As arquiteturas do tipo DenseNet podem ser observadas na Figura 13. A arquitetura usada neste trabalho foi a DenseNet-121.

3.4.11 U-Net

A U-Net (RONNEBERGER; FISCHER; BROX, 2015) é uma arquitetura de rede neural artificial desenvolvida para realizar a segmentação de imagens. Essa arquitetura adota uma configuração simétrica em formato de U, composta por uma parte de contração e uma parte de expansão. A parte de contração é constituída por camadas de convolução e *max pooling*, sendo utilizada para capturar o contexto da imagem. Por outro lado, a parte de expansão realiza a localização precisa e aumenta o tamanho dos mapas de características por meio de superamostragem. A interconexão entre as partes é estabelecida por meio de conexões salteadas. Na camada final da rede, é aplicada a função de ativação sigmoide para realizar a predição (RONNEBERGER; FISCHER; BROX, 2015).

Figura 13 – Arquiteturas DenseNet

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Fonte: [Huang et al. \(2017\)](#)

3.4.12 Transformers

As redes de arquitetura Transformers foram propostas em 2017 por [Vaswani et al. \(2017\)](#). Foi concebida para tarefas de processamento de linguagem natural, como a tradução automática de texto. As principais características da arquitetura são a capacidade de processar dados sequenciais longos de forma paralela e o mecanismo de atenção que permite que a rede neural consiga dar importância a múltiplos dados da entrada.

3.4.12.1 Arquitetura

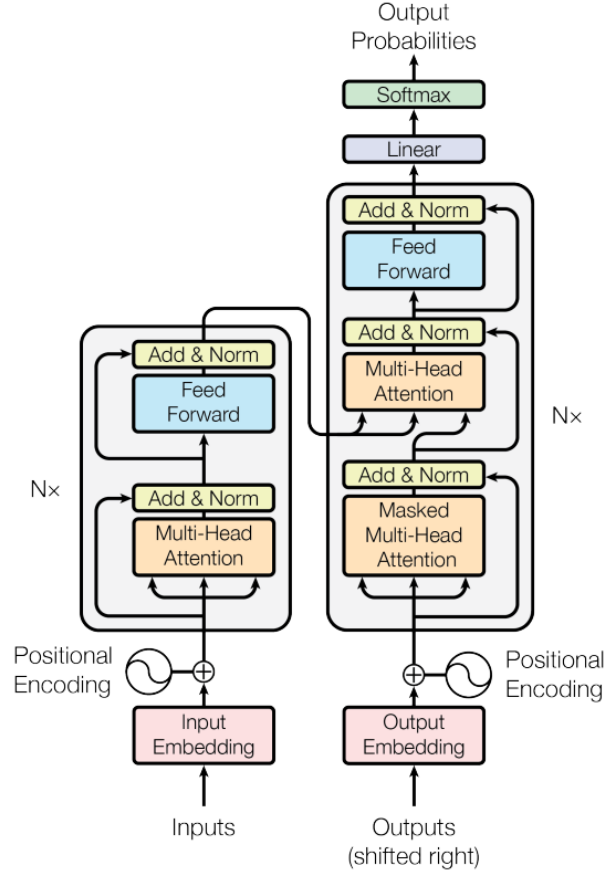
A arquitetura da rede *transformer* é do tipo codificador e decodificador. A parte codificadora gera uma sequência de símbolos representativos Z a partir de uma sequência X . Com Z , a parte decodificadora gera como saída uma sequência Y ([VASWANI et al., 2017](#)).

A parte codificadora consiste, primeiramente, de uma camada de codificação posicional (*positional encoding*), que transforma as palavras em vetores, gerando contexto baseado na posição das palavras. Além disso, esta parte é formada por 6 camadas idênticas, cada uma possui 2 subcamadas. A primeira subcamada é o mecanismo de autoatenção de múltiplas cabeças, e a segunda é uma camada *feedforward* completamente conectada. Utiliza-se uma camada de normalização e conexões residuais ao redor de cada subcamada ([VASWANI et al., 2017](#)).

A parte decodificadora também contém 6 camadas idênticas. Estas são semelhantes às camadas da parte codificadora, entretanto possuem mais uma subcamada de autoatenção de múltiplas cabeças que recebe a saída da parte

codificadora (VASWANI et al., 2017). Uma representação da arquitetura é mostrada na Figura 14.

Figura 14 – Arquitetura da rede Transfomer.



Fonte: Vaswani et al. (2017)

3.4.12.2 Codificação posicional

A fim de ter a informação das posições das palavras na sequência, cria-se uma representação da posição de cada uma, que é incorporada à representação de cada respectiva palavra. O cálculo da Codificação posicional é definido pelas Equações 3.8 e 3.9. Nestas equações, pos é a posição na sequência, d é a dimensão, 10000 é um valor definido por Vaswani et al. (2017), e i é um valor que indica alternância entre o seno e cosseno para o tamanho da sequência.

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right) \quad (3.8)$$

$$PE(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right) \quad (3.9)$$

3.4.12.3 Autoatenção

Autoatenção é um mecanismo de atenção que faz a relação de diferentes partes de uma sequência X a fim de gerar uma representação da sequência Y , ambas as sequências têm uma dimensão d (VASWANI et al., 2017).

O processo da autoatenção calcula os pesos de atenção que são usados para ponderar a relevância das partes da entrada entre si (VASWANI et al., 2017).

O mecanismo utiliza três vetores para calcular o valor de atenção (ou, *attention score* em inglês) de cada palavra em relação às demais palavras da sequência: Q (consulta), K (chaves), V (valor) (VASWANI et al., 2017).

Estes vetores são criados a partir da multiplicação da entrada com 3 matrizes com valores calculados durante o treinamento (WQ , WK , e WV). Calcula-se Q , K e V para toda entrada X . Em seguida, aplica-se o produto escalar entre o vetor de consulta de uma entrada x_i e todos os vetores de chave da sequência. Essa operação gera um vetor de *scores*.

O próximo passo é a aplicação de uma função de ativação softmax. Porém, o vetor é normalizado por $\frac{1}{\sqrt{d_k}}$, pois a função não tem bom desempenho para valores altos, eliminando gradientes e retardando o aprendizado. O valor obtido para cada entrada é multiplicado pelos respectivos vetores V . Os vetores resultantes são somados, gerando o score final. O processo então é repetido para cada entrada (x_{i+1}) (VASWANI et al., 2017). O cálculo pode ser observado na Equação 3.10.

$$\text{Attention} = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (3.10)$$

3.4.12.4 Atenção de múltiplas cabeças

A atenção de múltiplas cabeças pode ser vista como uma extensão da autoatenção simples. Desta forma, utiliza-se múltiplas cabeças de atenção paralelamente. A principal diferença é que os valores das matrizes WQ , WK e WV são diferentes em cada cabeça. Ao final do processo em cada cabeça, as matrizes são concatenadas e multiplicadas a uma nova matriz de peso W^0 . Então, a matriz resultante é passada para a camada *feedforward*. Este processo permite que cada cabeça consiga focar em partes diferentes da entrada, capturando informações mais complexas (VASWANI et al., 2017).

O cálculo de atenção de múltiplas cabeças pode ser observado nas Equações 3.11 e 3.12, onde QW_i^Q , KW_i^K , VW_i^V são as matrizes de transformação para uma cabeça i .

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (3.11)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.12)$$

3.4.13 ViT

Uma Vision Transformer (ViT) (DOSOVITSKIY et al., 2021) é uma arquitetura baseada na arquitetura de uma Transformer, que visa resolver problemas de visão computacional. A entrada da ViT é uma imagem $x \in \mathbb{R}^{H \times W \times C}$ dividida em patches $x_p \in \mathbb{R}^{N \times (P^2 C)}$. Ela passa por um processo de codificação de patches, conforme definido na Equação 3.13, onde D é o número de dimensões, N é o número de patches, a matriz de projeção de incorporação é representada por $E \in \mathbb{R}^{(P^2 \cdot C) \times D}$, e a matriz de posição de incorporação é representada por $E_{pos} \in \mathbb{R}^{N \times D}$.

$$z_0 = [x_p^1 E, x_p^2 E, \dots, x_p^N E] + E_{pos} \quad (3.13)$$

3.4.14 TransUnet

A TransUNet segue o padrão de codificação-decodificação similar à arquitetura de uma U-Net convencional, porém incorpora elementos que a diferenciam desta última. A codificadora é considerada híbrida, pois combina uma Rede Neural Convolucional (CNN) com um (ViT).

A arquitetura TransUNet, desenvolvida por (CHEN et al., 2021), utiliza uma ResNet50 (HE et al., 2016) para extrair características na fase de codificação, antes de passar para a etapa do Transformer. A saída de alguns níveis da CNN é então passada através de conexões de atalho (skip-connections) para os níveis correspondentes na decodificadora.

Devido à sua natureza híbrida, o mapa de características é usado como entrada para o transformer em vez da imagem original. Nesse cenário, os patches podem ter tamanho 1×1 .

Além disso, a arquitetura é composta por 12 camadas de transformers. Cada camada é formada por uma sequência de camada de normalização (LN), camada de atenção *multi-head* (MSA), outra camada de normalização e um perceptron multicamada (MLP). A operação da unidade do transformer na camada l pode ser definida pelas Equações 3.14 e 3.15.

$$z'_l = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1} \quad (3.14)$$

$$z_l = \text{MLP}(\text{LN}(z'_l)) + z'_l \quad (3.15)$$

A parte decodificadora consiste em blocos de *upsampling* em cascata, que têm como objetivo decodificar as características para gerar a imagem de saída no tamanho original da entrada. Cada bloco é composto por dois operadores de *upsampling*, seguidos por uma camada de convolução 3×3 e uma camada ReLU.

3.4.15 Árvores de decisão

As árvores de decisão podem ser compreendidas como um conjunto de regras que têm como finalidade separar dados em grupos ou prever um valor com base nas características e informações presentes. A nomenclatura “árvore” é devida ao processo pelo qual um dado é tratado, passando por um nó inicial que se ramifica em outros nós (LOH, 2011) (BREIMAN et al., 1984).

Para obter resultados mais robustos, é possível utilizar uma combinação de árvores. Uma técnica comumente empregada para essa finalidade é o *boosting*, que consiste em aprender o modelo de cada árvore sequencialmente, ajustando os dados e tratando os erros das árvores anteriores. Essa abordagem ajuda a melhorar a precisão do modelo final, ao mesmo tempo em que permite lidar com dados ruidosos e casos de *overfitting*.

3.4.16 XGBoost

O XGBoost (*eXtreme Gradient Boosting*) (CHEN; GUESTRIN, 2016) é um algoritmo de aprendizado de máquina supervisionado utilizado em árvores preditivas, baseado na técnica *Gradient tree boosting* (FRIEDMAN, 2001). Essa técnica consiste em utilizar a combinação de árvores de decisão fracas para formar um modelo robusto. Essas árvores de decisão fracas são modelos fracos, ou seja, apresentam resultados melhores somente do que a aleatoriedade.

O XGBoost pode ser usado tanto para tarefas de regressão quanto para classificação. A principal diferença entre o XGBoost e o *Gradient tree boosting* é a presença da regularização no primeiro, que melhora a generalização do modelo e reduz a probabilidade de *overfitting*. Outras vantagens que destacam o XGBoost são sua escalabilidade, extensibilidade, paralelismo e consumo de memória e recursos de hardware em relação a outros métodos (CHEN et al., 2015).

Neste trabalho, o XGBoost é usado como uma parte da etapa da classificação, utilizando a VGG-16 como extrator de características.

3.4.16.1 Parâmetros

O XGBoost possui hiperparâmetros que podem ser ajustados dependendo do objetivo para obtenção de melhores resultados. Mostram-se alguns destes na lista abaixo:

- Subamostragem : razão dos elementos usados no treinamento para cada iteração;
- Profundidade máxima: profundidade da árvore, quanto mais profunda, maiores são as chances de *overfitting*;
- Taxa de aprendizado: taxa de aprendizado durante o treinamento;
- Número de estimadores: número de árvores construídas no treinamento.

3.4.17 Active Learning

O *Active Learning* é empregado para selecionar dados importantes para anotação, visando aprimorar a capacidade de aprendizado do modelo. Geralmente, estes dados são mais difíceis de classificar e apresentam características mais significativas e únicas, gerando maior incerteza para o modelo de predição (REN et al., 2021; WAN et al., 2023). O processo começa com um conjunto inicial de dados rotulados para treinamento. O modelo resultante é aplicado sobre um conjunto de dados de teste não rotulados. Em seguida, os dados nos quais o modelo demonstra maior incerteza são revisados por um especialista, que os rotula. Por fim, esses dados são adicionados ao conjunto de treinamento. E, isto pode se repetir o quanto for necessário para se obter um modelo robusto (REN et al., 2021; BENGAR et al., 2022; WAN et al., 2023).

Em cada iteração, previsões são feitas no conjunto de teste após o treinamento. O nível de incerteza é avaliado usando a entropia, que é definida pela Equação 3.16, onde n é o número de classes no conjunto de dados e $p(y_i|x)$ é a probabilidade do dado x pertencer à classe y_i .

$$H(x) = - \sum_{i=1}^n p(y_i|x) \cdot \log_2(p(y_i|x)) \quad (3.16)$$

3.4.18 Data Augmentation

Devido à quantidade de imagens, há a possibilidade do método proposto não ser capaz de aprender a identificar corretamente as lesões. Assim, torna-se necessário o uso de *data augmentation*, abordagem para geração de novas imagens a partir de uma imagem existente, o que reduz a possibilidade de *overfitting* (YING, 2019). Esta técnica é realizada, normalmente, por meio de operações de transformação, tais como rotação, zoom e translação, as quais são aplicadas às imagens originais. Assim, o modelo consegue aprender com uma variedade maior de dados, aumentando sua capacidade de generalização.

3.5 Métricas de Avaliação

Nesta seção, são mostradas e explicadas as métricas utilizadas para a avaliação dos resultados de cada etapa do método desenvolvido.

As métricas de avaliação são geralmente extraídas a partir da análise da matriz de confusão, que organiza o resultado bruto obtido pelo método aplicado. Nesta, têm-se os valores de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos representados por vp , fp , vn , e fn , respectivamente.

A acurácia (ACC) é uma métrica que expressa o desempenho geral do modelo, avaliando quantas classificações foram realizadas corretamente dentre todas as possíveis. Em outras palavras, a acurácia mede a proporção de predições corretas feitas pelo modelo em relação ao número total de predições realizadas. É definida pela Equação 3.17.

$$ACC = \frac{vp + vn}{vp + fn + fp + fn} \quad (3.17)$$

A sensibilidade (SEN) mede o quão bom um modelo consegue identificar verdadeiros positivos (ou, pacientes que possuem endometriose) dentre todos os casos positivos. É definida pela Equação 3.18.

$$SEN = \frac{vp}{vp + fn} \quad (3.18)$$

Já a precisão (PREC) corresponde ao percentual de quantos são verdadeiros positivos dentre todos os classificados como positivos. Equação 3.19.

$$PREC = \frac{vp}{vp + fp} \quad (3.19)$$

A especificidade (SPE) representa a capacidade do modelo em classificar corretamente casos negativos (não possuem endometriose) dentre todos os realmente negativos. Equação 3.20.

$$ESP = \frac{vn}{vn + fp} \quad (3.20)$$

F1-score (F1) é a média harmônica entre a precisão e sensibilidade, podendo indicar uma melhor capacidade do modelo, e é definida pela Equação 3.21.

$$F1 = \frac{2 * SEN * PREC}{SEN + PREC} \quad (3.21)$$

O coeficiente Dice é uma das métricas usadas na avaliação de segmentação de

imagens, que mede o quão semelhantes são duas formas: o resultado da segmentação automática e a marcação real. O valor do Dice varia entre 0 e 1, onde 1 representa a semelhança completa entre as duas formas, e 0, formas totalmente diferentes. É calculado de acordo com a Equação 3.22, onde X e Y são os pixels da predição e do *ground truth*, respectivamente.

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3.22)$$

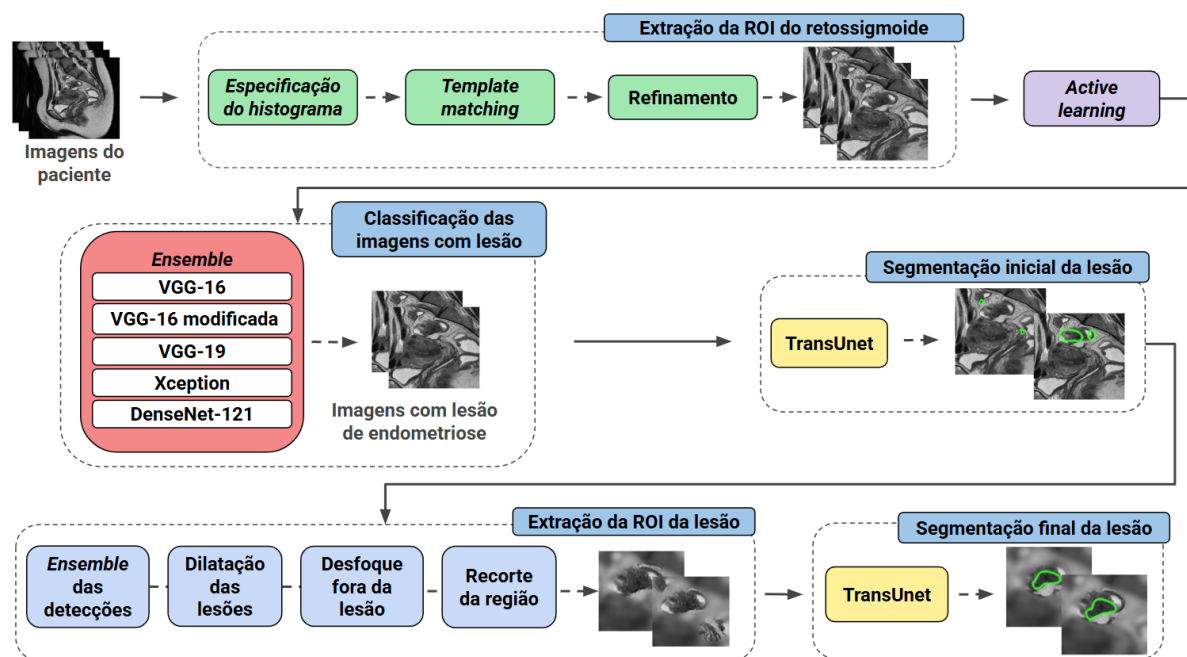
3.6 Considerações Finais

A fundamentação teórica do trabalho desenvolvido foi apresentada neste capítulo. Aqui, os conceitos necessários para a compreensão do método, a base e os elementos importantes do problema foram explicados. No próximo capítulo, os materiais utilizados e o método proposto por este trabalho serão apresentados e discutidos.

4 Método Proposto

O método proposto para classificar e segmentar lesões de endometriose em RM é apresentado neste capítulo. A primeira etapa do método, a extração da ROI do retossigmoide, define uma área de interesse que contém as duas regiões, ajudando a reduzir informações irrelevantes e otimizar o uso dos recursos computacionais. Em seguida, o *active learning* é aplicado para selecionar as imagens mais representativas para o treinamento das RNAs. A classificação indica em quais imagens do exame de um paciente há lesões, indicando se o paciente é considerado saudável. As prováveis regiões das lesões são determinadas através de uma etapa de segmentação inicial. Em seguida, realiza-se uma etapa para obter outra região de interesse onde estas se encontram. Uma última segmentação é empregada nessas áreas, a fim de identificar a área das lesões. O método é ilustrado na Figura 15.

Figura 15 – Fluxograma do método proposto.



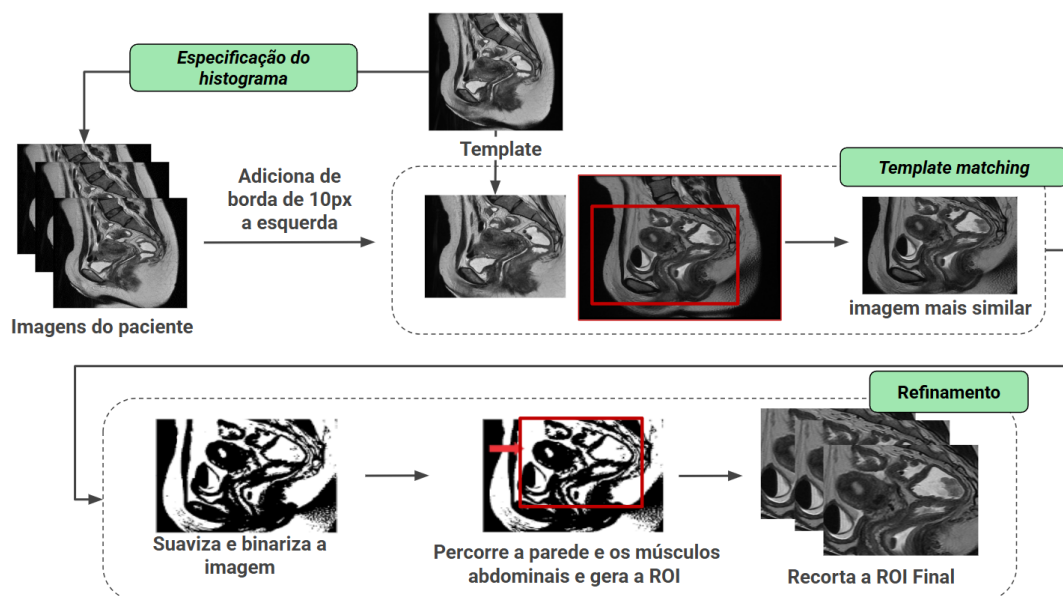
Fonte: O Autor

4.1 Extração da ROI do Retossigmoide

Identificar lesões de endometriose é uma tarefa desafiadora, até mesmo para radiologistas experientes, pois as lesões podem ter várias formas e texturas, além de se assemelharem a outras estruturas nas imagens de RM. Para evitar detecções fora da região pélvica e focar na área do retossigmoide, o método inclui uma etapa

especificamente dedicada à identificação dessas áreas. Essa etapa também contribui para reduzir o tamanho da imagem, otimizando os recursos computacionais nas etapas subsequentes. O fluxo do processo de extração dessa ROI é apresentado na Figura 16.

Figura 16 – Etapa de Identificação da ROI.



Fonte: O Autor

Para a realização desta etapa, selecionou-se uma imagem da base de imagens para servir como imagem de referência (Figura 17). Na imagem escolhida, é possível ver com clareza os órgãos da região pélvica e distingui-los devido à distribuição dos tons de cinza.

Figura 17 – Imagem de referência utilizada na especificação do histograma.

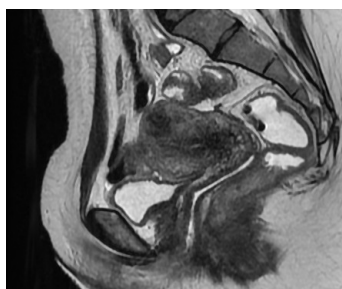


Fonte: O Autor

O processo de definição da ROI começa com a aplicação da especificação do histograma sobre todas as imagens do exame para que elas fiquem semelhantes à imagem de referência, principalmente aquelas nas quais a visibilidade da forma dos órgãos é parecida.

Em seguida, define-se um *template* para a aplicação do *template matching*. O *template* é um recorte da imagem escolhida como referência na subetapa anterior. O recorte usado como tal contém a área que abrange o retossigmoide. Embora a estrutura pélvica seja semelhante entre os pacientes, pequenas diferenças, como a disposição de alguns órgãos, o tamanho e as variações nos níveis de cinza, podem dificultar a operação entre as imagens. E essas diferenças podem resultar em perdas da região do retossigmoide. Por isso, adiciona-se informação externa à borda da região da pelve, incorporando a parede abdominal anterior na imagem. Com isso, adiciona-se uma borda de 10 pixels de largura à esquerda das imagens do exame, preenchida com valor zero (pixels pretos), para simular a região externa à área abdominal do *template*, já que essa região externa está ausente em uma minoria dos casos. Assim, tem-se uma identificação mais precisa da ROI. A Figura 18 apresenta o *template* utilizado.

Figura 18 – *Template* utilizado no *template matching*.



Fonte: O Autor

Utiliza-se a técnica *template matching* para percorrer todas as imagens de um exame e identificar a mais semelhante, definindo, conseqüentemente, a região correspondente. Durante a execução dessa técnica, a imagem do *template* é usada em tamanhos diferentes: o original e 102% do tamanho original. Esses tamanhos foram escolhidos com base em testes e na observação das imagens da base, considerando pequenas variações no tamanho do retossigmoide. Um tamanho menor poderia resultar em uma região inadequada. Os dois tamanhos de *template* são aplicados, e seleciona-se a imagem do exame que gerou maior similaridade entre o *template* e a região a ser detectada.

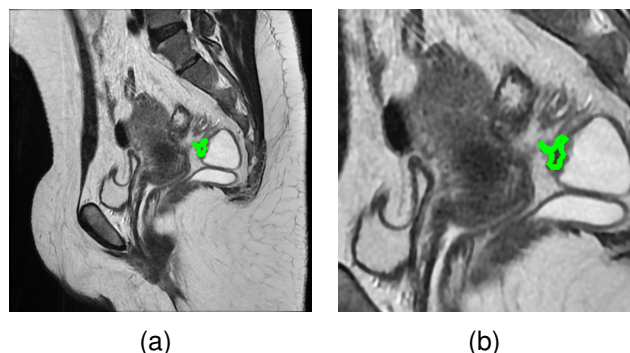
Após encontrar a imagem mais similar, aplica-se a subetapa de refinamento para remover áreas externas desnecessárias, como a região da parede abdominal, deixando apenas a região dos órgãos de interesse. Para isso, primeiramente, aplica-se uma suavização com o filtro gaussiano na imagem e realiza-se a binarização da imagem com o método de Otsu (OTSU et al., 1975) para diferenciar as áreas claras das escuras. Em seguida, percorre-se a imagem da esquerda para a direita, e de cima para baixo até metade da imagem, para encontrar o pixel preto mais à esquerda, que indica o início da parede abdominal. A partir deste ponto, percorre-se a imagem no eixo x , atravessando a

parede abdominal (pixels brancos) e o músculo abdominal (pixels pretos), até chegar à região interna da pelve onde os órgãos se encontram (pixels brancos).

A ROI final é definida por meio de recorte de 256x256 a partir do pixel mais superior à esquerda, englobando o reto e o sigmoide. Por fim, o recorte é aplicado nas mesmas coordenadas para todas as imagens do exame. A Figura 19 fornece um exemplo de resultado antes e depois da extração da ROI do retossigmoide. Dessa forma, o processo para a extração da ROI do retossigmoide para um exame é o seguinte:

1. Definição da imagem de referência e do *template*.
2. Especificação do histograma
 - a) Para cada imagem de um exame:
 - i. Ajustar histograma para se assemelhar ao *template*.
3. Adição de borda
 - a) Adicionar uma borda de 10 pixels de largura na esquerda com pixels pretos.
4. Aplicação do *template matching*
 - a) Para cada imagem:
 - i. Aplicar *template matching* usando o *template* original e uma versão com 102% do tamanho.
 - b) Selecionar a imagem que apresenta maior similaridade.
5. Refinamento
 - a) Aplicar filtro gaussiano para suavização
 - b) Aplicar binarização de Otsu para segmentar regiões claras e escuras.
 - c) Percorrer pixels da esquerda para a direita na metade superior até encontrar pixel preto mais a direita.
 - d) percorrer pelos pixels brancos (parede abdominal) no eixo x até encontrar um pixel preto (músculo abdominal).
 - e) percorrer pelos pixels pretos no eixo x até encontrar um pixel branco (região dos órgãos).
6. Definição da ROI
 - a) Determinar a ROI como uma região equivalente a 50% da imagem na altura e largura, a partir da coordenada x do pixel branco e altura baseada na região detectada pelo *template matching*.
 - b) Recortar a região definida sobre todas as imagens do exame.

Figura 19 – Exemplo de imagem antes e depois da extração da ROI do retossigmoide. (a) e (b) apresentam imagens de uma paciente antes e após a extração da ROI do retossigmoide, respectivamente. Marcação manual da endometriose em verde.



4.2 Active learning

Devido ao número limitado de imagens com endometriose disponíveis, à demanda de tempo para a marcação adequada de novas amostras e à singularidade de algumas lesões de endometriose, seja em sua forma ou textura, percebe-se que a seleção aleatória de imagens no conjunto de treinamento poderia não refletir a realidade. Assim, o modelo enfrentaria maior dificuldade com novos casos que apresentassem características distintas.

O *Active Learning* é empregado neste trabalho de forma diferenciada para contornar os possíveis problemas causados pela seleção aleatória. As principais diferenças são a exclusão do oráculo e a utilização de dados anotados no teste. Ou seja, os dados previamente anotados mais representativos são selecionados através do cálculo das incertezas para compor o conjunto de treinamento, permitindo o uso das amostras mais representativas e aprimorando a capacidade de aprendizado do modelo.

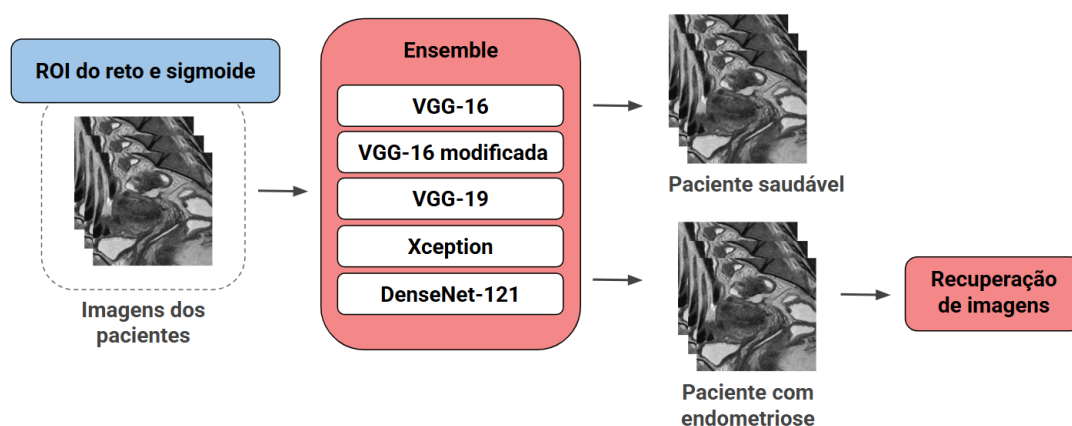
Neste trabalho, o conjunto de treinamento inicial é definido de forma aleatória e é expandido a cada iteração. O conjunto de validação, também selecionado aleatoriamente, permanece fixo ao longo do processo. Todos os demais pacientes são incluídos no conjunto de teste. Em cada iteração, são feitas previsões no conjunto de teste. Os níveis de incerteza são calculados usando entropia.

Após o cálculo da incerteza das imagens de teste, são selecionadas as cinco imagens com maior incerteza para cada classe, doente ou saudável. A partir do nome de cada imagem selecionada, o respectivo paciente é identificado e todas as suas imagens são transferidas para o conjunto de treinamento. Posteriormente, o modelo é treinado novamente usando este conjunto de treinamento atualizado até o critério de parada estabelecido, com aproximadamente 20% das imagens para cada classe no teste.

4.3 Classificação das Imagens com Lesão

O objetivo da classificação é determinar quais pacientes apresentam lesões e em quais imagens essas lesões podem ser encontradas (Figura 20). Para alcançar esse objetivo, é empregada uma abordagem de *ensemble* de redes, que combina múltiplos modelos para classificação usando um processo de média da probabilidade dos modelos (GANAIÉ et al., 2022). Um paciente é considerado afetado pela doença se duas ou mais imagens forem identificadas como possuindo lesão de endometriose. Isso melhora o desempenho e reduz a variância entre os modelos (JU; BIBAUT; LAAN, 2018). As arquiteturas de redes neurais para esta tarefa incluem VGG-16, DenseNet-121, VGG-19, Xception e uma versão modificada de VGG-16, proposta neste trabalho.

Figura 20 – Etapa de Classificação.



Fonte: O Autor

Na versão modificada da VGG-16, as camadas densas são substituídas por Camadas Convolucionais Recorrentes (RCL) (LIANG; HU, 2015). Essa arquitetura possibilita o fluxo de informações entre as camadas, reduzindo o número de parâmetros do modelo, e permite a construção de uma rede mais profunda. A presença de trajetórias mais longas na RCL permite que o modelo aprenda características complexas do contexto ao redor (LIANG; HU, 2015; ALOM et al., 2019). Essa arquitetura pode ser vista na Figura 21.

A rede é inicializada com pesos pré-treinados da ImageNet na parte da VGG16 e, em seguida, é ajustada (treinada) para a tarefa específica, descongelando todas as camadas. Ela é utilizada como extratora de características até a camada totalmente conectada, gerando representações que servem como entrada para um classificador XGBoost. Além de mitigar o problema do *overfitting*, o XGBoost, quando combinado com redes neurais como extratores de características, apresentou resultados promissores (LIEW; HAMEED; CLOS, 2021), superando o desempenho da rede neural utilizada

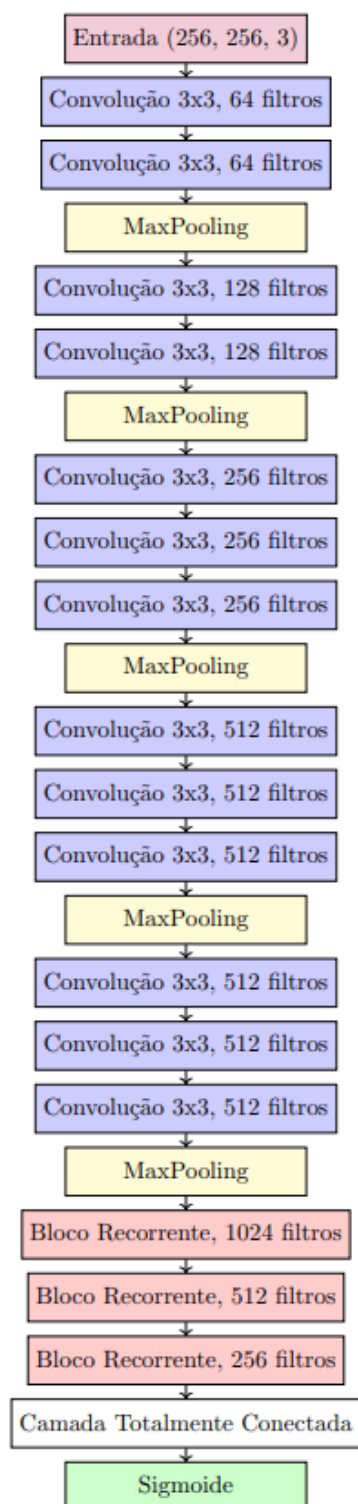


Figura 21 – VGG-16 modificada

isoladamente, conforme relatado na literatura (RAHMAN et al., 2021). O XGBoost é então treinado utilizando essas características para classificar as imagens.

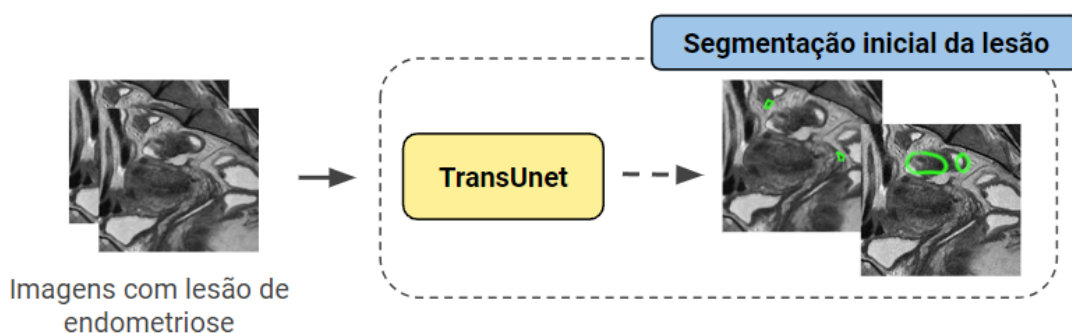
As demais redes também são inicializadas com pesos pré-treinados. Ao final, é realizado o *ensemble* das redes, e o paciente é classificado como doente ou saudável. Após a definição dos pacientes com endometriose, busca-se identificar mais imagens que

possam conter lesões associadas à doença. Para isso, os vizinhos imediatamente mais próximos de cada imagem previamente classificada como apresentando lesões de endometriose também são classificados dessa forma. Adicionalmente, classificam-se as 10 imagens mais próximas à esquerda e à direita cuja probabilidade, determinada pelo *ensemble*, seja superior a 0,4 como doentes. Este valor de probabilidade foi escolhido através da avaliação dos resultados, considerando que ele ainda pode indicar a presença de endometriose, dada a dificuldade de classificação da doença.

4.4 Segmentação Inicial da Lesão

A identificação da área lesionada pela endometriose é crucial tanto para monitorar a progressão da doença quanto para auxiliar no planejamento cirúrgico, caso necessário. A TransUnet é utilizada para gerar uma segmentação inicial das lesões nas imagens de endometriose a partir da etapa de classificação, com o objetivo de identificar possíveis áreas afetadas. A TransUnet se beneficia dos transformers para extrair contexto global, facilitando a obtenção de informações sobre as áreas ao redor da lesão (DOSOVITSKIY et al., 2021; CHEN et al., 2021). Além disso, a arquitetura em forma de “U” da U-Net, juntamente com as skip-connections, permite a captura de contexto e localização (RONNEBERGER; FISCHER; BROX, 2015). A TransUnet demonstrou melhores resultados em tarefas de segmentação multi-órgão e cardíaca (CHEN et al., 2021) em comparação com arquiteturas como V-Net (MILLETARI; NAVAB; AHMADI, 2016), ResNet U-Net (RONNEBERGER; FISCHER; BROX, 2015) e ResNet Attention U-Net (SCHLEMPER et al., 2019). O método é representado na Figura 22, e um exemplo de saída da etapa é mostrado na Figura 23.

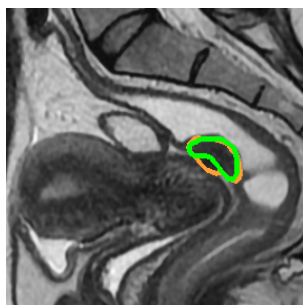
Figura 22 – Etapa de segmentação inicial.



Fonte: O Autor

A partir dos experimentos realizados, observou-se que ainda há perda de partes da lesão, além de dificuldades na identificação de lesões pequenas, ou até mesmo uma superestimação delas. Com base nessa análise, novas etapas foram adicionadas para refinar o processo de segmentação.

Figura 23 – Ilustração da segmentação inicial. Marcação do especialista em verde, e segmentação da etapa de segmentação inicial em laranja.



Fonte: O Autor

4.5 Extração da ROI da Lesão

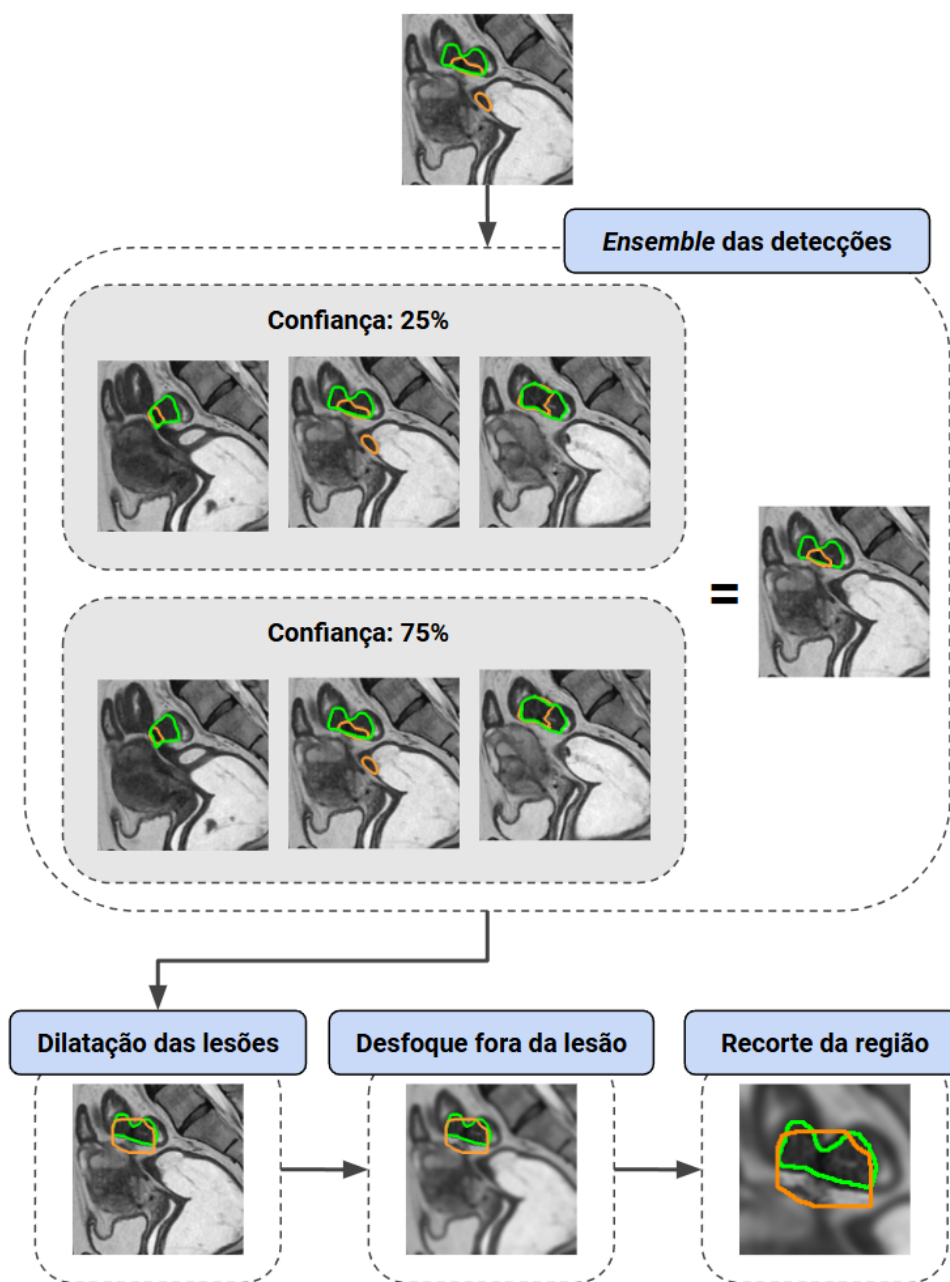
Como a segmentação inicial pode, em alguns casos, perder parte da lesão ou até mesmo não detectar a lesão em algumas imagens, desenvolveu-se uma etapa para aproveitar ao máximo as informações encontradas na fase anterior, antes de enviar a imagem para o refinamento da segmentação da lesão. Essa etapa é chamada de extração da ROI da lesão (Figura 24).

Para extrair a região da lesão na imagem processada, utiliza-se uma abordagem de *ensemble* das máscaras de predição por meio de votação majoritária. Essa votação considera as máscaras geradas para a imagem anterior, a imagem atual e a imagem posterior, utilizando predições com níveis de confiança de 25% e 75%, para cada imagem. No caso de existir apenas duas imagens, considera-se a anterior ou a posterior.

Após obter a máscara resultante da votação, aplica-se uma dilatação, utilizando um elemento estruturante quadrado de 30×30 , para expandir ligeiramente a área, capturando informações próximas à lesão ou até mesmo partes dela não detectadas. Em seguida, aplica-se um filtro gaussiano à imagem original, restrito à região externa à máscara dilatada com elemento estruturante de 15×15 . Isso reduz a nitidez nas áreas fora da máscara sem que haja muita perda de informação de localização e contexto, e tenta evitar que essas regiões sejam interpretadas como lesões na etapa seguinte.

Posteriormente, é criada uma *bounding box* ao redor da lesão dilatada para garantir a captura de informações contextuais. Caso alguma dimensão da *bounding box* resultante seja menor que 128 pixels, ela é ajustada para atingir esse valor mínimo. Por fim, a *bounding box* é ajustada para formar um quadrado, tomando como base a maior dimensão, e é então utilizada como entrada para a etapa de segmentação da lesão.

Figura 24 – Representação da etapa de extração da ROI da lesão com exemplo real. As marcações manuais da lesão de endometriose estão em verde, e as automáticas, em laranja.



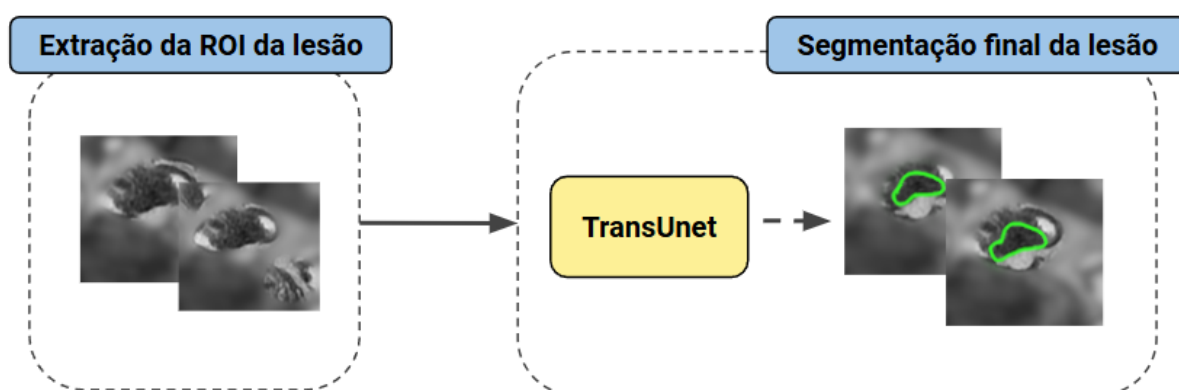
Fonte: O Autor

4.6 Segmentação Final da Lesão

A segmentação final da lesão é a última etapa do método. Esta segunda fase de segmentação pode ser vista como um refinamento, com o objetivo de aprimorar os resultados da primeira segmentação realizada. Com isso, obtém-se uma representação mais precisa do tamanho e dos limites da lesão, permitindo ao especialista avaliar melhor a gravidade da doença.

Devido aos resultados obtidos na primeira etapa de segmentação e às vantagens de sua arquitetura, mencionadas anteriormente, a TransUnet é utilizada novamente para segmentar a lesão. Nesta fase, a entrada consiste em uma imagem menor, contendo menos áreas irrelevantes à lesão, extraídas na etapa anterior. Essa abordagem assegura que a TransUnet possa concentrar sua atenção de forma mais refinada na ROI. A Figura 25 ilustra esta etapa.

Figura 25 – Etapa de segmentação da endometriose.



Fonte: O Autor

4.7 Considerações Finais

Este capítulo apresentou o método desenvolvido para a classificação de pacientes, a segmentação das lesões de endometriose e o uso do *active learning*, detalhando as etapas do processo e ressaltando a importância de cada uma delas. A aplicação dessas etapas na ordem apresentada visa reduzir falsos positivos, obter melhor localização da lesão e otimizar o uso dos recursos computacionais. O capítulo seguinte trata dos resultados alcançados e da discussão sobre estes.

5 Resultados e Discussões

Neste capítulo, são apresentados os resultados obtidos para cada etapa do método proposto e a base de imagens utilizada. Para a realização dos testes com o método, emprega-se a linguagem de programação Python 3.7 em conjunto com bibliotecas de processamento de imagem e estruturas de aprendizado profundo como XGBoost, Keras, Tensorflow e OpenCV. Os experimentos foram realizados em Linux e Windows.

Quanto ao hardware, foram utilizados dois computadores de alto desempenho para a realização dos experimentos. Um dos computadores possui uma CPU Intel Core i7-6700HQ com velocidade de 2 GHz, 16 GB de memória RAM e uma placa de vídeo Nvidia GeForce GTX 1070. O outro computador conta com 32 GB de RAM, CPU Intel Core i7-7700K com velocidade de 4,20 GHz e uma placa de vídeo Nvidia GeForce GTX 1080 Ti.

Para a avaliação quantitativa dos resultados, utilizou-se a acurácia (ACC), sensibilidade (SEN), precisão (PREC), especificidade (ESP), F1-score (F1) e Dice.

As funções de *loss* (ou perda, em português) avaliadas foram: Focal ([LIN et al., 2017](#)), Tversky ([SALEHI; ERDOGMUS; GHOLIPOUR, 2017](#)), Dice ([MILLETARI; NAVAB; AHMADI, 2016](#)) e a combinação BCE e Tversky (BCETversky). Estas funções foram escolhidas por tratarem bases de dados desbalanceadas.

Para as etapas que utilizam aprendizado profundo utilizou-se a técnica de *early stopping* ([BROWNLIE, 2018](#)), que consiste em interromper o treinamento da rede quando a métrica monitorada deixa de melhorar. Também foi empregado o decaimento da taxa de aprendizado de 10%, seguindo o mesmo princípio. Ambas monitoram o valor obtido na função de *loss*, usada para quantificar a perda entre o valor predito e o valor verdadeiro.

5.1 Base de Imagens

Para este estudo, utilizou-se uma base de dados privada composta por imagens de ressonância magnética ponderadas em T2 (RM) da pelve feminina. O conjunto de dados foi desenvolvido em parceria com a Clínica Fonte Imagem Medicina Diagnóstica e o Departamento de Ginecologia do Hospital Universitário Pedro Ernesto, localizado no Rio de Janeiro, Brasil. As imagens foram adquiridas utilizando um aparelho de ressonância magnética SIGNA™ Artist 1.5T da GE Healthcare.

A base de imagens inclui 179 exames de ressonância magnética, sendo 107 com

lesões de endometriose e 72 considerados saudáveis. O conjunto de dados abrange exames com lesões de endometriose profunda, com tamanho mínimo de dois centímetros, localizadas na região do retossigmoide. Todos os exames presentes nesta base estão totalmente anonimizados, garantindo a proteção da privacidade dos pacientes.

Cada exame na base de imagens consiste em aproximadamente 32 fatias de ressonância magnética no formato DICOM, resultando em um total de 5.730 imagens. Destas imagens, 452 estão anotadas com a presença da doença. Cada imagem possui dimensões de 512×512 pixels. A Figura 26 ilustra um exemplo de imagem da base, com a marcação verde indicando a área afetada pela endometriose, conforme marcado por um especialista.

Figura 26 – Exemplo de imagens da base marcada. Marcação da lesão em verde.

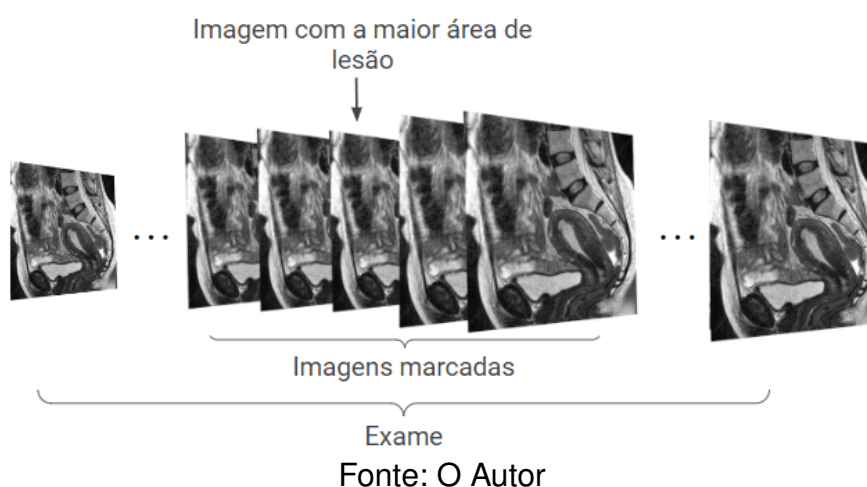


5.1.1 Protocolo de Marcação das Imagens

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 também são marcadas. Porém, nem sempre é possível marcar essas cinco imagens devido à visibilidade da lesão; da mesma forma, às vezes, é viável marcar mais imagens. Dessa forma, algumas imagens podem ter mais ou menos marcações do que outras. A Figura 27 ilustra o protocolo de marcação.

O trabalho de marcação das lesões foi realizado por três especialistas: um com mais de 25 anos de experiência, e os outros dois com aproximadamente três anos de experiência no diagnóstico e tratamento de endometriose. Cada marcação feita por um especialista foi revisada e avaliada por outro.

Figura 27 – Representação do protocolo de marcação das imagens.



5.1.2 Preparação da Base de imagens

Para todas as etapas que exigem o treinamento de alguma RNA, as imagens foram divididas em três conjuntos: treino, validação e teste. Todas as imagens de um mesmo exame foram atribuídas exclusivamente a um dos conjuntos, a fim de garantir que não haja viés nos resultados.

5.2 Extração da ROI do Retossigmoide

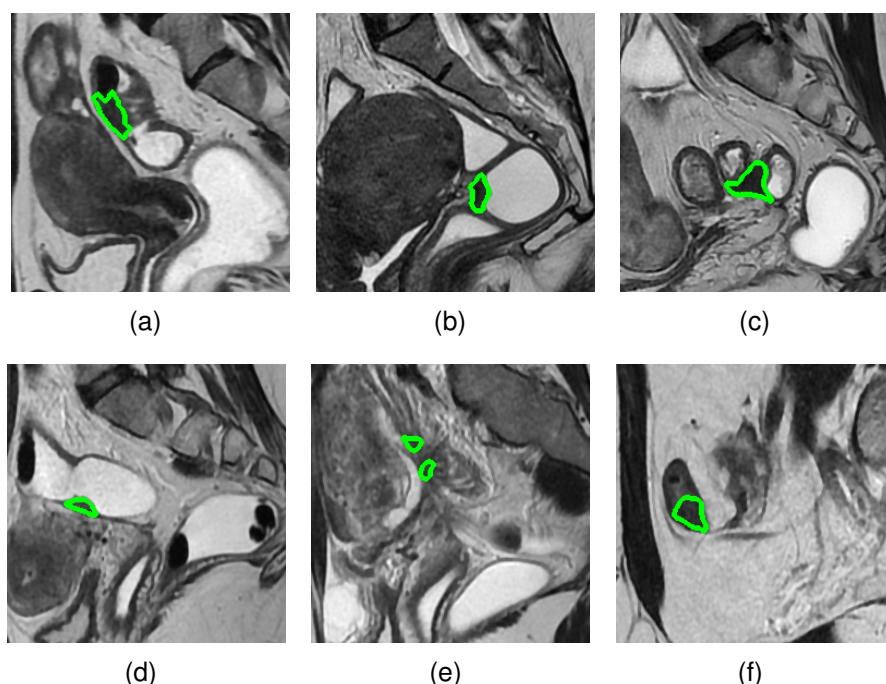
O objetivo principal da etapa atual do processo é minimizar a ocorrência de falsos positivos fora da região da pelve, sem comprometer a detecção da área do retossigmoide afetada pela lesão. Para avaliar essa fase, foi realizada uma análise de sensibilidade focada nas lesões presentes nas imagens. Esse procedimento garantiu que todas as lesões fossem corretamente incluídas dentro da região de interesse, resultando em uma sensibilidade de 100%. A Figura 28 mostra exemplos de resultados obtidos nesta etapa.

5.3 Active Learning

Como a base de imagens foi ampliada durante o desenvolvimento, a seleção das imagens para treinamento com Active Learning foi realizada quando a base continha 156 pacientes, sendo 72 saudáveis e 86 com endometriose. Nas etapas seguintes, toda a base foi utilizada, e os novos exames anotados foram incorporados ao treinamento.

A primeira iteração do treinamento com o active learning foi realizada com um pequeno conjunto de pacientes, selecionados aleatoriamente da base de imagens, sendo 20% saudáveis e 20% diagnosticados com endometriose. A mesma quantidade é usada no conjunto de validação, e os demais no conjunto de teste. A arquitetura utilizada

Figura 28 – Exemplos de resultados da extração da ROI do retossigmoide. (a), (b), (c), (d), (e) e (f) apresentam exemplos de imagens de pacientes distintos. Marcação manual da lesão em verde.



foi a VGG-16 modificada. O treinamento é concluído quando há 18 pacientes com endometriose e 12 pacientes saudáveis (20% de cada classe) no conjunto de teste.

Antes de utilizar o *active learning*, foram realizados testes preliminares com a seleção aleatória dos conjuntos de treinamento, validação e teste. Dos 158 pacientes, 18 com endometriose e 12 saudáveis foram utilizados para o conjunto de teste. Nos experimentos, empregaram-se tanto a arquitetura VGG-16 original quanto uma versão modificada proposta neste trabalho, VGG-16 MOD. Os resultados obtidos com a seleção aleatória e com o *active learning*, tanto por imagem quanto por paciente, são apresentados nas Tabelas 2 e 3, respectivamente, onde AL indica os resultados do *active learning*. A seleção automática apresentou os melhores resultados, tanto por imagem quanto por paciente, em comparação com a seleção aleatória.

Tabela 2 – Resultados da classificação de imagens com e sem *active learning* (AL)

Modelo	ACC	SEN	SPE	PREC	F1
VGG-16	0,797	0,795	0,797	0,285	0,419
VGG-16 MOD	0,836	0,443	0,876	0,267	0,333
VGG-16 (AL)	0,808	0,943	0,797	0,270	0,420
VGG-16 MOD (AL)	0,874	0,771	0,882	0,342	0,474

Com os resultados apresentados, o uso do *active learning* mostra-se promissor para selecionar automaticamente imagens para o conjunto de treino, conseguindo identificar imagens mais representativas para a construção do modelo.

Tabela 3 – Resultados da classificação de pacientes com e sem *active learning* (AL)

Modelo	ACC	SEN	SPE	PREC	F1
VGG-16	0,567	0,889	0,083	0,593	0,711
VGG-16 MOD	0,7	0,778	0,583	0,737	0,757
VGG-16 (AL)	0,767	0,944	0,4167	0,72	0,837
VGG-16 MOD (AL)	0,767	0,944	0,5	0,739	0,829

5.4 Classificação das Imagens com Lesão

Experimentos foram realizados utilizando diversas arquiteturas de redes neurais para classificar pacientes com endometriose e identificar imagens de exames contendo lesões. Isso é feito por meio de um *ensemble* de cinco redes neurais.

Para a realização dos testes do método, o conjunto de teste definido pelo *active learning* tornou-se o conjunto de validação, enquanto o conjunto de validação passou a ser utilizado como conjunto de teste. Isso ocorre porque, no conjunto de validação, não houve nenhum tipo de seleção automática, o que evita possíveis vieses e garante que ele represente um conjunto sem padrões pré-definidos.

Para a construção do conjunto de treinamento na etapa de classificação, foram removidas as imagens anteriores e posteriores às marcações dos especialistas em cada exame de pacientes com endometriose. Esse processo foi adotado devido ao protocolo de marcação utilizado, no qual algumas imagens não marcadas poderiam conter lesões ou partes delas, o que poderia comprometer o treinamento adequado do modelo.

Para exames com mais de quatro imagens marcadas, foram removidas duas imagens anteriores e duas posteriores a cada marcação. Já para exames com menos de quatro imagens marcadas, as três anteriores e posteriores foram removidas. Um exemplo desse processo, aplicado a um exame com cinco marcações, é ilustrado na Figura 29. Adicionalmente, algumas imagens de pacientes saudáveis foram removidas aleatoriamente para reduzir a diferença entre as duas classes.

Para o conjunto de validação, foram utilizadas todas as imagens de pacientes saudáveis, bem como as imagens de pacientes com endometriose — sejam saudáveis ou não — provenientes do processo descrito anteriormente. No entanto, nenhum processo de remoção foi aplicado ao conjunto de teste a fim de representar dados de entrada reais. A Tabela 4 apresenta a distribuição das imagens entre os conjuntos de treino, validação e teste.

A definição dos hiperparâmetros foi realizada por meio de *Grid Search* e pela análise visual da evolução do treinamento ao longo das épocas para cada rede utilizada. A Tabela 5 apresenta os parâmetros avaliados, enquanto a Tabela 6 exibe os hiperparâmetros definidos para cada rede. A escolha dos parâmetros foi baseada na

Figura 29 – Exemplo do processo de construção do conjunto de treinamento na classificação de imagens da lesão para um exame.



Fonte: O Autor

Tabela 4 – Distribuição de imagens nos conjunto de treino, validação e teste

Conjunto de Dados	Saudável	Endometriose
Treino	1117	299
Validação	777	70
Teste	882	83

menor *loss* obtida durante o treinamento. Os parâmetros para cada rede são mostrados na Tabela 6. Os parâmetros da XGBoost são mostrados na Tabela 7. Para o treinamento, utilizou-se uma paciência de 25 épocas para o *early stopping* e 15 épocas para o decaimento da taxa de aprendizado, com monitoramento da função de *loss* ao longo do processo.

Tabela 5 – Parâmetros de função de perda, tamanho do *Batch*, taxa de aprendizado e otimizador avaliados na classificação.

Função de Perda	Batch	Taxa de Aprendizado	Otimizador
BCETversky	8	0,1	Adam
Tversky	16	0,01	SGD
Focal	32	0,001	Adadelta
-	-	0,0001	-

Tabela 6 – Parâmetros de função de perda, tamanho do *Batch*, taxa de aprendizado e otimizador selecionados por meio de *Grid Search* para cada rede na classificação.

Modelo	Função de Perda	Batch	Taxa de aprendizado	Otimizador
VGG-16	Focal	32	0,001	Adadelta
VGG-19	Focal	32	0,001	Adadelta
DenseNet-121	Focal	32	0,01	SGD
Xception	Focal	32	0,1	SGD
VGG-16 MOD	Focal	32	0,01	SGD

Tabela 7 – Parâmetros do modelo XGBoost.

Parâmetro	Valor
Taxa de aprendizado	0,03
Profundidade máxima	5
Número de estimadores	225
Subamostragem	1

Para a classificação também foi utilizado o aumento durante a execução do treinamento (*on-the-fly*, no inglês). A Tabela 8 lista as operações empregadas e seus valores. Este foi definido a partir da análise de experimentos preliminares e da base de imagens, a fim de não causar grandes distorções nas imagens. As operações são executadas de forma aleatória e de acordo com os valores presentes na tabela.

Tabela 8 – Parâmetros de aumento de dados para a classificação

Tipo de Aumento de Dados	Valor
Rotação	20° a 20°
Deslocamento Horizontal	-0,2 a 0,2%
Deslocamento Vertical	-0,1 a 0,1%
Brilho	0,95% a 1,05%
Espelhamento Horizontal	Falso ou verdadeiro
Cisalhamento	-5° a 5°
Zoom	0,9% a 1,05%

A Tabela 9 apresenta os resultados obtidos na classificação por imagem, utilizando o modelo referente à época com o maior valor de F1-score no conjunto de validação durante o processo de treinamento. Deste modo, haveria a possibilidade de atingir um F1-score maior também no conjunto de teste. Nessa tabela, observa-se que os resultados variam entre as redes avaliadas. Quando a abordagem de *ensemble* foi empregada, houve uma melhoria nos resultados, atingindo valores mais altos de F1-score e precisão. Já na Tabela 10, são apresentados os resultados da classificação por paciente, considerando um paciente como doente caso ao menos duas de suas imagens sejam classificadas como doentes. Assim como na classificação por imagem, a abordagem de *ensemble* também se destacou, proporcionando os melhores resultados. As matrizes de confusão do resultado com *ensemble* para imagem e paciente são mostradas nas Tabelas 11 e 12, respectivamente.

Tabela 9 – Métricas para a classificação por imagem.

Modelo	ACC	SEN	PREC	ESP	F1
VGG16	0,914	0,6024	0,5	0,9433	0,5464
VGG19	0,8912	0,6506	0,4154	0,9138	0,507
DenseNet121	0,9026	0,5422	0,4455	0,9365	0,4891
Xception	0,8839	0,5181	0,3739	0,9184	0,4343
VGG Mod	0,8725	0,7711	0,381	0,8821	0,51
<i>Ensemble</i>	0,9223	0,7108	0,5364	0,9422	0,6114

Tabela 10 – Métricas para a classificação por paciente.

Rede	ACC	SEN	PREC	ESP	F1
VGG16	0,8	0,8889	0,8	0,6667	0,8421
VGG19	0,8	1,0	0,75	0,5	0,8571
DenseNet12	0,71	0,7778	0,7368	0,5833	0,7568
Xception	0,8333	1,0	0,7826	0,5833	0,878
VGG Mod	0,7	1,0	0,6667	0,25	0,8
<i>Ensemble</i>	0,9333	1,0	0,9	0,8333	0,9474

Tabela 11 – Matriz de confusão da classificação por imagem para o *Ensemble*.

	Endometriose	Saudável
Endometriose	59	24
Saudável	51	831

Tabela 12 – Matriz de confusão da classificação por paciente para o *Ensemble*.

	Endometriose	Saudável
Endometriose	18	0
Saudável	2	10

Ao utilizar a estratégia de recuperação de imagens vizinhas às classificadas inicialmente como possuindo endometriose, foi possível recuperar mais imagens e aumentar a sensibilidade da classificação por imagens. Porém, com isso, algumas imagens saudáveis são classificadas como doentes. Assim, 18 imagens de pacientes saudáveis foram classificadas como tendo endometriose. Estas imagens serão processadas pelas etapas de segmentação inicial, extração de ROI e segmentação final da lesão. E as imagens falso positivas (ou saudáveis classificadas como doente) podem ser filtradas durante o restante do método. A matriz de confusão final da classificação por imagem é observada na Tabela 13.

5.5 Segmentação Inicial da Lesão

Tabela 13 – Matriz de confusão da classificação por imagem final para o *Ensemble*.

	Endometriose	Saudável
Endometriose	75	8
Saudável	103	779

Na base utilizada para a segmentação inicial, além de todas as imagens com endometriose, foram selecionadas aleatoriamente algumas imagens de pacientes saudáveis que apresentavam as regiões do retossigmoide visíveis. Essa divisão foi motivada pela quantidade de lesões presentes e pela proporção em pixels entre a área de interesse (lesão) e o fundo (tecido saudável). A distribuição de imagens é apresentada na Tabela 14. Adicionalmente, a segmentação inicial da lesão foi avaliada de três modos:

- I. Considerando os resultados provenientes diretamente da classificação;
- II. Considerando os resultados da classificação, mas com a remoção de imagens classificadas como saudáveis de pacientes diagnosticados com endometriose, pois, devido ao protocolo de marcação, essas imagens poderiam conter lesões não identificadas;
- III. O caso ideal, no qual apenas imagens com endometriose são encaminhadas para essa etapa.

Tabela 14 – Distribuição de imagens nos conjunto de treino, validação e teste

Conjunto de Dados	Quantidade de imagens
Treino	366
Validação	85
Teste (modo I)	178
Teste (modo II)	92
Teste (modo III)	83

Assim como na classificação das imagens com lesão, foi utilizado o aumento de dados durante o treinamento. As operações realizadas e seus respectivos valores podem ser observados na Tabela 15. Esses valores foram definidos de forma a evitar alterações excessivas nas imagens, garantindo a preservação das características relevantes.

A TransUnet foi selecionada para a tarefa devido às suas vantagens em relação a outras arquiteturas. Os melhores hiperparâmetros foram definidos a partir dos valores apresentados na Tabela 16. Para esta etapa, os parâmetros selecionados incluem um tamanho de *batch* de 2, a função de *loss* BCETversky, uma taxa de aprendizagem de 0,0001 e o otimizador Adam. O treinamento contou com uma paciência de 20 épocas para o *early stopping* e 15 épocas para o decaimento da taxa de aprendizado, com monitoramento da função de *loss* ao longo do processo.

Tabela 15 – Operações de aumento de dados para a segmentação inicial da lesão

Tipo de Aumento de Dados	Valor
Rotação	30°
Deslocamento Horizontal	0,15%
Deslocamento Vertical	0,15%
Cisalhamento	0,02%
Zoom	0,8%-1,2%

Tabela 16 – Parâmetros de função de perda, tamanho do *Batch*, taxa de aprendizado e otimizador para segmentação

Função de <i>loss</i>	Batch	Taxa de Aprendizado	Otimizador
Dice	2	0,01	Adam
Tversky	4	0,001	SGD
Focal	8	0,0001	Adadelta
BCETversky	16	0,00001	-

Os resultados obtidos pela TransUNet foram comparados com os da U-Net utilizando os *backbones* VGG-16 e DenseNet121 com pesos pré-treinados, visando validar a escolha da arquitetura. Conforme apresentado na Tabela 17, a TransUNet demonstrou desempenho superior nas métricas relacionadas aos pixels da lesão. O *backbone* VGG-16 registrou o menor valor de Dice, enquanto a DenseNet121, apesar de apresentar um valor de Dice superior ao da VGG-16, mostrou uma perda de sensibilidade maior em relação às outras arquiteturas avaliadas.

Tabela 17 – Métricas de desempenho para diferentes modelos de segmentação

Modelo	SEN	PREC	ESP	DICE
TransUNet	0,6458	0,5376	0,9937	0,5360
UNet + VGG-19	0,6162	0,4378	0,9916	0,4501
UNet + DenseNet121	0,5905	0,5110	0,9950	0,5083

Devido aos experimentos indicarem dificuldades na identificação de lesões, especialmente as menores, testou-se alterar o tamanho das imagens usadas na entrada da rede para avaliar o impacto do redimensionamento na identificação e segmentação das lesões. A Tabela 18 apresenta a comparação dos resultados obtidos pela TransUNet com diferentes tamanhos de entrada. Observou-se que o aumento do tamanho da imagem resultou em uma melhora nos valores de Dice e em uma redução na perda de detecção das lesões. Com imagens de tamanho 256×256 , a rede não conseguiu identificar lesões em 12 imagens. Já com o tamanho 288×288 , esse número foi reduzido para sete.

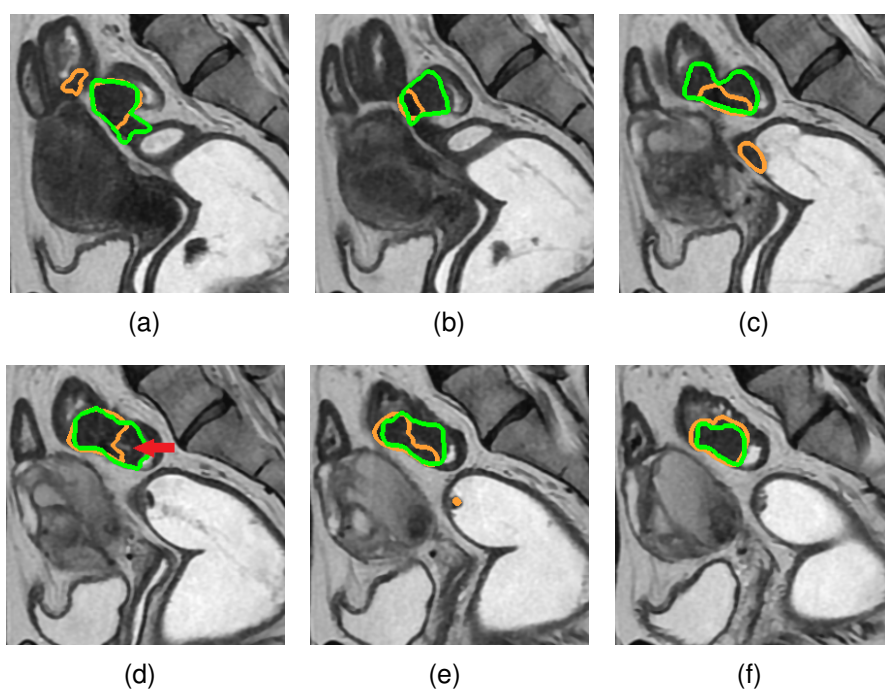
A Figura 30 apresenta um exemplo de segmentação inicial referente à paciente 22. Nela, observa-se a ocorrência de falsos positivos em áreas distintas do retossigmoide, algumas localizadas em posições próximas em diferentes cortes de imagem. Além disso, é evidente uma subestimação da lesão. Na Figura 30(d), destaca-se uma região

Tabela 18 – Métricas de desempenho para o TransUNet com diferentes tamanhos de imagem

Modelo	Tamanho da Imagem	SEN	PREC	ESP	DICE
TransUNet	224 × 244	0,6346	0,5192	0,9921	0,5172
TransUNet	256 × 256	0,6458	0,5376	0,9937	0,5360
TransUNet	288 × 288	0,6917	0,6230	0,9953	0,6199
TransUNet	320 × 320	0,6707	0,5715	0,9949	0,5917

esbranquiçada na parte da lesão que não foi segmentada, indicada pela seta vermelha, o que possivelmente contribuiu para a falha na identificação dessa área.

Figura 30 – Exemplo da segmentação inicial da lesão de endometriose para a paciente 22. (a), (b), (c), (d), (e) e (f) exemplos de imagens do paciente. As marcações manuais estão em verde, e as automáticas, em laranja.



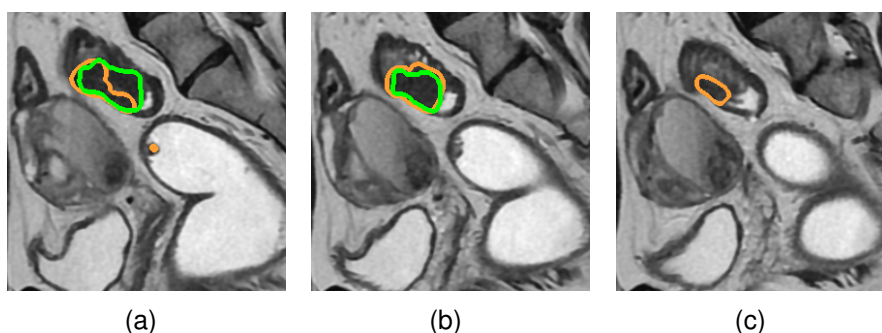
Os resultados obtidos para cada modo de avaliação podem ser observados na Tabela 19. As imagens saudáveis têm um impacto significativo nas métricas no modo I, especialmente quando comparadas ao modo II, que apresenta resultados semelhantes ao modo III, o caso ótimo. Esse fato pode indicar a presença de partes das lesões em imagens não marcadas, como consequência do protocolo de marcação. Na Figura 31, é apresentado um caso em que as duas imagens anteriores possuíam marcação de lesão, e uma lesão foi identificada, aproximadamente, na mesma posição. Casos como esse devem ser avaliados por um especialista.

Das 18 imagens classificadas como pertencentes a pacientes saudáveis, 10 exibiram segmentações indicando a presença de uma possível lesão, apesar da ausência da doença. A Figura 32 apresenta um exemplo, destacando o resultado de segmentação

Tabela 19 – Métricas de desempenho para o TransUNet para diferentes modos de avaliação

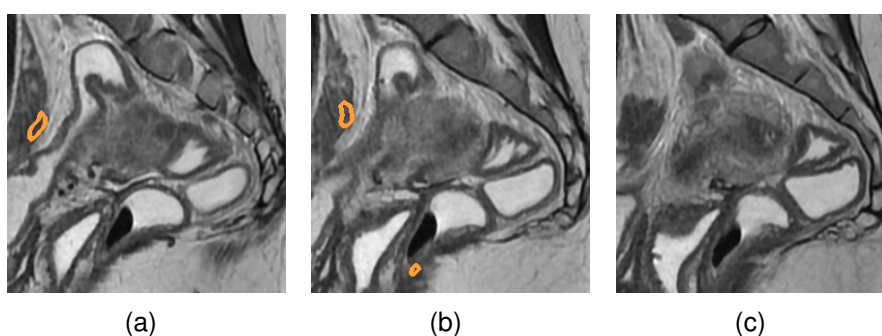
Modelo	Modo de avaliação	SEN	PREC	ESP	DICE
TransUNet	I	0,5119	0,4798	0,9958	0,4785
TransUNet	II	0,6503	0,5918	0,9955	0,5889
TransUNet	III	0,6917	0,6230	0,9953	0,6199

Figura 31 – Exemplo de erro segmentação inicial da lesão de endometriose para a paciente 22. (a) e (b) mostram caso de acerto e (c) mostra um caso de erro das imagens da paciente. As marcações manuais estão em verde, e as automáticas, em laranja.



obtido para uma paciente saudável. A análise deste caso e dos outros exemplificados anteriormente demonstra a ocorrência de falsas detecções pelo modelo e até a perda de lesão, evidenciando a necessidade de etapas de refinamento após a segmentação.

Figura 32 – Exemplo da segmentação inicial da lesão de endometriose para a paciente 80. (a), (b) e (c) apresentam exemplos de imagens da paciente saudável. As marcações manuais estão em verde, e as automáticas estão em laranja.



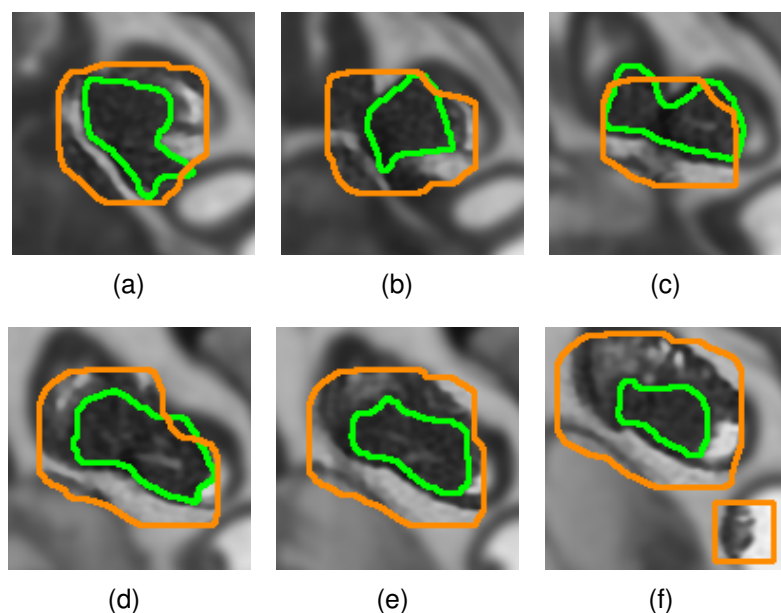
5.6 Extração da ROI da Lesão

Na etapa de extração da ROI da lesão, após a realização da subetapa de *ensemble* das detecções das imagens de um exame, aplica-se uma dilatação sobre a máscara gerada com o objetivo de expandir a região ao redor da lesão. Em seguida, aplica-se um filtro gaussiano para desfocar e reduzir a presença de texturas em áreas que possam se assemelhar à lesão, como o útero e outras regiões do retossigmoide.

Quando compara-se a marcação manual com a marcação automática expandida, obteve-se um valor de 0,9870 na sensibilidade, considerando todas as imagens com lesões. Mas na avaliação da sensibilidade obtida com ROI, temos o valor de 1, indicando que todas as lesões estão completamente dentro da ROI.

O resultado do processo de extração da ROI da lesão para a paciente 22 é apresentado na Figura 33. Nota-se, em alguns casos, que pequenas partes das lesões são perdidas, mas ainda há a possibilidade de ocorrência de falsos positivos. Em relação aos resultados da mesma paciente na etapa anterior, mostrados nas Figuras 30(e) e 30(f), observa-se que um falso positivo foi removido e outro foi gerado nas Figuras 33(e) e 33(f), respectivamente, devido ao processo de *ensemble* das detecções utilizado.

Figura 33 – Exemplo da extração da ROI da lesão para a paciente 22. (a), (b) (c), (d), (e) e (f) mostram exemplos de imagens do paciente. As marcações manuais estão em verde, e as automáticas, em laranja.



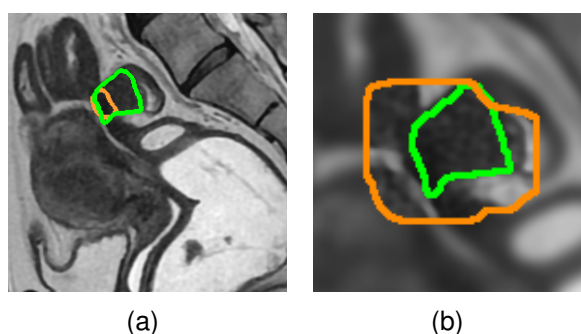
Na Figura 34, apresenta-se a comparação entre a segmentação inicial e o resultado da extração da ROI da lesão. Embora a segmentação inicial perca parte da lesão, é possível recuperar quase toda a lesão na região não suavizada da etapa de extração da ROI.

Quanto às imagens de pacientes saudáveis, entre as 10 nas quais foram detectadas candidatas a lesões na etapa de segmentação inicial, apenas seis continuaram possuindo algum tipo de segmentação por conta do *ensemble* das detecções desta etapa.

5.7 Segmentação Final da Lesão

Com base nos experimentos anteriores, optou-se por utilizar novamente a TransUnet para a segmentação das lesões de endometriose. Para o treinamento, foram

Figura 34 – Comparação da segmentação inicial da lesão com a extração da ROI da lesão para a paciente 22. (a) segmentação inicial da lesão e (b) extração da ROI. As marcações manuais estão em verde, e as automáticas, em laranja.



utilizados recortes de 128×128 pixels, extraídos a partir da dilatação da lesão marcada pelo especialista, com um desfoque aplicado fora da região dilatada. E, somente imagens com lesões foram adicionadas ao conjunto de treinamento.

A Tabela 20 mostra a distribuição das imagens para os conjuntos de treino, validação e teste. Assim como na segmentação inicial da lesão, esta etapa foi avaliada de três modos:

- I. Considerando os resultados provenientes diretamente da classificação;
- II. Considerando os resultados da classificação, mas com a remoção de imagens classificadas como saudáveis de pacientes diagnosticados com endometriose, pois, devido ao protocolo de marcação, essas imagens poderiam conter lesões não identificadas;
- III. O caso ideal, no qual apenas imagens com endometriose são encaminhadas para essa etapa.

Tabela 20 – Distribuição de imagens nos conjunto de treino, validação e teste

Conjunto de Dados	Quantidade de imagens
Treino	299
Validação	70
Teste (modo I)	178
Teste (modo II)	92
Teste (modo III)	83

As mesmas operações de aumento de dados utilizadas na etapa de segmentação inicial da lesão foram empregadas durante o treinamento, e podem ser observadas também na Tabela 21. Quanto aos hiperparâmetros, a otimização resultou nos seguintes valores: taxa de aprendizagem de 0,0001, tamanho de *batch* de 8 e uma função de perda

combinando BCE e Tversky. Estes foram definidos a partir dos valores apresentados na Tabela 22. Foi adotada uma paciência de 20 épocas para o *early stopping* e de 15 para o decaimento da taxa de aprendizado, com monitoramento da função de *loss*.

Tabela 21 – Operações de aumento de dados para a segmentação final da lesão

Tipo de Aumento de Dados	Valor
Rotação	30°
Deslocamento Horizontal	0,15%
Deslocamento Vertical	0,15%
Cisalhamento	0,02%
Zoom	0,8%-1,2%

Tabela 22 – Parâmetros de função de *loss*, tamanho do *Batch*, taxa de aprendizado e otimizador para segmentação

Função de <i>loss</i>	Batch	Taxa de Aprendizado	Otimizador
Dice	2	0,01	Adam
Tversky	4	0,001	SGD
Focal	8	0,0001	Adadelta
BCETversky	16	0,00001	-

Os resultados quantitativos obtidos estão apresentados na Tabela 23. Assim como observado na segmentação inicial da lesão, as imagens não marcadas de pacientes com endometriose impactam os resultados do modelo. Entretanto, os resultados dos modos II e III são próximos e demonstram que o método apresenta um bom desempenho ao diferenciar imagens realmente saudáveis daquelas com endometriose. Das seis imagens saudáveis provenientes da etapa anterior, cinco apresentaram a segmentação de alguma área identificada como suposta lesão. Ainda na tabela 23, também são exibidos os resultados obtidos no modo II sem a subetapa de suavização da etapa de extração da ROI da lesão, onde é possível observar uma queda em todas as métricas em relação ao modo II com suavização.

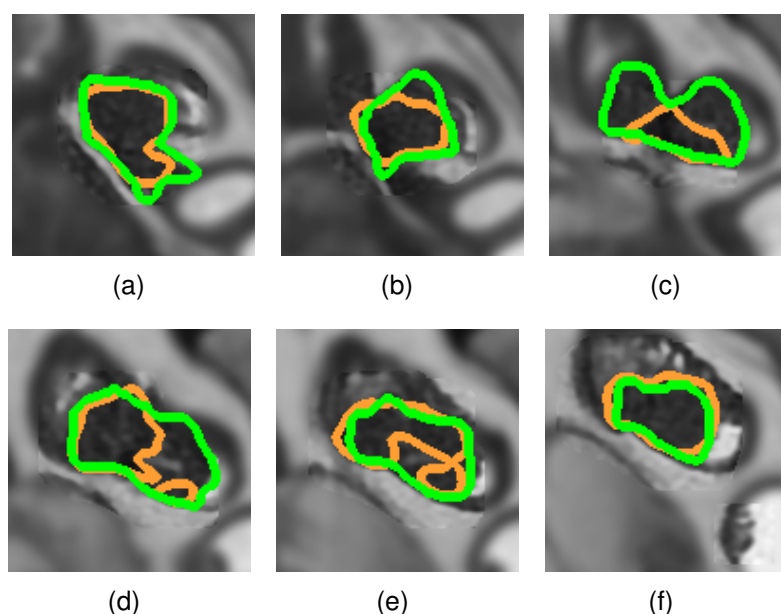
Tabela 23 – Métricas de desempenho do TransUNet para segmentação da lesão da endometriose

Modelo	Modo de avaliação	SEN	PREC	ESP	DICE
TransUNet	I	0,5542	0,6059	0,9892	0,5663
TransUNet	II	0,6660	0,7155	0,9869	0,6655
TransUNet	III	0,6731	0,7365	0,9862	0,6689
TransUNet	II (sem suavização)	0,5805	0,6122	0,9794	0,5556

Na Figura 35, são mostrados exemplos de resultados da paciente 22 obtidos nesta etapa, na qual é possível observar uma boa segmentação da forma na maioria das imagens. Na Figura 36, compara-se a segmentação inicial, a extração da ROI e a segmentação final da lesão para algumas imagens dessa mesma paciente. Observa-se

uma melhoria na segmentação da lesão de maneira geral, com destaque para o exemplo das Figuras 36(a), 36(b) e 36(c), que mostra a segmentação inicial, extração da ROI e segmentação da lesão, respectivamente. Nesse caso, foi possível recuperar grande parte da lesão, resultando em um aumento do valor de Dice de 0,3860 na segmentação inicial para 0,7833 na segmentação final da lesão. As Figuras 36(d), 36(e) e 36(f) apresentam um exemplo onde foi eliminado um falso positivo. Nas Figuras 36(g), 36(h) e 36(i), também é possível notar uma melhoria na segmentação da lesão, embora ainda haja perda da parte onde há uma área esbranquiçada, indicada pela seta vermelha.

Figura 35 – Exemplo da segmentação final da lesão de endometriose para a paciente 22. (a), (b) (c), (d), (e) e (f) exemplos de imagens do paciente. As marcações manuais estão em verde, e as automáticas, em laranja.

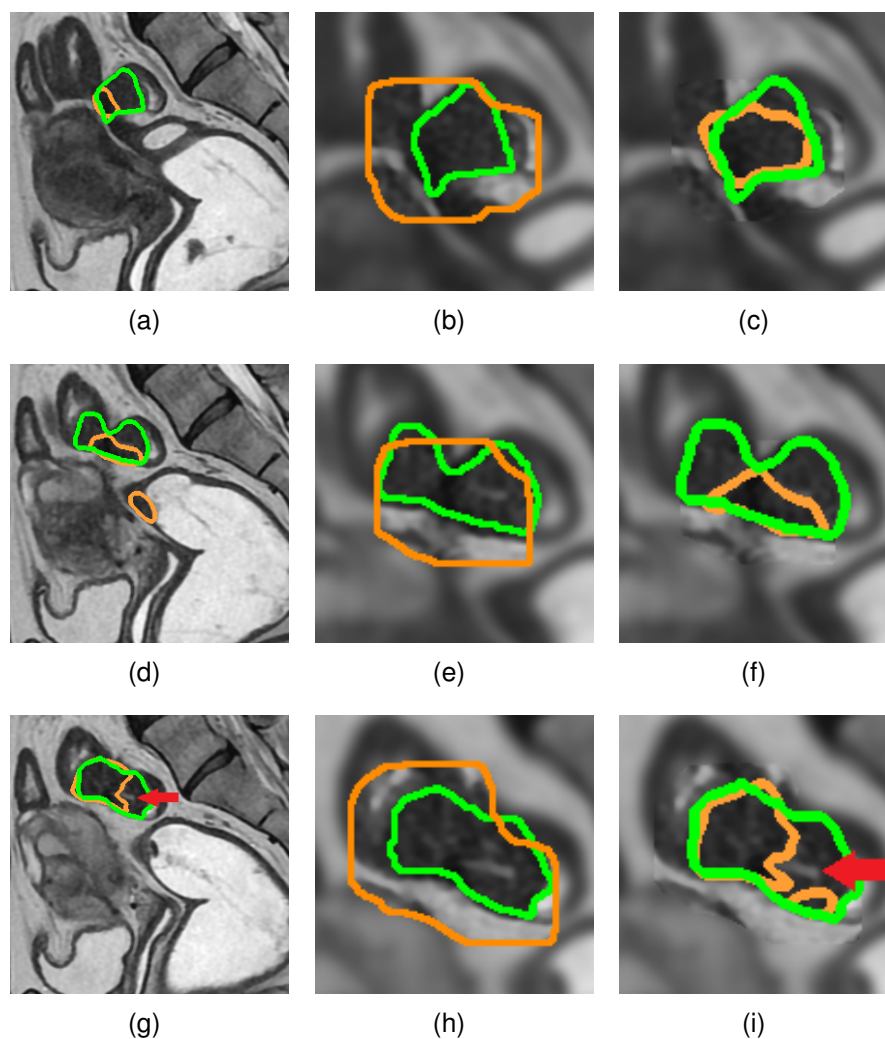


Considerando que apenas as imagens com endometriose chegam à etapa de segmentação (modo III de avaliação), dentre as 83 imagens do teste, em apenas 3 não foram encontradas lesões. Assim, constata-se que, em pelo menos 96,38% dos casos, houve alguma intersecção entre a marcação manual e automática do método proposto. Neste caso, o método consegue atingir 0,8074 no índice Dice como média entre a melhor segmentação de cada paciente.

5.8 Discussão

Nesta seção, é feita uma discussão sobre os resultados obtidos através do método proposto desde a extração das ROI da região do retossigmoide até a segmentação da lesão. Assim, disserta-se sobre os possíveis motivos que levaram à obtenção dos resultados apresentados, incluindo a análise da base de dados, do pré-processamento das imagens e a aplicação do modelo de aprendizado profundo. Além disso, serão

Figura 36 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 22. (a), (d) e (g) mostram a segmentação inicial da lesão, (b), (e) e (h) representam a extração da ROI da lesão, e (c), (f) e (i) apresentam a segmentação final da lesão em três imagens do exame, respectivamente. As marcações manuais estão em verde, e as automáticas, em laranja.



analisadas as vantagens e limitações do método proposto, considerando aspectos como a precisão, a sensibilidade e a especificidade alcançadas.

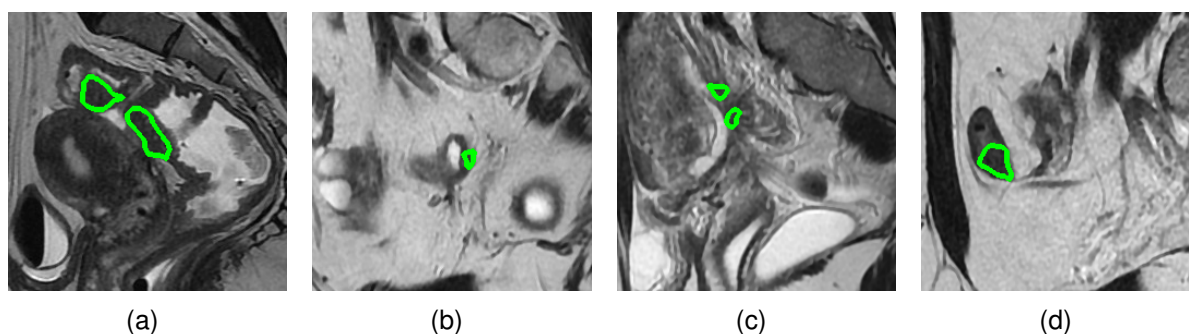
5.8.1 Extração da ROI do Retossigmoide

A etapa de identificação da ROI do retossigmoide permite que as fases subsequentes se concentrem especificamente nesta região, especialmente na fase de classificação, em que não se utiliza uma máscara que forneça a localização e a forma das lesões. Além disso, essa etapa oferece a vantagem de gerar imagens de tamanho uniforme, sem a necessidade de redimensionar algumas delas para o treinamento, o que exige dimensões iguais entre todas as imagens. A extração da ROI também reduz o

tamanho da imagem sem a necessidade de redimensionamento, diminuindo o custo computacional do processamento e acelerando o tempo de treinamento das RNAs utilizadas (THAMBAWITA et al., 2021; SABOTTKE; SPIELER, 2020).

Apesar disso, essa etapa poderia ser substituída por uma rede de detecção, como a YOLO (REDMON, 2016), ou por uma rede de segmentação, como a própria TransUnet. No entanto, isso exigiria a marcação da região do retossigmoide, o que não é uma tarefa trivial. Em algumas imagens, a região não é claramente definida, o que demanda uma análise detalhada. A Figura 37 apresenta exemplos da região do retossigmoide com lesões, que aparecem em formas variadas. Nota-se que, devido à falta de definição dos seus limites em alguns casos, como mostrado na Figura 37(c), e à baixa similaridade entre elas, torna-se praticamente inviável realizar a marcação adequada para a segmentação da área.

Figura 37 – Exemplos de resultados da extração da ROI do retossigmoide. (a), (b), (c) e (d) apresentam imagens de diferentes pacientes. Marcação manual da lesão em verde.



5.8.2 Active Learning

O uso do *active learning* demonstrou resultados promissores na construção de um modelo mais robusto. Essa abordagem permite selecionar automaticamente, com base no critério de incerteza, as imagens que compõem o conjunto de treinamento, garantindo que elas sejam mais representativas. Isso representa uma vantagem especialmente em bases com um número reduzido de imagens, mas com significativa diversidade de características. Essa estratégia pode potencializar o desempenho do modelo ao focar em exemplos mais informativos, promovendo um aprendizado superior em comparação com métodos que utilizam seleção aleatória, alcançando resultados melhores.

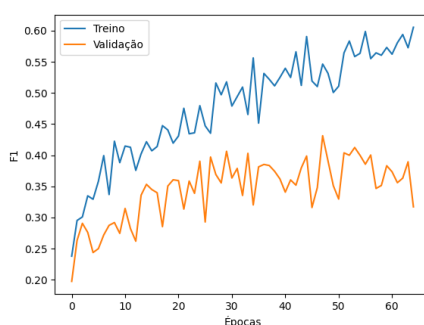
5.8.3 Classificação das Imagens com Lesão

Os resultados de classificação podem ser diretamente influenciados pela composição da base de imagens. Observa-se que algumas imagens saudáveis foram incorretamente classificadas como contendo endometriose pelo método automatizado.

Durante os testes, nota-se que algumas imagens rotuladas como saudáveis, localizadas próximas das imagens rotuladas como não saudáveis (com endometriose) nos exames de pacientes com endometriose, foram classificadas como não saudáveis. Isso sugere que pode haver vestígios de lesões nessas imagens, ou que elas compartilham características muito semelhantes. A utilização de um *ensemble* de modelos demonstrou melhorias na classificação, quando comparado ao uso de modelos individuais, além de aprimorar a identificação de pacientes com endometriose.

A ViT também foi uma das arquiteturas testadas, entretanto, o resultado não foi satisfatório em relação às demais, alcançando um valor máximo de F1-score de 0,40 na validação e de 0,38 no teste. Na Figura 38, pode-se observar o monitoramento da *loss* desta rede, na qual nota-se uma leve queda no aprendizado após o valor máximo de F1-score ser atingido.

Figura 38 – Monitoramento do F1-score da ViT.



5.8.4 Segmentação da Lesão

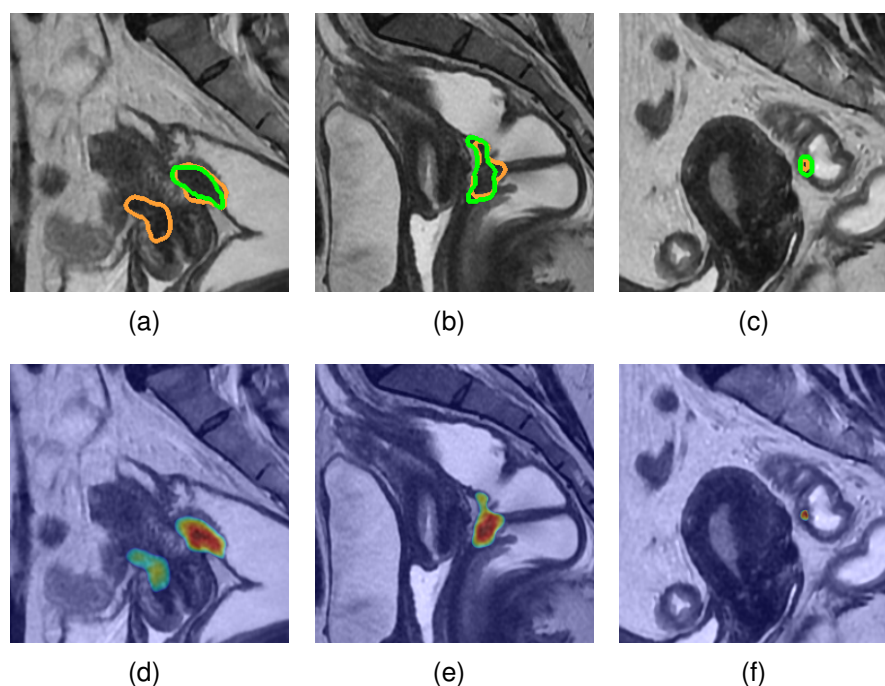
Nesta seção, trata-se das principais etapas realizadas para segmentar a lesão: segmentação inicial da lesão, extração da ROI da lesão e segmentação final.

Na segmentação inicial, observa-se dificuldade em identificar lesões pequenas ou áreas menores das lesões nas imagens. Contudo, esse não é o único desafio: diversas regiões possuem características semelhantes às das lesões, seja em textura ou forma, o que torna a identificação difícil, até mesmo para especialistas experientes, resultando em alguns falsos positivos. No entanto, o método é capaz de localizar lesões maiores, às vezes com detecções parciais, e lesões pequenas com áreas bem definidas, tornando-se uma etapa viável para a localização inicial das lesões.

Avaliou-se também o modelo por meio do mapa de calor gerado pela Grad-CAM na segmentação inicial da lesão, com o objetivo de analisar as regiões consideradas relevantes pela rede. A análise desses mapas revelou que a rede consegue, na maioria dos casos, manter o foco em regiões específicas do retossigmoide que apresentam características semelhantes às lesões. No entanto, observou-se que, em alguns casos, a

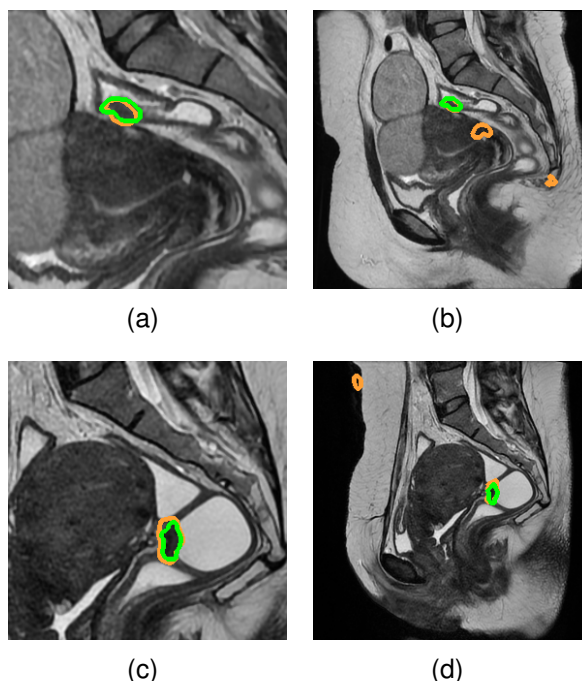
rede também direciona sua atenção para regiões do útero, que visualmente possuem similaridade com as lesões. Alguns exemplos estão apresentados na Figura 39. Nela, são mostrados três casos de pacientes distintos, nos quais todos focam na região da lesão, exceto na Figura 39(d), que também destaca uma parte do útero, embora com menor intensidade.

Figura 39 – Exemplos de resultados da segmentação inicial da lesão em diferentes pacientes e seus respectivos mapas de calor. As marcações manuais estão representadas em verde e as automáticas em laranja nos resultados em (a), (b) e (c). (d), (e) e (f) apresentam os respectivos mapas de calor associados.



Durante os testes, também foi avaliada a segmentação inicial da lesão sobre a imagem completa com a TransUnet. Observou-se que ocorreram segmentações em regiões além do retossigmoide, incluindo áreas fora da região pélvica. Esse comportamento pode ser explicado pela presença de várias áreas com textura semelhante às lesões, além da variação no posicionamento destas. Alguns casos são ilustrados na Figura 40. Na Figura 40(b), observa-se que, além das áreas escuras das segmentações obtidas, as regiões ao redor também apresentam semelhança com as de casos reais de lesão. Já na Figura 40(d), houve um caso em que a segmentação ocorreu externamente à paciente. Esse erro ocorreu devido à região não ser totalmente preta (pixels com valor zero) e conter pixels com valores semelhantes aos da região da lesão, o que levou o modelo a realizar uma segmentação incorreta.

Figura 40 – Comparação da segmentação inicial com e sem extração da ROI do retossigmoide. As marcações manuais estão representadas em verde e as automáticas, em laranja. (a) e (c) apresentam os resultados com a extração da ROI, enquanto (b) e (d) exibem os resultados correspondentes das mesmas imagens sem a extração da ROI.

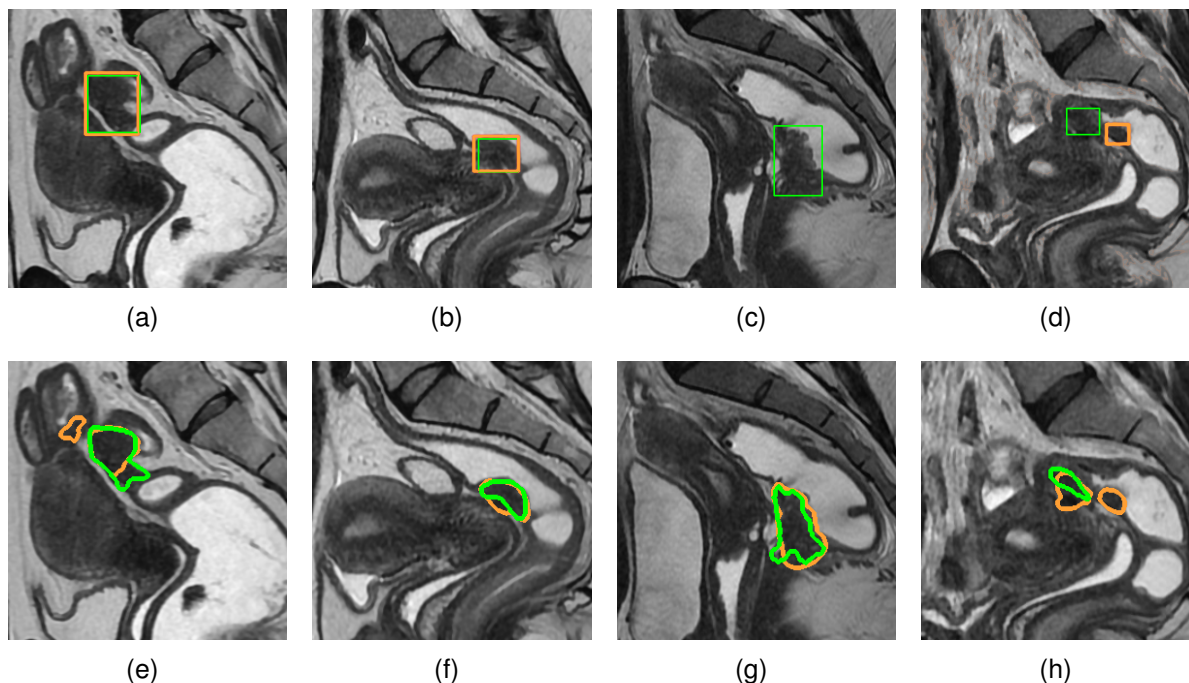


Além da avaliação das redes para segmentação, foi testada a YOLOv9 (WANG; YE; LIAO, 2025) como uma etapa inicial de detecção, com o objetivo de substituir as etapas de segmentação inicial e extração da ROI da lesão. No entanto, dos 83 casos de lesões de endometriose, a YOLOv9 conseguiu detectar algum candidato a lesão em apenas 29 imagens, sendo que 23 dessas detecções foram corretas. A Figura 41 mostra algumas imagens de exemplos deste experimento. De modo geral, a arquitetura se saiu bem em alguns casos onde as lesões são grandes e bem definidas, porém isto não é um padrão entre as imagens, o que explica a baixa sensibilidade.

A etapa de extração da ROI da lesão tem como objetivos recuperar lesões pequenas não identificadas no passo anterior, reduzir falsos positivos por meio do ensemble de detecções, recuperar partes da lesão não detectadas com a ajuda da dilatação e evitar o reaparecimento de falsos positivos distantes das lesões por meio do desfoque. Além disso, visa mitigar a criação de novos falsos positivos causados pela menor contextualização gerada pelo recorte.

Na etapa de segmentação final da lesão, busca-se aprimorar os resultados por meio de um novo uso da TransUnet. Com a combinação desta fase com a extração da ROI da lesão, foi possível recuperar lesões pequenas e reduzir lesões superestimadas. Nas Figuras 42(a), 42(b) e 42(c) nota-se que houve uma pequena redução da área

Figura 41 – Comparação entre exemplos de resultados da detecção da lesão com a YOLOv9 e da Segmentação inicial da lesão. (a), (b), (c) e (d) resultados da detecção, e (e), (f), (g) e (h) resultados da segmentação inicial. As marcações manuais estão em verde, e as automáticas, em laranja.



segmentada na etapa final. As Figuras 42(d), 42(e) e 42(f) apresentam um caso em que a segmentação inicial não detectou a lesão, possivelmente devido ao seu pequeno tamanho. No entanto, por meio da subetapa de *ensemble* de detecções na etapa de extração da ROI da lesão realizada com a imagem anterior a esta no exame (Figura 42(a)), foi possível recuperar a região da lesão.

Observou-se que, mesmo com parte da lesão em desfoque, ainda é possível identificar parcialmente a lesão nessa área. Na Figura 43, é possível observar que o método foi capaz de recuperar parte da lesão que nem havia sido detectada na segmentação inicial, apesar do desfoque.

Contudo, alguns erros relacionados a falsos positivos ainda persistem, especialmente devido à dificuldade na identificação da lesão, particularmente no útero, onde a textura é visualmente similar à da lesão. Na Figura 44, observa-se que não houve segmentação da lesão inicialmente, e que por conta do *ensemble* das detecções na extração da ROI da lesão, uma nova região falso positiva foi gerada (Figuras 44(a), 44(b) e 44(c)). Apesar disso, a TransUnet foi capaz de segmentar a lesão na etapa final. Nas Figuras 44(g), 44(h) e 44(i), houve uma redução da área falso positiva, gerada inicialmente.

A Figura 45 apresenta três exemplos de mapas de calor de pacientes distintos referentes à segmentação final da lesão de endometriose. As imagens geradas mostram

Figura 42 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 75. (a) e (d) mostram a segmentação inicial da lesão, (b) e (e) representam a extração da ROI da lesão, e (c) e (f) apresentam a segmentação final da lesão em duas imagens do exame. As marcações manuais estão em verde, e as automáticas, em laranja.

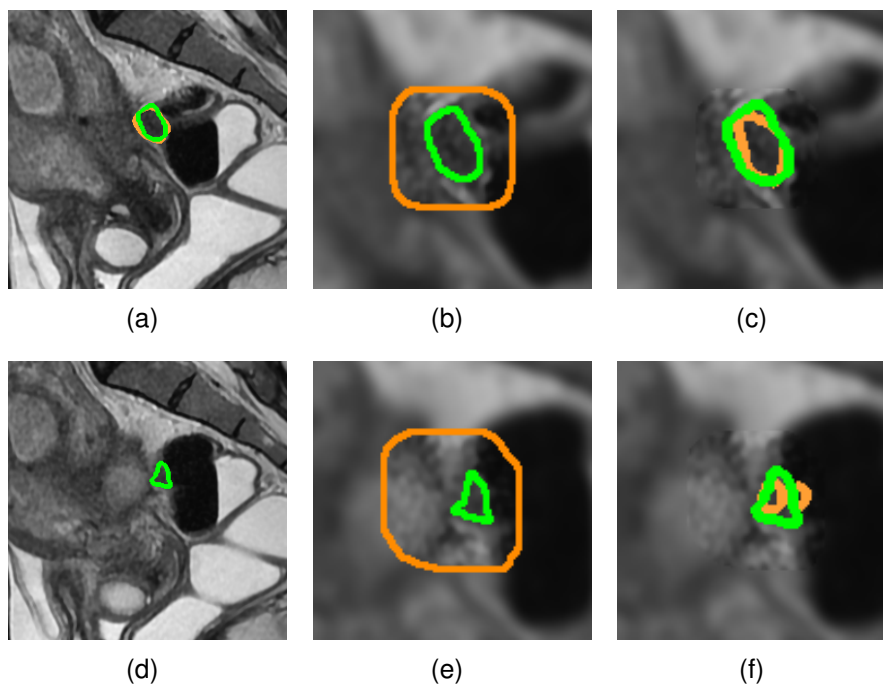
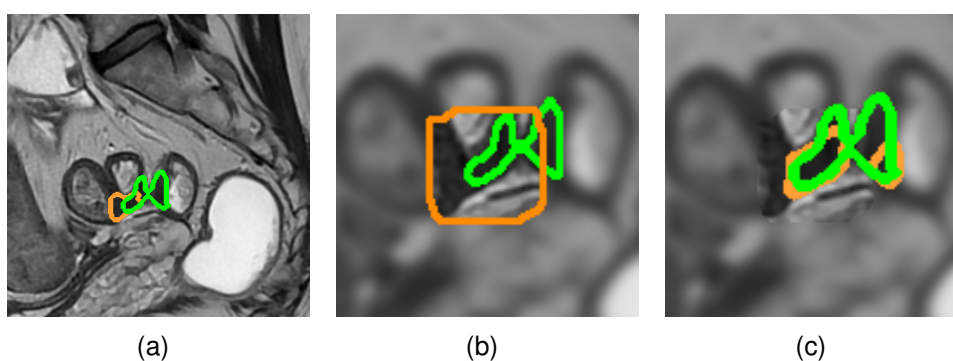


Figura 43 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 99. (a) mostra a segmentação inicial da lesão, (b) representa a extração da ROI da lesão, e (c) apresenta a segmentação final da lesão em uma imagem do exame. As marcações manuais estão em verde, e as automáticas, em laranja.

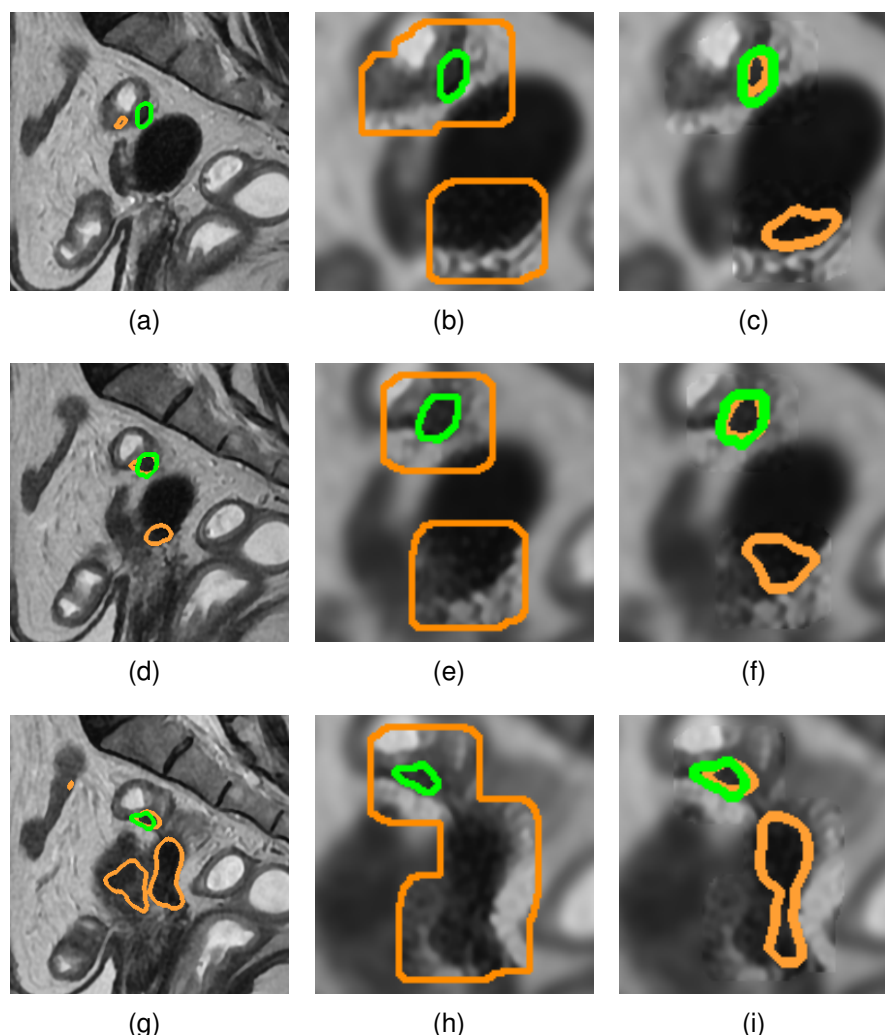


um foco maior nas bordas das lesões em alguns casos, o que pode contribuir para uma melhor identificação do formato das mesmas.

5.8.5 Limitações

Após uma análise dos resultados quantitativos e visuais obtidos pelo método de detecção de endometriose utilizando aprendizado profundo, foram identificados

Figura 44 – Comparação da segmentação inicial, extração da ROI da lesão e segmentação final da lesão para a paciente 23. (a), (d) e (g) mostram a segmentação inicial da lesão, (b), (e) e (h) representam a extração da ROI da lesão, e (c), (f) e (i) apresentam a segmentação final da lesão em três imagens do exame. As marcações manuais estão em verde, e as automáticas, em laranja.

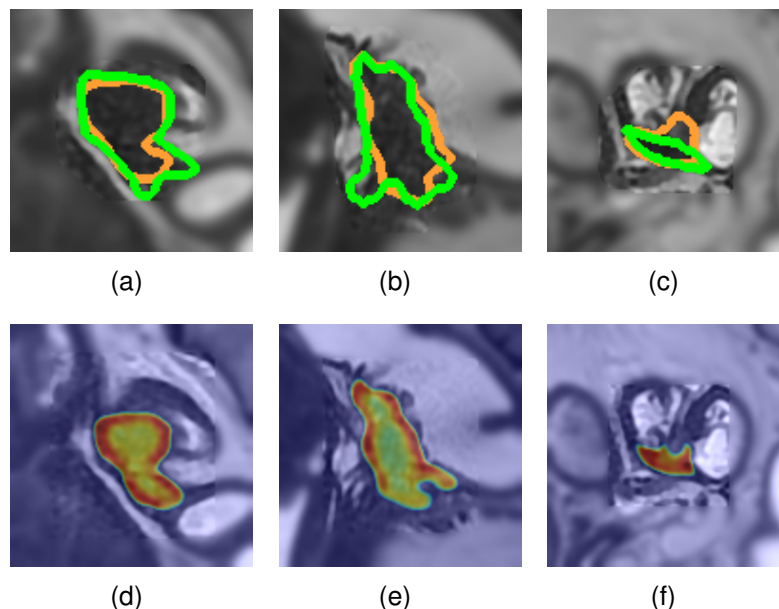


os principais erros e suas possíveis causas durante o processo de classificação e segmentação das lesões. De maneira geral, esses erros foram mais frequentes nas fatias em que as lesões apresentavam áreas menores, ou seja, nas fatias onde as extremidades das lesões estavam visíveis, mas com menor definição de sua extensão.

Uma das principais limitações do método proposto é a falta de representatividade na base de imagens utilizada. A inclusão de novas imagens, com maior diversidade de lesões e melhor definição da região do retossigmoide, pode contribuir de forma significativa para o aprimoramento do método e para a obtenção de melhores resultados.

Como foi apresentado no Capítulo 3, onde são mostrados exemplos de lesões presentes na base de imagens (Seção 3.2.1) e na Figura 37, a base utilizada no método contém uma variedade de lesões com diferentes formatos e texturas, o que dificulta a

Figura 45 – Exemplos de resultados da segmentação final da lesão em diferentes pacientes e seus respectivos mapas de calor. As marcações manuais estão representadas em verde e as automáticas em laranja nos resultados em (a), (b) e (c). (d), (e) e (f) apresentam os respectivos mapas de calor associados.



definição de um padrão consistente. Além disso, a localização das lesões na região do retossigmoide, bem como a forma desta última, também varia, tornando o processo de classificação e segmentação ainda mais complexo.

A depender da imagem do exame, a visualização da região retossigmoide pode variar significativamente, apresentando desde limites bem definidos até apenas partes visíveis. Em alguns casos, a presença de líquido na região contribui para uma definição visual mais clara da lesão, evidenciando seus limites. Em outros casos, porém, a textura da lesão é muito semelhante à das áreas ao redor, dificultando sua identificação.

As marcações podem apresentar imprecisões, mesmo quando realizadas por especialistas, e o diagnóstico por imagem pode não atingir 100% de precisão, como apontado na literatura. Considerando lesões de dois centímetros ou mais e uma acurácia de 90% por parte do especialista (WYKES; CLARK; KHAN, 2004), a margem de erro para a maior parte da lesão é considerada baixa. Contudo, áreas menores da lesão podem passar despercebidas ou não ser devidamente anotadas, devido à dificuldade de identificação, sendo uma das possíveis causas para os resultados sobre imagens consideradas saudáveis de pacientes com endometriose, especialmente na segmentação. Durante alguns dos testes realizados, algumas imagens foram reavaliadas novamente pelos especialistas, pois alguns dos falsos positivos detectados automaticamente eram semelhantes às lesões reais.

5.8.6 Considerações Finais

Neste capítulo, foram mostrados e discutidos os resultados obtidos e a influência do material, do método e da base de imagens utilizada. De modo geral, destacam-se os seguintes pontos:

- O trabalho alcançou resultados positivos em um campo relativamente pouco explorado;
- Os resultados obtidos com as etapas de extração de ROI mostraram-se essenciais para obter melhores resultados;
- O uso de *active learning* permite a seleção de imagens mais representativas para o treinamento;
- Vários experimentos foram conduzidos para encontrar as etapas apropriadas para os objetivos deste trabalho;
- O método proposto mostra bons resultados para classificar se uma paciente tem ou não endometriose;
- O método proposto pode localizar e delinear bem uma lesão em uma das fatias do exame, geralmente onde a maior parte da lesão é visível, permitindo que o especialista analise a lesão e sua progressão.

6 Conclusão

Assim como o diagnóstico da endometriose é complexo mesmo para especialistas experientes, métodos automáticos para identificação e segmentação da doença enfrentam diversos desafios. A natureza heterogênea das lesões, que podem variar em tamanho, forma e localização, dificulta a padronização dos critérios diagnósticos e exige abordagens avançadas baseadas em inteligência artificial e aprendizado profundo. Nesse contexto, o desenvolvimento de métodos automatizados como uma ferramenta de segunda opinião pode contribuir para uma avaliação ágil, reduzindo erros e auxiliando na detecção precoce. Um diagnóstico antecipado é fundamental para a qualidade de vida das pacientes.

A escassez de estudos que abordem a classificação e segmentação automatizada de lesões de endometriose em imagens de ressonância magnética é significativa. Entre os trabalhos encontrados, a maioria foca em apenas uma dessas tarefas e utiliza outros tipos de imagens. Esse cenário também reflete a falta de bases públicas de imagens de endometriose disponíveis para o desenvolvimento de métodos. Essas limitações dificultam a realização de comparações diretas com outros estudos.

Dessa forma, este trabalho apresentou um método abrangente que vai desde a classificação de imagens com endometriose até a segmentação das lesões, combinando técnicas de processamento de imagens com redes neurais artificiais (RNAs), utilizando uma base de imagens privada. O método obteve resultados promissores, demonstrando um potencial significativo para o avanço no auxílio automático ao diagnóstico de endometriose. O desempenho alcançado foi de aproximadamente 90% de acurácia para a classificação e 66% de Dice para a segmentação. Considerando o melhor caso por paciente, o valor de Dice ultrapassou 80%. Esses resultados podem ser aprimorados com a expansão da base de imagens, já que o número de imagens com endometriose é limitado em relação às imagens saudáveis. Além disso, uma base de dados mais representativa poderia melhorar o reconhecimento das diversas formas de lesões.

De modo geral, o método proposto apresentou resultados satisfatórios para um problema ainda pouco explorado, introduzindo uma abordagem inovadora para seu tratamento. Com este método, busca-se oferecer um suporte significativo ao diagnóstico, beneficiando tanto os pacientes quanto os radiologistas.

6.1 Trabalhos Futuros

Os resultados obtidos, tanto na classificação quanto na segmentação, demonstram-se promissores, evidenciando o potencial do método proposto. No entanto, melhorias ainda podem ser implementadas para aprimorar sua eficiência. A seguir, destacam-se

algumas sugestões para trabalhos futuros:

- Realizar a marcação da região do retossigmoide em formato de *bounding box*, viabilizando a aplicação de redes de detecção, como a YOLO;
- A utilização de uma rede multiescala, como a DuAT (TANG et al., 2023), para mitigar o problema em lesões pequenas;
- Expandir a base de imagens para aprimorar a representatividade e a eficácia do método;
- Avaliar o uso de técnicas de *ensemble* para combinar resultados da segmentação de treinamentos distintos, em função dos tamanhos das imagens;
- Investigar o uso de redes generativas (YEĞİN; AMASYALI, 2024), com foco na reconstrução da máscara, a partir da saída binária da segmentação inicial.

6.2 Produção científica

A Tabela 24 apresenta os artigos publicados relacionados à endometriose.

Tabela 24 – Artigos publicados sobre endometriose.

Título	Tipo	Qualis	Status
Figueredo, W. K. R., Silva, A. C., de Paiva, A. C., Diniz, J. O. B., Brandão, A., Oliveira, M. A. P. Automatic segmentation of deep endometriosis in the rectosigmoid using deep learning. <i>Em: Image and Vision Computing</i>. Ano: 2024.	Periódico	A2	Publicado
Figueredo, W. K., da Silva, I. F., Diniz, J. O., Silva, A. C., de Paiva, A. C., Salomão, A. C. B., de Oliveira, M. A. Abordagem computacional baseada em deep learning para o diagnóstico de endometriose profunda através de imagens de ressonância magnética <i>Em: Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)</i>. Ano: 2023.	Conferência	A4	Publicado
Pinto, D. M., Figueredo, W. K., da Silva, I. F., Silva, A. C., de Paiva, A. C., Salomao, A. C., de Oliveira, M. A. Segmentação Automática de Endometriose Profunda em Imagens de Ressonância Magnética Baseada em Swin-Unet. <i>Em: Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)</i> . Ano: 2024.	Conferência	A4	Publicado

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.

ALAN, J.; JIA, X. et al. Remote sensing digital image analysis: an introduction. **Remote sensing digital image analysis: an introduction**, 2006.

ALOM, M. Z.; YAKOPCIC, C.; HASAN, M.; TAHA, T. M.; ASARI, V. K. Recurrent residual U-Net for medical image segmentation. **Journal of Medical Imaging**, SPIE, v. 6, n. 1, p. 014006, 2019. Disponível em: <<https://doi.org/10.1117/1.JMI.6.1.014006>>.

ANDREUCCI, C. G. L.; SOUZA, L. M.; LOMBARDI, L.; ANDREUCCI, H. T. C. Principais achados radiológicos de endometriose em ressonância magnética. **Atas de Ciências da Saúde (ISSN 2448-3753)**, v. 2, n. 3, 2014.

AZEVEDO, L. L. A. d. Métodos estatísticos em aprendizado de máquinas para problemas de classificação. 2018.

BAILEY, H. R.; OTT, M. T.; HARTENDORP, P. Aggressive surgical management for advanced colorectal endometriosis. **Diseases of the colon & rectum**, Springer, v. 37, p. 747–753, 1994.

BALLEYGUIER, C.; CHAPRON, C.; DUBUISSON, J.; KINKEL, K.; FAUCONNIER, A.; VIEIRA, M.; HÉLÉNON, O.; MENU, Y. Comparison of magnetic resonance imaging and transvaginal ultrasonography in diagnosing bladder endometriosis. **The Journal of the American Association of Gynecologic Laparoscopists**, Elsevier, v. 9, n. 1, p. 15–23, 2002.

BARBOSA, V. A. d. F. et al. Sistemas inteligentes baseados em deep-wavelet e redes neurais convolucionais para apoio ao diagnóstico de câncer de mama usando imagens termográficas. Universidade Federal de Pernambuco, 2022.

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.

BELLEIS, P.; GIACOMETTI, C. F. Utilização da cannabis medicinal no tratamento da endometriose. **BrJP**, SciELO Brasil, v. 6, p. 95–96, 2023.

BELLEIS, P.; JR, J. A. D.; PODGAEC, S.; GONZALES, M.; BARACAT, E. C.; ABRÃO, M. S. Aspectos epidemiológicos e clínicos da endometriose pélvica: uma série de casos. **Revista da Associação Médica Brasileira**, SciELO Brasil, v. 56, p. 467–471, 2010.

BENGAR, J. Z.; WEIJER, J. van de; FUENTES, L. L.; RADUCANU, B. Class-balanced active learning for image classification. In: **Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision**. [S.l.: s.n.], 2022. p. 1536–1545.

- 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.
- BIANCHINI, Â. R. et al. Arquitetura de redes neurais para o reconhecimento facial baseado no neocognitron. Universidade Federal de São Carlos, 2001.
- BLOCH, F. Nuclear induction. **Physical review**, APS, v. 70, n. 7-8, p. 460, 1946.
- BRADLEY, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. **Pattern recognition**, Elsevier, v. 30, n. 7, p. 1145–1159, 1997.
- BRANDÃO, A. C.; SILVA, A. O. Diseases of the female pelvis: advances in imaging evaluation. **Magnetic Resonance Imaging Clinics**, Elsevier, v. 21, n. 2, p. 447–469, 2013.
- BREIMAN, L.; FRIEDMAN, J.; STONE, C. J.; OLSHEN, R. A. **Classification and Regression Trees**. Andover, England, UK: Taylor & Francis, 1984. ISBN 978-0-41204841-8.
- BROWNLIE, J. **Better deep learning: train faster, reduce overfitting, and make better predictions**. [S.l.]: Machine learning mastery, 2018.
- BRUNELLI, R. **Template matching techniques in computer vision: theory and practice**. New York, USA: John Wiley & Sons, 2009.
- CAMPAGNACCI, R.; PERRETTA, S.; GUERRIERI, M.; PAGANINI, A.; SANCTIS, A. D.; CIAVATTINI, A.; LEZOCHE, E. Laparoscopic colorectal resection for endometriosis. **Surgical Endoscopy and Other Interventional Techniques**, Springer, v. 19, p. 662–664, 2005.
- CARAÇA, D.; PODGAEC, S.; BARACAT, E.; ABRÃO, M. Mecanismos fisiopatológicos da dor pélvica na endometriose profunda. **Diagn Tratamento**, v. 16, n. 2, p. 57–61, 2011.
- CARMO, R. **Reto**. 2022. Acessado em 1 de julho de 2022. Disponível em: <<https://www.kenhub.com/pt/library/anatomia/reto>>.
- CHAICHIAN, S.; ABOLGHASEMI, J.; OMIDI, F. N.; RIMAZ, S.; NAJMI, Z.; MEHDIZADEH-KASHI, A.; MOAZZAMI, B. Factors affecting endometriosis in women of reproductive age: The differences between the results of neural network and logistic regression. **Shiraz E-Medical Journal**, Kowsar, v. 19, n. 9, 2018.
- 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.
- 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**. New York, NY, USA: Association for Computing Machinery, 2016. p. 785–794.
- CHEN, T.; HE, T.; BENESTY, M.; KHOTILOVICH, V.; TANG, Y.; CHO, H.; CHEN, K. et al. Xgboost: extreme gradient boosting. **R package version 0.4-2**, v. 1, n. 4, p. 1–4, 2015.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE Computer Society, 2017. p. 1800–1807. ISSN 1063-6919. Disponível em: <<https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.195>>.

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

DINIZ, J. O. B. et al. Métodos para segmentação de medula espinhal e esôfago em tomografia computadorizada de planejamento à radioterapia. Universidade Federal do Maranhão, 2021.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBERN, D.; ZHAI, X.; UNTERTHINER, T.; DEGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S.; USZKOREIT, J.; HOULSBY, N. An image is worth 16x16 words: Transformers for image recognition at scale. In: **International Conference on Learning Representations**. [s.n.], 2021. Disponível em: <<https://openreview.net/forum?id=YicbFdNTTy>>.

ENDOMETRIOSE online. **Endometriose – Quais as causas?** 2012. Acessado em 15 de Janeiro de 2021. Disponível em: <<https://endometrioseonline.com.br/endometriose-quais-as-causas/>>.

FERREIRA, J. L. et al. Segmentação automática da próstata em imagens de ressonância magnética utilizando redes neurais convolucionais, mapa probabilístico e treinamento adversário. Universidade Federal do Maranhão, 2019.

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

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 2001.

GANAI, M. A.; HU, M.; MALIK, A.; TANVEER, M.; SUGANTHAN, P. Ensemble deep learning: A review. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 115, p. 105151, 2022.

GHAZI, N.; ARJMAND, M.; AKBARI, Z.; MELLATI, A. O.; SAHEB-KASHAF, H.; ZAMANI, Z. 1h nmr-based metabolomics approaches as non-invasive tools for diagnosis of endometriosis. **International Journal of Reproductive BioMedicine**, Shahid Sadoughi University of Medical Sciences and Health Services, v. 14, n. 1, p. 1, 2016.

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.

GONZALEZ, R.; WOODS, R. **Processamento Digital de Imagens. Book**. São Paulo, Brasil: São Paulo: Pearson Pretience Hall, 2010.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning. book in preparation for mit press. URL; <http://www.deeplearningbook.org>, 2016.

GUERRIERO, S.; PASCUAL, M.; AJOSSA, S.; NERI, M.; MUSA, E.; GRAUPERA, B.; RODRIGUEZ, I.; ALCAZAR, J. L. Artificial intelligence (ai) in the detection of rectosigmoid deep endometriosis. **European Journal of Obstetrics & Gynecology and Reproductive Biology**, Elsevier, v. 261, p. 29–33, 2021.

HACKER, D. **013 CNN VGG 16 and VGG 19**. 2022. Acessado em 10 de novembro de 2023. Disponível em: <<https://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/>>.

HAGE, M. C. F. N. S.; IWASAKI, M. Imagem por ressonância magnética: princípios básicos. **Ciência Rural**, SciELO Brasil, v. 39, p. 1275–1283, 2009.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre, RS: Bookman Editora, 2001.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.]: IEEE Computer Society, 2016. p. 770–778.

HSU, A. L.; KHACHIKYAN, I.; STRATTON, P. Invasive and non-invasive methods for the diagnosis of endometriosis. **Clinical obstetrics and gynecology**, NIH Public Access, v. 53, n. 2, p. 413, 2010.

HUA, K.-L.; HSU, C.-H.; HIDAYATI, S. C.; CHENG, W.-H.; CHEN, Y.-J. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. **OncoTargets and therapy**, Taylor & Francis, p. 2015–2022, 2015.

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.

JU, C.; BIBAUT, A.; LAAN, M. van der. The relative performance of ensemble methods with deep convolutional neural networks for image classification. **Journal of Applied Statistics**, Taylor & Francis, v. 45, n. 15, p. 2800–2818, 2018.

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.

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.

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**. Springer, 2020. (Lecture Notes in Computer Science, v. 11962), p. 439–450. Disponível em: <https://doi.org/10.1007/978-3-030-37734-2_36>.

LEIBETSEDER, A.; SCHOEFFMANN, K.; KECKSTEIN, J.; KECKSTEIN, S. Endometriosis detection and localization in laparoscopic gynecology. **Multimedia Tools and Applications**, Springer, p. 1–25, 2022.

LEVINE, b. **What Is Bowel Endometriosis?** 2022. Acessado em 10 de novembro de 2022. Disponível em: <<https://www.everydayhealth.com/endometriosis/bowel-endometriosis/>>.

- LIANG, M.; HU, X. Recurrent convolutional neural network for object recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.]: IEEE Computer Society, 2015. p. 3367–3375.
- LIEW, X. Y.; HAMEED, N.; CLOS, J. An investigation of xgboost-based algorithm for breast cancer classification. **Machine Learning with Applications**, Elsevier, v. 6, p. 100154, 2021.
- LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLÁR, P. Focal loss for dense object detection. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2980–2988.
- LOH, W.-Y. Classification and regression trees. **Wiley interdisciplinary reviews: data mining and knowledge discovery**, Wiley Online Library, v. 1, n. 1, p. 14–23, 2011.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.
- MARCHIORI, E.; ZANETTI, G.; RODRIGUES, R. S.; SOUZA, L. S.; JR, A. S. S.; FRANCISCO, F. A. F.; HOCHHEGGER, B. Endometriose pleural: achados na ressonância magnética. **Jornal Brasileiro de Pneumologia**, SciELO Brasil, v. 38, p. 797–802, 2012.
- MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: IEEE. **2016 fourth international conference on 3D vision (3DV)**. [S.l.], 2016. p. 565–571.
- 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.
- NOGUEIRA, A. C. R.; SANTIAGO, M. T.; BAHIA, C. P.; SOARES, H. H. P. Tratamento da endometriose pélvica: uma revisão sistemática. **Revista Científica UNIFAGOC-Saúde**, v. 3, n. 2, p. 38–43, 2018.
- OLIVE, D.; SCHWARTZ, L. Endometriosis. **The New England journal of medicine**, Wiley Online Library, v. 328, n. 24, p. 1759–1769, 1993.
- OTSU, N. et al. A threshold selection method from gray-level histograms. **Automatica**, v. 11, n. 285-296, p. 23–27, 1975.
- PACHECO, C. A. R.; PEREIRA, N. S. Deep learning conceitos e utilização nas diversas áreas do conhecimento. **Revista Ada Lovelace**, v. 2, p. 34–49, 2018.
- 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.
- PODGAEC, S. Manual de endometriose: 2014/2015. **São Paulo, SP: Federação Brasileira das Associações de Ginecologia e Obstetrícia**, 2014.

- 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.
- 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.
- RAHMAN, M.; CAO, Y.; SUN, X.; LI, B.; HAO, Y. Deep pre-trained networks as a feature extractor with xgboost to detect tuberculosis from chest x-ray. **Computers & Electrical Engineering**, Elsevier, v. 93, p. 107252, 2021.
- REDMON, J. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016.
- REN, P.; XIAO, Y.; CHANG, X.; HUANG, P.-Y.; LI, Z.; GUPTA, B. B.; CHEN, X.; WANG, X. A survey of deep active learning. **ACM computing surveys (CSUR)**, ACM New York, NY, v. 54, n. 9, p. 1–40, 2021.
- REZATOFIGHI, H.; TSOI, N.; GWAK, J.; SADEGHIAN, A.; REID, I.; SAVARESE, S. Generalized intersection over union. June 2019.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. Cham, 2015. p. 234–241.
- 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.
- SABOTTKE, C. F.; SPIELER, B. M. The effect of image resolution on deep learning in radiography. **Radiology: Artificial Intelligence**, Radiological Society of North America, v. 2, n. 1, p. e190015, 2020.
- SALEHI, S. S. M.; ERDOGMUS, D.; GHOLIPOUR, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: SPRINGER. **International workshop on machine learning in medical imaging**. [S.l.], 2017. p. 379–387.
- SANAR. **Ressonância Magnética: O que É, Como É Feita e Mais - Sanar Medicina**. 2019. (Acessado em 09 de outubro de 2020). Disponível em: <<https://www.sanarmed.com/dica-de-radiologia-ressonancia-magnetica>>.
- 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.
- SCARDAPANE, A.; LORUSSO, F.; BETTOCCHI, S.; MOSCHETTA, M.; FIUME, M.; VIMERCATI, A.; PEPE, M.; ANGELELLI, G.; IANORA, A. S. Deep pelvic endometriosis: accuracy of pelvic mri completed by mr colonography. **La Radiologia Medica**, Springer Milan, p. 1–16, 2013.
- SCHLEMPER, J.; OKTAY, O.; SCHAAP, M.; HEINRICH, M.; KAINZ, B.; GLOCKER, B.; RUECKERT, D. Attention gated networks: Learning to leverage salient regions in medical images. **Medical image analysis**, Elsevier, v. 53, p. 197–207, 2019.

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

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.

SILVA, A. C. **Algoritmos para Diagnóstico Assistido de nódulos Pulmonares Solitários EM Imagens de Tomografia Computadorizada**. Tese (Doutorado) — Pontifical Catholic University of Rio de Janeiro, Brazil, 2004.

SILVA, I. F. S. d. et al. Segmentação automática de estruturas cardíacas em imagens de cine-ressonância magnética de eixo curto usando redes neurais convolucionais. Universidade Federal do Maranhão, 2023.

SILVA, L. A. d. Abordagem de aprendizado profundo para extração de quadros significativos em volumes de tomografia computadorizada. Universidade Federal do Amazonas, 2023.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: . [S.l.]: Computational and Biological Learning Society, 2015. p. 1–14.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 1–9.

TANG, F.; XU, Z.; HUANG, Q.; WANG, J.; HOU, X.; SU, J.; LIU, J. Duat: Dual-aggregation transformer network for medical image segmentation. In: SPRINGER. **Chinese Conference on Pattern Recognition and Computer Vision (PRCV)**. [S.l.], 2023. p. 343–356.

TAYLOR, H. Diagnosing endometriosis: is laparoscopy the gold standard. **Supplement to OBG Management**, v. 1, p. 52, 2019.

THAMBAWITA, V.; STRÜMKE, I.; HICKS, S. A.; HALVORSEN, P.; PARASA, S.; RIEGLER, M. A. Impact of image resolution on deep learning performance in endoscopy image classification: An experimental study using a large dataset of endoscopic images. **Diagnostics**, MDPI, v. 11, n. 12, p. 2183, 2021.

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.

VARELLA, M. **Reto**. 2021. Acessado em 1 de novembro de 2021. Disponível em: <https://drauziovarella.uol.com.br/corpo-humano/reto/>.

- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.
- VISALAXI, S.; MUTHU, T. S. Automated prediction of endometriosis using deep learning. **International Journal of Nonlinear Analysis and Applications**, International Journal of Nonlinear Analysis and Applications, v. 12, n. 2, p. 2403–2416, 2021.
- VISALAXI, S.; PUNNOOSE, D.; MUTHU, T. S. Lesion extraction of endometriotic images using open computer vision. In: IEEE. **2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)**. Tamil Nadu, India, 2021. p. 747–751.
- VISALAXI, S.; SUDALAIMUTHU, T. Automated segmentation of endometriosis using transfer learning technique. **F1000Research**, F1000 Research Limited, v. 11, n. 360, p. 360, 2022.
- WAN, T.; XU, K.; YU, T.; WANG, X.; FENG, D.; DING, B.; WANG, H. A survey of deep active learning for foundation models. **Intelligent Computing**, AAAS, v. 2, p. 0058, 2023.
- WANG, C.-Y.; YEH, I.-H.; LIAO, H.-Y. M. Yolov9: Learning what you want to learn using programmable gradient information. In: SPRINGER. **European conference on computer vision**. [S.l.], 2025. p. 1–21.
- WYKES, C. B.; CLARK, T. J.; KHAN, K. S. Accuracy of laparoscopy in the diagnosis of endometriosis: a systematic quantitative review. **BJOG: An International Journal of Obstetrics & Gynaecology**, Wiley Online Library, v. 111, n. 11, p. 1204–1212, 2004.
- 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.
- YASUMOTO, K.; SUZUKI, A.; MATSUMURA, N.; BABA, T.; OIWA, Y.; MIYAMOTO, S.; MATSUURA, M.; OBAMA, K.; KIDO, A.; UMEOKA, S. et al. Cancer risk and management in a woman with peutz-jeghers syndrome. In: SPRINGER. **International Cancer Conference Journal**. [S.l.], 2012. v. 1, p. 1–14.
- YEĞİN, M. N.; AMASYALI, M. F. Generative diffusion models: A survey of current theoretical developments. **Neurocomputing**, Elsevier, v. 608, p. 128373, 2024.
- YING, X. An overview of overfitting and its solutions. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2019. v. 1168, n. 2, p. 022022.
- 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.
- ZHOU, B.; LAPEDRIZA, A.; XIAO, J.; TORRALBA, A.; OLIVA, A. Learning deep features for scene recognition using places database. **Advances in neural information processing systems**, v. 27, 2014.
- ZWAS, F.; LYON, D. Endometriosis-an important condition in clinical gastroenterology. **Digestive Diseases and Sciences**, PLENUM PUBL CORP 233 SPRING ST, NEW YORK, NY 10013, v. 36, n. 3, p. 353–364, 1991.