

# **Segmentação de Endometriose Profunda em Imagens de Ressonância Magnética utilizando Redes Convolucionais como Codificadores para Arquitetura U-net**

**Marcos V. Farias<sup>1</sup>, Weslley K. R. Figueiredo<sup>1</sup>, Aristófanes C. Silva<sup>1</sup>  
Anselmo C. de Paiva<sup>1</sup>, Alice C. C. B. Salomão<sup>2</sup>, Marco A. P. de Oliveira<sup>3</sup>**

<sup>1</sup>Núcleo de Computação Aplicada – Universidade Federal do Maranhão (UFMA)  
Av. dos Portugueses, 1966 - Vila Bacanga, São Luís - MA, 65080-805

<sup>2</sup>Clínica Fonte de Imagem - Rio de Janeiro, RJ - Brasil

<sup>3</sup>Universidade do Estado do Rio de Janeiro (UERJ) - Rio de Janeiro, RJ - Brasil

{marvinfar852, weslley.kelson, ari, paiva}@nca.ufma.br,

{brandaosalomao, endometriose}@gmail.com

**Abstract.** Deep endometriosis is a globally prevalent pathology that affects around 20% of women of reproductive age and afflicts the patient with symptoms ranging from abdominal pain to bleeding and infertility. Its diagnosis can be made by imaging tests, which, combined with machine learning models, represent a significant improvement in this process. This paper brings an automatic approach to endometriosis segmentation in magnetic resonance images, obtaining a Dice of 51% and a Recall of 66,58%.

**Resumo.** A endometriose profunda é uma patologia de escala global, afetando cerca de 20% das mulheres em idade reprodutiva. Seus sintomas vão de dores abdominais até sangramento e infertilidade quando em estágios avançados. Seu diagnóstico pode ser realizado por exames de imagem, os quais, combinados com modelos de aprendizado de máquina, representam uma poderosa ferramenta para o profissional em ambiente clínico. O presente trabalho traz uma abordagem automática para segmentação de endometriose em imagens de ressonância magnética, obtendo Dice de 51% e Sensibilidade de 66,58%.

## **1. Introdução**

A endometriose é uma doença que acomete a pelve feminina, sendo caracterizada pela presença do tecido endometrial fora da cavidade uterina [Gomes and Alves 2018]. Trata-se de uma patologia infecciosa e de prevalência em torno de 20% das mulheres em idade reprodutiva [Gomes and Alves 2018, Nácul and Spritzer 2010]. O quadro clínico da paciente com endometriose é bastante variável, podendo ser assintomático ou, na maioria dos casos, sintomático [Nácul and Spritzer 2010]. Os principais sintomas são dismenorreia, dor pélvica crônica, infertilidade, dispareunia de profundidade, sintomas intestinais e urinários cíclicos, como dor ou sangramento ao evacuar/urinar durante o período menstrual [Bellelis et al. 2010]. Diversas teorias foram postuladas para tentar explicar a etiopatogenia desta patologia, sendo a mais aceita a estabelecida por Sampson em 1927, conhecida como teoria da mestruação retrógrada. Esta teoria defende que no momento da

menstruação, parte do sangue eliminado passa pelas tubas uterinas vai para a cavidade abdominal transportando células que possuem a capacidade de crescer em diversas regiões da cavidade abdominal. Quando o sistema imunológico, que é responsável pela defesa do organismo, não consegue eliminar essas células, elas podem se instalar e proliferar, levando ao estabelecimento da doença [Silva et al. 2021, Gomes and Alves 2018].

A dificuldade no diagnóstico da endometriose é um aspecto relevante, já que, por se tratar de uma patologia sem sintomas característicos, esta é facilmente confundida com outras enfermidades. A classificação das lesões é dada, após a avaliação de sua morfologia, em endometriose superficial, quando a profundidade das lesões que penetram no espaço retroperitoneal ou na parede de órgãos pélvicos for menor que 5 mm, e profunda, quando a profundidade for maior que 5 mm [Caraça et al. 2011]. Atualmente, a videolaparoscopia é considerada padrão ouro no diagnóstico da endometriose, apesar de ser um método invasivo. O procedimento é realizado por meio de pequenas incisões na região abdominal, local onde é inserido o aparelho com uma câmera que guiará o médico durante a cirurgia [Gomes and Alves 2018]. Embora o diagnóstico definitivo da endometriose necessite de intervenção cirúrgica, preferencialmente por videolaparoscopia, diversos achados nos exames físico, de imagem e laboratoriais já podem predizer, com alto grau de confiabilidade, que a paciente apresenta endometriose [Nácul and Spritzer 2010].

Sendo a endometriose uma patologia por vezes assintomática e de difícil diagnóstico, além de, na maioria dos casos, existir uma falta de consenso referente a qual conduta seguir diante do quadro clínico das pacientes, entende-se que a ressonância magnética (**RM**) se destaca como um excelente método pré-cirúrgico de diagnóstico [Gomes and Alves 2018]. Sua importância no diagnóstico da endometriose está na identificação e avaliação da extensão das lesões subperitoneais, não-visíveis à laparoscopia, apresentando acurácia, sensibilidade e especificidade acima de 90% para endometriose profunda [Coutinho Junior et al. 2008].

O presente trabalho busca apresentar um método para a segmentação de lesões de endometriose profunda em imagens de **RM**. Foram realizadas comparações entre arquiteturas convolucionais empregadas como *backbones* de codificação de uma rede U-net, evocando diferentes estratégias de extração de características.

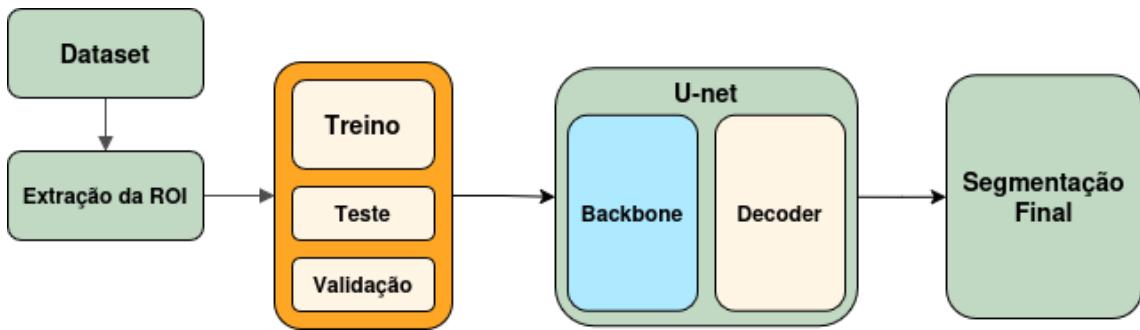
Este trabalho está estruturado do seguinte modo: o método proposto é apresentado na Seção 2; os experimentos realizados e resultados obtidos estão presentes na Seção 3; por fim, a Seção 4 contém conclusões e trabalhos futuros.

## 2. Método

A presente seção apresenta em detalhes o fluxo realizado para a segmentação da endometriose, sendo este compreendido das etapas de: extração da Região de Interesse (**ROI**), aumento de dados via difusão, pré-processamento das imagens geradas e testes com os diferentes backbones. O esquema desta abordagem pode ser visualizado na Figura 1.

### 2.1. Extração da ROI

Capturar uma **ROI** que compreenda com acuidade a região das lesões nas imagens é crucial, não só para prover ao modelo um dado que facilite a sua aprendizagem, visto que os exames de **RM** muitas vezes comprehendem informações não relevantes, mas também para a redução do tamanho das imagens e consequentemente do custo computacional.

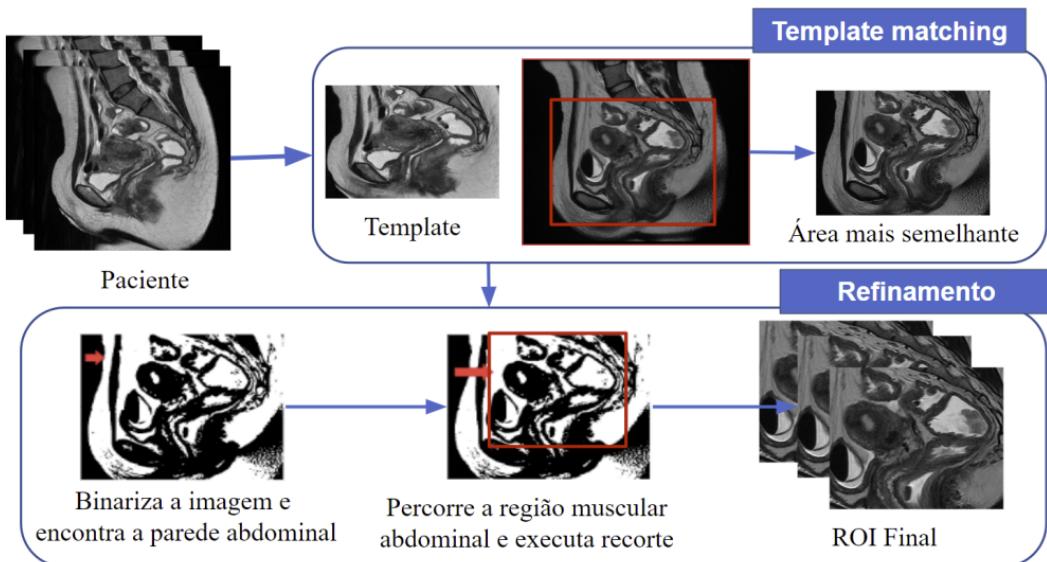


**Figura 1.** Fluxograma dos experimentos realizados.

Para o presente problema, as marcações de lesão estão internas a uma região conhecida como reto-sigmóide, que compreende o final do intestino grosso. O procedimento adotado para a identificação e extração da **ROI** é o proposto por [Figueiredo et al. 2023]. Primeiramente é aplicada a técnica de *Template Matching* para identificar uma região mínima entre as imagens que englobe todos os órgãos do paciente. A imagem utilizada como template é um recorte de um exame selecionado por apresentar uma melhor visualização do reto-sigmóide. É feita então a binarização da imagem resultante para que as estruturas escuras e claras fiquem bem definidas. Em seguida, busca-se pela parede abdominal por meio de uma análise pixel a pixel, a partir do pixel mais superior à esquerda que não for branco. Com esse pixel, tem-se a parte exterior do abdômen. A partir desse ponto, percorre-se os pixels brancos no eixo x até encontrar novamente um pixel preto, que representa a região do músculo abdominal. Continua-se a percorrer o eixo x até encontrar um pixel branco, que é a região onde se encontram os órgãos. A partir desse pixel, faz-se um corte na imagem considerando a altura da imagem de entrada encontrada pelo template matching e o valor de x desse pixel [Figueiredo et al. 2023]. A nova imagem é então obtida a partir da saída dos passos anteriores recortando 256x256 pixels a partir do canto superior esquerdo. A Figura 2 detalha o processo descrito.

## 2.2. Arquitetura U-net para segmentação

Segmentação semântica compreende uma área da visão computacional que objetiva atribuir os pixels de uma imagens a diferentes classes, de modo que essas classes representem diferentes objetos de interesse na imagem. Dentre os modelos para segmentação, a U-net [Ronneberger et al. 2015] se destaca por se basear em uma abordagem de codificação-decodificação, sendo muito utilizada neste âmbito. Sua principal característica é a capacidade de compactar os dados de entrada em um espaço latente e então expandir essa representação vetorial com a finalidade de segmentar a região desejada. Outro ponto fundamental de sua arquitetura são as *skip connections*, conexões que acontecem entre o codificador e o decodificador da rede. Essas conexões possibilitam o fluxo de informações entre esses blocos, o que ajuda com a preservação de detalhes mais finos ao longo do processamento do modelo. Adicionalmente, arquiteturas convolucionais profundas continuam representando o estado da arte para uma enorme gama de problemas nas áreas de visão computacional e processamento de imagens. Essa dominância das *CNNs* não foi coincidência, uma vez que a estratégia de uma “janela deslizante” é intrínseca para diversos problemas de processamento visual [Liu et al. 2022]. Desse modo, redes convolucionais são amplamente utilizadas como bloco de codificação padrão, também chamado



**Figura 2. Diagrama extraído de [Figueiredo et al. 2023] ilustrando a extração da ROI.**

de *backbone*. Durante o presente estudo, foram realizados experimentos com as seguintes *convnets*: A ConvNext, VGG19 e EfficientNetV2.

A ConvNext [Liu et al. 2022] é uma arquitetura convolucional que busca imbuir alguns comportamentos visualizados em modelos baseados em atenção (*transformers*). Partindo de uma ResNet [He et al. 2015], são propostas diversas modificações arquiteturais como: modificação da quantidade de blocos por etapa da rede; utilização de convoluções por ponto; substituir a função de ativação ReLU pela GELU e empregar *Layer Normalization* em detrimento do *Batch Normalization*. Os autores defendem que aproximar os modelos convolucionais das características apresentadas por um modelo *transformer*, em especial a *Swin Transformer* [Liu et al. 2021], possibilitando que esse modelo alcance desempenho similar enquanto mantém sua fácil implementação e custo computacional reduzido.

A VGG19 [Simonyan and Zisserman 2015] por sua vez, explora o impacto da expansão da profundidade de uma rede convolucional, adicionando mais camadas de codificação compostas de pequenos *kernels* (3x3). Cada estágio de extração de características é composto por 3 operações de convolução seguidas de um *max-pooling*, que reduz a dimensionalidade dos mapas de características através da amostragem dos maiores valores presentes. A versão “VGG19” é composta então por uma rede VGG que possui 19 estágios de codificação, tendo como o saída o vetor de características resultante desse processo.

O modelo da EfficientNet v2, proposto por [Tan and Le 2021], realiza melhorias na rede EfficientNet [Tan and Le 2020], arquitetura já apresentada pelos autores em trabalho anterior. A nova versão apresentada modifica a estrutura da rede original a fim de melhorar o custo computacional e tempo de treinamento. Algumas das melhorias implementadas por eles são: Utilização dos blocos de codificação MBConv e FusedMBConv, os quais ponderam os canais do mapa de características mais relevantes através de sua

relação com os demais; utilizar *kernels* 3x3 para diminuir o número de operações, apesar de aumentar o número de blocos de codificação para compensar. Assim, o modelo totaliza 6 estágios de extração de características, cada um contendo um número diferente de camadas convolucionais.

### 2.3. Métricas de Avaliação

Ao compararmos a segmentação gerada por nossos modelos e a gerada pelo especialista nos deparamos com as seguintes situações: verdadeiros positivos (**VP**) são pixels corretamente identificados como endometriose; falsos positivos (**FP**) são pixels incorretamente identificados como endometriose; falsos negativos (**FN**) são pixels incorretamente indicados como não sendo endometriose. A partir dessas informações podemos calcular métricas que nos indicam quão efetivo nosso modelo foi na tarefa desempenhada. As métricas utilizadas neste trabalho são o **Dice** [1], a **Precisão** [2], a **Sensibilidade** [3] e a **Intersecção sobre a União (IoU)** [4].

$$Dice = \frac{2 * VP}{VP + FP + FN} \quad (1)$$

$$Precisão = \frac{VP}{VP + FP} \quad (2)$$

$$Sensibilidade = \frac{VP}{VP + FN} \quad (3)$$

$$IoU = \frac{VP}{VP + FP + FN} \quad (4)$$

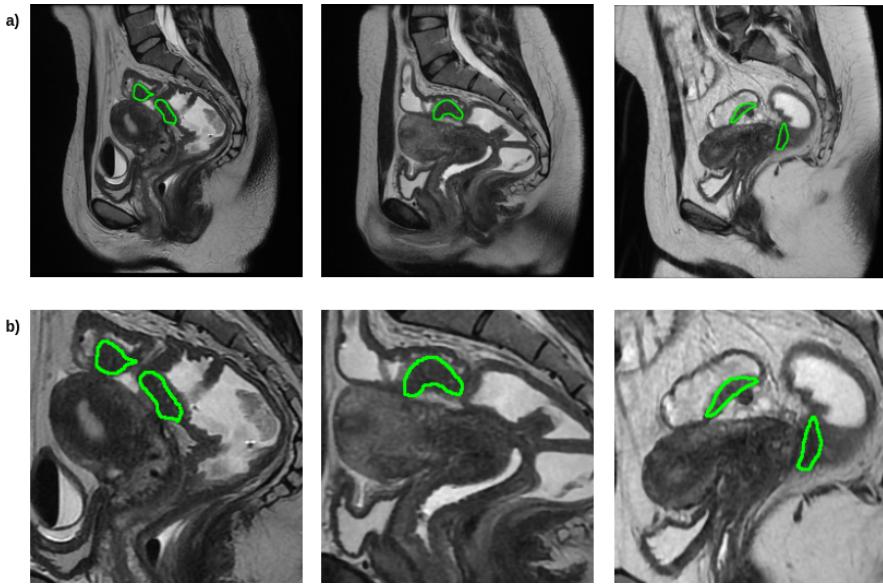
## 3. Experimentos

### 3.1. Base de Dados

A base de imagens utilizada neste trabalho foi disponibilizada pela Clínica Fonte Imagem (Rio de Janeiro) coordenada pela Dra. Alice Brandão, que também atuou como especialista na marcação das lesões nas imagens. A aquisição dos exames se deu através de ressonância magnética, resultando em volumes do abdômen inferior dos pacientes com resolução 512x512 no plano sagital. Para a tarefa de segmentação, só nos são interessantes imagens que possuam a presença de endometriose, portanto, imagens que não possuíssem marcações da patologia não foram incluídas. Ao todo, a base é constituída de exames de 105 pacientes, totalizando 451 fatias com lesão, sendo que cada paciente possui entre 1 e 9 imagens presentes nesse conjunto. É importante citar que somente lesões com diâmetro superior a 2cm foram marcadas. A Figura 3 apresenta algumas fatias de exames presentes na base antes e depois da extração da ROI.

### 3.2. Configurações experimentais

Para a realização dos experimentos a base de dados foi dividida nos conjuntos de treino, validação e teste, com 274, 95 e 82 imagens de dimensão convertida para 128x128, respectivamente. Todas os resultados apresentados foram realizados em *hold-out*, onde



**Figura 3. Exemplos de imagens presentes na base de imagens: a) apresenta as imagens antes da extração da região de interesse correspondente ao reto-sigmóide e b) apresenta os exames após essa etapa.**

realizou-se 3 execuções de cada experimento para captura da média e desvio padrão dos mesmos.

Os hiperparâmetros utilizados em todos os experimentos (com os diferentes *backbones*) foram *batch size* de 16, otimizador AdamW com taxa de aprendizado  $10^{-4}$  e 200 épocas de treinamento. Com relação às funções de perda, foram realizados testes com a *Dice Loss* e *Tversky Loss*. A *Dice Loss*, além de amplo uso na literatura relacionada a segmentação, possui uma lógica simples que verifica o nível de sobreposição entre a lesão real e a predição do modelo, determinando assim um coeficiente de erro para ser retropropagado para a rede. No entanto, para problemas onde existe um desbalanceamento entre as classes abordadas (lesão e fundo no caso deste trabalho), a *Tversky Loss* apresenta boas propriedades de controle sobre essa condição. Tratando-se de uma generalização da função *Dice*, a função de perda *Tversky* possui a capacidade de, através de seus parâmetros  $\alpha$  e  $\beta$ , penalizar de forma mais acentuada o modelo pela geração de falsos positivos ou falsos negativos. Para o contexto de imagens médicas é interessante que se tenha um baixo número de falsos negativos, que são casos onde o modelo não consegue identificar a patologia em um paciente doente. Assim, os parâmetros da função de perda foram configurados como  $\alpha = 0.7$  e  $\beta = 0.3$ , onde  $\alpha$  representa o peso dado aos falsos negativos.

Os experimentos apresentados nesta seção foram conduzidos em uma GPU NVIDIA GeForce RTX 3060 utilizando o *framework PyTorch*. É importante ressaltar que as arquiteturas empregadas como *backbones* utilizam pesos pré-treinados na base **ImageNet1k**.

### 3.3. Resultados

As métricas obtidas nos experimentos realizados podem ser visualizadas na Tabela 1.

Os resultados obtidos indicam uma certa vantagem na utilização da *Tversky Loss*, o que, dado o pequeno tamanho das lesões, é esperado. É importante observar que a

**Tabela 1. Resultados obtidos**

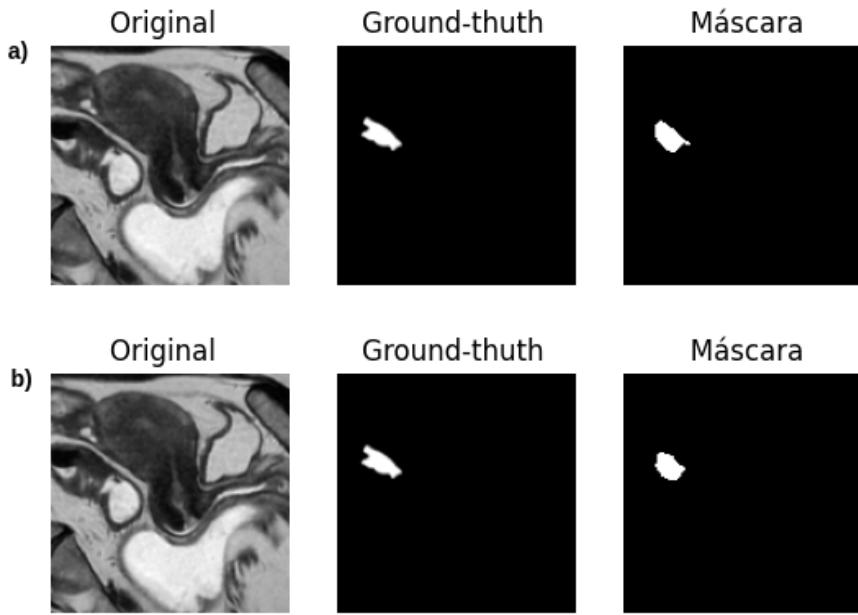
<b>Backbone</b>	<b>Dice</b>	<b>Precisão</b>	<b>Sensibilidade</b>	<b>IoU</b>
<b>Dice Loss</b>				
ConvNext	49,95% $\pm$ 3,3	52,79% $\pm$ 1,8	53,74% $\pm$ 4,9	29,36% $\pm$ 1,8
VGG19	49,22% $\pm$ 1,5	53,60% $\pm$ 0,8	62,73% $\pm$ 3,1	32,22% $\pm$ 0,7
EfficientNet v2	40,88% $\pm$ 2,8	51,28% $\pm$ 5,3	50,70% $\pm$ 2,9	26,83% $\pm$ 2,5
<b>Tversky Loss</b>				
ConvNext	46,84% $\pm$ 1,8	51,90% $\pm$ 3,5	60,42% $\pm$ 1,6	30,54% $\pm$ 1,3
<b>VGG19</b>	<b>50,99% <math>\pm</math> 1,0</b>	<b>54,21% <math>\pm</math> 1,3</b>	<b>66,58% <math>\pm</math> 1,8</b>	<b>32,90% <math>\pm</math> 0,7</b>
EfficientNet v2	45,30% $\pm$ 3,2	51,05% $\pm$ 1,8	57,63% $\pm$ 5,5	29,05% $\pm$ 1,9

métrica mais afetada pelo uso desse função de perda foi a sensibilidade, cuja qual é penalizada por falsos negativos. No caso da ConvNext por exemplo, o modelo apresentou melhora em somente duas das quatro métricas (sensibilidade e IoU). Nesse sentido, é possível observar que a *Tversky Loss* influenciou positivamente nossos modelos para o objetivo esperado, perder o mínimo de lesões possível. Outrossim, o modelo que apresenta a melhor performance, a VGG19, utilizou essa função objetivo durante seu treinamento. Comparando exclusivamente as métricas obtidas por esse modelo, visualizamos que houve um ganho significativo de cerca de 4 pontos percentuais na sensibilidade em relação ao modelo que que utiliza a função *Dice*.

A VGG19, enquanto backbone de melhor performance entre os testados, possui a arquitetura mais pesada em número de parâmetros, sendo 144M deles treináveis ao longo de sua estrutura, em comparação a 24M da EfficientNet v2 e 89M da ConvNext base. De modo que, para o problema abordado, redes mais profundas parecem desempenhar um melhor papel como extrator de características, pelo menos entre as arquiteturas convolucionais. Isso é positivo em termos de métrica, mas um backbone mais leve e de métricas satisfatórias pode ser relevante para ambientes onde o custo computacional e tempo de inferência é crítico.

A Figura 4 apresenta um estudo de caso demonstrando uma melhor precisão do modelo utilizando a VGG19. É possível verificar que ele consegue se adequar melhor, mesmo que levemente, ao formato da lesão real, comportamento que se repete em outros exemplares da base, corroborando esse ponto. Ao mesmo tempo, o modelo ConvNext é capaz de detectar corretamente o local onde a lesão se encontra, mesmo não sendo capaz de compreender seus detalhes. No entanto, para o especialista em ambiente clínico é relevante que se tenha uma segmentação mais precisa possível, já que isso irá auxiliar de maneira mais significativa em exames exploratórios ou em práticas cirúrgicas para a remoção da lesão.

Entre as predições dos modelos é possível identificar certa dificuldade em acertar lesões que não estejam bem centralizadas, o que pode estar ligado a diversos fatores como



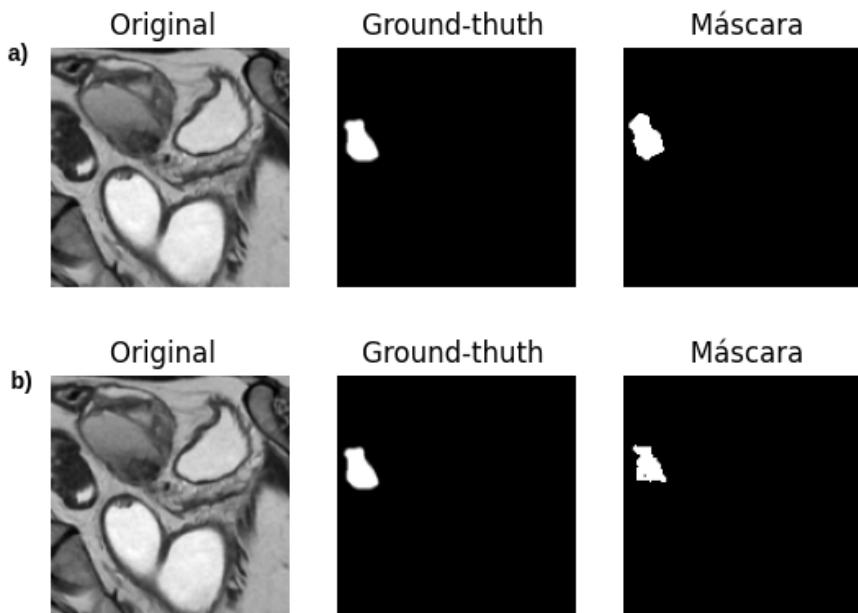
**Figura 4. Estudo de Caso entre as previsões realizadas entre o modelo a) VGG19 e b) ConvNext, ambos treinados com a Tversky Loss.**

a geração de uma ROI inapropriada, contraste desfavorável nas bordas da imagem, baixo número de exemplares com esse comportamento no conjunto de treino etc. É interessante observar que, embora a EfficientNet v2 tenha apresentado as piores métricas, ela ainda se sobressai em alguns casos, mesmo utilizando a *Dice Loss*. Na Figura 5, é possível identificar uma máscara de maior qualidade sendo predita pelo modelo de piores métricas, o que pode indicar que as diferentes arquiteturas convolucionais possuem distintos aspectos positivos e falhas, abrindo espaço para utilização de várias dessas redes de maneira complementar.

**Tabela 2. Comparação com a Literatura**

Método	Dice	Precisão	Sensibilidade	IoU
<b>Melhores resultados</b>				
Método Proposto	$50,99\% \pm 1,0$	$54,21\% \pm 1,3$	$66,58\% \pm 1,8$	$32,90\% \pm 0,7$
[Figueroedo et al. 2023]	$66,89\% \pm -$	$73,65\% \pm -$	$67,31\% \pm -$	$-\% \pm -$
[Pinto et al. 2024]	$48,07\% \pm -$	$54,10\% \pm -$	$61,70\% \pm -$	$37,20\% \pm -$

O trabalho de [Figueroedo et al. 2023] representa o estado da arte para o problema de segmentação de endometriose em imagens de RM. O método apresentado pelo autor possui um grande número de etapas, sendo capaz de atingir melhores resultados ao custo de elevar significativamente o nível de complexidade da sua abordagem. No entanto, seu método para extração da ROI é algo extremamente relevante, intensificando ainda mais o valor desta pesquisa. O trabalho de [Pinto et al. 2024] também representa uma interessante proposta sobre o tema, pois este utiliza mecanismos de atenção como forma de tornar a segmentação mais precisa. Adicionalmente, seu trabalho também foca na



**Figura 5. Estudo de Caso entre as predições realizadas entre o modelo a) EfficientNet v2 treinada com *Dice Loss* e b) VGG19 treinada com a *Tversky Loss*.**

construção de mecanismos para verificar a conectividade das predições das lesões de um mesmo paciente. Isso permite com que as falsas lesões identificadas pelo modelo sejam parcialmente removidas, aumentando assim a Precisão final dos seus resultados. Dentro deste panorama, a abordagem convolucional aqui proposta, apesar de simples, consegue apresentar resultados coerentes com aquilo que é praticado na literatura, sendo esta última ser bem escassa.

#### 4. Conclusão

Este trabalho propôs um método de segmentação de endometriose profunda em imagens de ressonância magnética utilizando redes convolucionais aliadas a um arquitetura U-net. Os resultados obtidos demonstram potencial na utilização dessas tecnologias para o auxílio na identificação de endometriose em ambiente clínico, representando uma proposta interessante nesse sentido. A utilização de outras arquiteturas, no entanto, deve ser explorada não só como maneira de enriquecer a experimentação, mas também para que se possa estudar a utilização de várias arquiteturas de segmentação trabalhando conjuntamente para refinar as predições finais. O emprego de filtros para melhora do contraste e remoção de ruído pode ser uma boa abordagem diante da atual conjuntura, possibilitando melhor evidenciação das regiões de interesse. Finalmente, é relevante colocar que a base de dados utilizada neste trabalho contém um baixo número de amostras, podendo ser um aspecto passível de melhora num momento futuro.

#### 5. Agradecimento(s)

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil – Código de Financiamento 001, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brasil, e à Fundação de Amparo à Pesquisa e

Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) (Brasil) pelo apoio financeiro.

## Referências

- Bellelis, P., Dias Jr, J. A., Podgaec, S., Gonzales, M., Baracat, E. C., and Abrão, M. S. (2010). Aspectos epidemiológicos e clínicos da endometriose pélvica: uma série de casos. *Revista da associação médica brasileira*, 56:467–471.
- Caraça, D. B., Podgaec, S., Baracat, E. C., and Abrão, M. S. (2011). Mecanismos fisiopatológicos da dor pélvica na endometriose profunda. *Diagn Tratamento*, 16(2):57–61.
- Coutinho Junior, A. C., Lima, C. M. A. d. O., Coutinho, E. P. D., Ribeiro, É. B., Aidar, M. N., and Gasparetto, E. L. (2008). Ressonância magnética na endometriose pélvica profunda: ensaio iconográfico. *Radiologia Brasileira*, 41:129–134.
- Figueredo, W. K., da Silva, I. F., Diniz, J. O., Silva, A. C., de Paiva, A. C., Salomão, A. C. B., and de Oliveira, M. A. (2023). Abordagem computacional baseada em deep learning para o diagnóstico de endometriose profunda através de imagens de ressonância magnética. In *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 138–149. SBC.
- Gomes, N. A. and Alves, K. (2018). A ressonância magnética no diagnóstico de endometriose profunda com acometimento intestinal: relato de caso. *Unilus ensino e pesquisa*, 15(38):25–39.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows.
- Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986.
- Nácul, A. P. and Spritzer, P. M. (2010). Aspectos atuais do diagnóstico e tratamento da endometriose. *Revista Brasileira de ginecologia e obstetrícia*, 32:298–307.
- Pinto, D., Figueredo, W., Silva, I., Silva, A., Paiva, A., Salomão, A., and Oliveira, M. (2024). Segmentação automática de endometriose profunda em imagens de ressonância magnética baseada em swin-unet. In *Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 471–482, Porto Alegre, RS, Brasil. SBC.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.
- Silva, J. C. R., Valerio, F. P., Herren, H., Troncon, J. K., Garcia, R., and Poli Neto, O. B. (2021). Endometriose: aspectos clínicos do diagnóstico ao tratamento. *Femina*, pages 134–141.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Tan, M. and Le, Q. V. (2020). Efficientnet: Rethinking model scaling for convolutional neural networks.

Tan, M. and Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training.