Segmentação Automática de Endometriose Profunda em Imagens de Ressonância Magnética Baseada em Swin-Unet

Daniel M. Pinto[1], Weslley K. R. Figueredo[1], Italo F. S. da Silva[1], Aristófanes C. Silva[1], Anselmo C. de Paiva[1], Alice C. C. B. Salomão[2], Marco A. P. de Oliveira[3]

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

[2]Clínica Fonte de Imagem - Rio de Janeiro, RJ - Brasil

[3]Universidade do Estado do Rio de Janeiro (UERJ) - Rio de Janeiro, RJ - Brasil

{daniel.pinto, weslley.kelson, francyles, ari, paiva}@nca.ufma.br,
{brandaosalomao, endometriose}@gmail.com

Abstract. *Deep endometriosis is the disease characterized by the presence of endometrium outside the uterine cavity, causing acute discomfort for affected individuals. Non-invasive image-based methods for assessing the degree of disease progression are effective but time-consuming for specialists. This work proposes an automatic method for segmenting endometriosis lesions in magnetic resonance images using a Swin-Unet. The method achieved a precision of* $45,6\%$*, sensitivity of* $61,9\%$*, dice of* $47,7\%$ *and jaccard of* $36,2\%$*. At least one image per patient was segmented with good quality in 17 out of 18 patients used for testing.*

Resumo. *A endometriose profunda é a doença caracterizada pela presença do endométrio fora da cavidade uterina, causando agudo desconforto para as pessoas afetadas. Métodos não invasivos baseados em imagem para a aferição do grau de evolução da doença são eficazes mas custosos em tempo dos especialistas. Este trabalho propõe um método automático de segmentação de lesões de endometriose em imagens de ressonância magnética utilizando uma Swin-Unet. O método alcançou uma precisão de* $45,6\%$*, sensibilidade* $61,9\%$*, dice de* $47,7\%$ *e jaccard de* $36,2\%$*. Foi segmentada com boa qualidade ao menos uma imagem por paciente em 17 dos 18 pacientes utilizados para teste.*

1. Introdução

A endometriose é uma doença caracterizada pela presença do endométrio, originário no útero, fora da cavidade uterina e causa lesões nos órgãos em volta. Uma das áreas comumente afetadas é a do reto e sigmoide [Schneider et al. 2016]. A doença é declarada profunda quando a lesão se estende por mais de 5 milímetros. Ela impacta a qualidade de vida e produtividade de trabalho de mulheres em idade reprodutiva, atingindo de 5% a 10% das mulheres em todo o mundo e podendo causar fortes dores pélvicas e no reto, além de infertilidade [Schneider et al. 2016, Manganaro et al. 2012].

O tratamento da doença frequentemente envolve cirurgia, portanto vê-se necessário um método preciso para diagnóstico e análise de custo-benefício para a operação

[Schneider et al. 2016]. Os métodos mais comuns para o diagnóstico de endometriose são a laparoscopia e a ressonância magnética (RM) [Manganaro et al. 2012]. Essa análise é notavelmente desafiadora pelo fato das lesões serem visualmente semelhantes ao tecido à sua volta, além de seu formato ser variável [Schneider et al. 2016, Leibetseder et al. 2022], assim demandando muito tempo de análise do especialista. Por isso, o desenvolvimento de um método automático de auxílio ao diagnóstico da endometriose é extremamente necessário.

Há muita evidência que RM é o melhor método não invasivo para avaliar o grau de presença da endometriose profunda, se aferido por um especialista experiente [Schneider et al. 2016]. Detecção e diagnóstico auxiliado por computador (CAD) auxilia o especialista a interpretar as informações de imagens médicas, identificar a localização e formato de potenciais lesões para que o profissional possa validá-las como presença da doença [Kimori 2011]. Nessa perspectiva, o aprendizado profundo tem demonstrado grande potencial em tarefas de visão computacional aplicada a imagens de RM [Lundervold and Lundervold 2019].

A Vision Transformer (ViT) é uma arquitetura de rede de aprendizado profundo introduzida por [Dosovitskiy et al. 2020]. A Swin Transformer é uma arquitetura aprimorada de ViT, proposta por [Liu et al. 2021], mais eficiente que suas antecessoras. Hoje, redes neurais baseadas em Swin Transformer fazem parte do estado-da-arte da visão computacional. O trabalho de [Cao et al. 2022] compara com outras arquiteturas de CNN e Transformer o desempenho da Swin-Unet, uma ViT baseada em Swin Transformer, para a segmentação de imagens médicas de tomografia computadorizada. [Cao et al. 2022] pré-treinaram a rede na ImageNet e demonstraram sua maior eficácia em comparação às outras arquiteturas.

No contexto da endometriose há uma carência de trabalhos que utilizem aprendizado profundo em RM. O trabalho de [Figueredo et al. 2023] apresenta uma das poucas abordagens promissoras com essa técnica para classificação automática de fatias de RM de pacientes que apresentam endometriose no reto e sigmoide, bem como pacientes saudáveis. Porém, ainda são escassos os trabalhos de segmentação da endometriose nesse tipo de imagem, baseados em aprendizado profundo.

Assim, este trabalho tem como objetivo propor um método baseado em Swin-Unet pré-treinada para a tarefa de segmentação automática da endometriose profunda na região do reto e sigmoide a partir de imagens de ressonância magnética já classificadas como possuindo lesão. O trabalho está organizado da seguinte forma: a Seção 2 detalha a base de imagens, o método proposto e as métricas de avaliação do método. A Seção 3 apresenta a análise dos resultados da segmentação automática. A Seção 4 apresenta as conclusões do trabalho.

2. Materiais e Método

Esta seção apresenta a base de imagens de RM utilizada, o método proposto de segmentação das lesões de endometriose e as métricas utilizadas para avaliá-lo. O método proposto, ilustrado na Figura 1, consiste em três etapas. Primeiro, é extraída a região de interesse (ROI) das imagens para diminuir o processamento. Em seguida, utiliza-se a Swin-Unet para segmentar as lesões. Então, é realizada uma eliminação de falsos positivos da segmentação através de uma técnica baseada em conectividade entre as segmentações dos

pacientes.

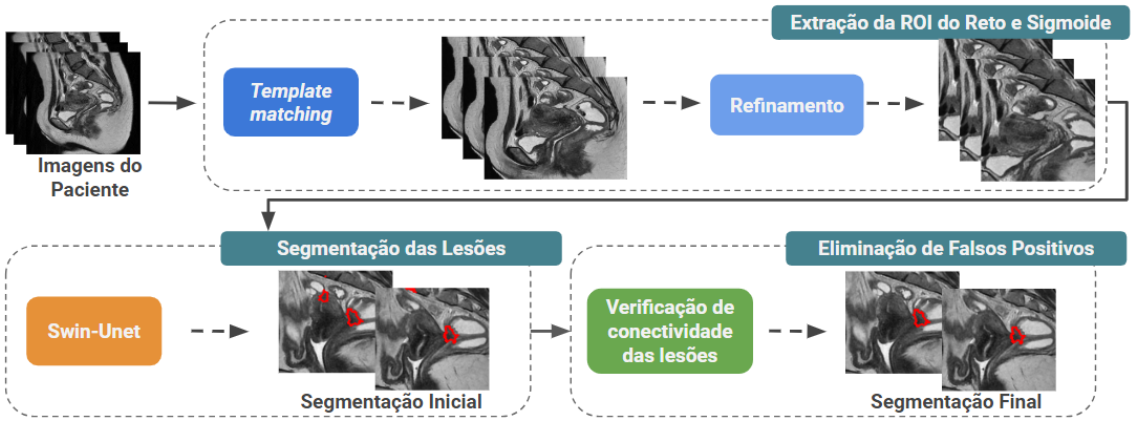


Figura 1. Esquema do método proposto.

2.1. Base de Imagens

A base de imagens é a mesma utilizada por [Figueredo et al. 2023], agora com novos volumes de pacientes que apresentam endometriose. Ela foi disponibilizada pela Clínica Fonte Imagem (Rio de Janeiro), coordenada pela Dra. Alice Brandão, especialista que realizou as marcações. A base é composta de volumes de ressonância magnética, de 512x512 pixels, no plano sagital, do abdômen inferior de 105 pacientes que apresentam lesões de endometriose.

As lesões foram marcadas manualmente pelo especialista, partindo da fatia com a maior lesão e marcando as fatias anteriores e posteriores enquanto fosse possível ver lesões, gerando máscaras binárias com as marcações. Lesões de tamanho inferior a 2 cm não foram marcadas. Para as etapas de segmentação deste trabalho foram utilizadas apenas as fatias que contêm lesão, resultando em uma base de 451 imagens, cada paciente apresentando de 1 a 9 dessas fatias. Na Figura 2 são apresentados exemplos das imagens da base.

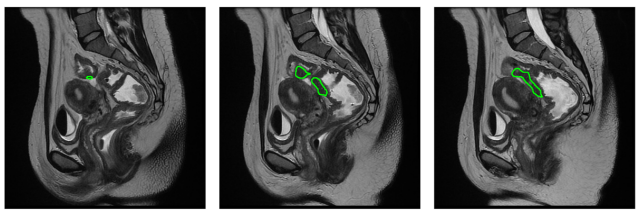


Figura 2. Fatias consecutivas de uma paciente que apresentam lesão. As marcações das lesões pelo especialista estão em verde.

2.2. Extração da Região de Interesse (ROI)

As imagens possuem muitas áreas que não agregam informação à detecção de endometriose. Portanto, a base de imagens passou por uma etapa de pré-processamento para recorte das imagens para manter apenas a região do reto e sigmoide, onde encontram-se as lesões, utilizando o método baseado em processamento de imagens proposto por

[Figueredo et al. 2023]. O primeiro passo do método é a realização de um *template matching*. O segundo passo é um refinamento para excluir da ROI a parede abdominal, que foi utilizada pelo *template* para fornecer informações posicionais.

Foi selecionada uma imagem central de um dos volumes de RM da base como *template* a ser usado em todos os pacientes. O *template matching* encontra a região mais semelhante ao *template* dentre todas as imagens do paciente. Então, o passo de refinamento mantém na ROI apenas o que está à direita da parede abdominal e padroniza seu tamanho. Binariza-se a imagem pela mediana e percorre-se da esquerda para a direita os pixels das regiões homogeneizadas, até que a seja encontrado o final do músculo abdominal, que marca o limite mais à esquerda da ROI. Finalmente, a partir do canto superior esquerdo da região definida, esta é redimensionada para 256x256 pixels. As coordenadas e dimensões da ROI então são replicadas para todas as demais imagens do paciente.

2.3. Segmentação das Lesões

Nesta etapa é realizada a segmentação das lesões de endometriose pela Swin-Unet [Cao et al. 2022] treinada na base de imagens. As transformers vêm apresentando resultados promissores em tarefas de visão computacional [Liu et al. 2021]. Elas utilizam blocos baseados em auto-atenção em sua estrutura, originalmente desenvolvidos para o processamento de linguagem natural. A imagem de entrada é dividida em recortes quadrados de mesmo tamanho, que são transformados em vetores unidimensionais e, em seguida, inseridos na rede como sequências de *tokens*. O mecanismo de auto-atenção da rede aprende a determinar a correlação desses *tokens*, tornando a ViT capaz de relacionar todas as áreas de uma imagem entre si.

A Swin Transformer é uma arquitetura mais eficiente de ViT, proposta por [Liu et al. 2021]. Seus principais componentes são os blocos *window based multi-head self-attention* (W-MSA) e *shifted window based multi-head self-attention* (SW-MSA). W-MSA realiza a auto-atenção apenas em conjuntos de recortes da imagem agrupados por proximidade, em janelas. SW-MSA utiliza a técnica *shifted windows* (janelas deslocadas) para permitir que as informações espaciais sejam comunicadas entre as janelas. Após uma sequência de W-MSA e SW-MSA, é realizada a fusão dos recortes de cada janela, enfileirados e concatenados. Arquiteturas baseadas em Swin Transformer hoje fazem parte do estado-da-arte na visão computacional, com sua vantagem sendo seu menor custo de processamento comparado às transformers anteriores. Assim, faz-se promissora a aplicação de uma dessas arquiteturas para o objetivo deste trabalho.

A Swin-Unet é uma Swin Transformer aplicada no formato de U-Net para a tarefa de segmentação, ou seja, tanto o *encoder* quanto o *decoder* típicos de uma U-Net são formados por blocos de Swin Transformer, e existem conexões de salto entre os blocos do *encoder* e os do *decoder*, bem como um gargalo conectando ambos. No *decoder*, ao invés de camadas de fusão de patches, há camadas de expansão de patches. A arquitetura da rede é apresentada na Figura 3. A saída tem dois canais, um para o fundo e um para a lesão. A versão da Swin Transformer utilizada neste trabalho é a *Tiny* [Liu et al. 2021].

[Cao et al. 2022] demonstraram alta eficácia do pré-treinamento da rede ao aplicá-la para imagens médicas, comportamento comum às ViT. Portanto, ao treinar a Swin-Unet utilizada nesta etapa do método, seus pesos foram inicializados a partir do pré-treinamento realizado na ImageNet pelos autores da rede.

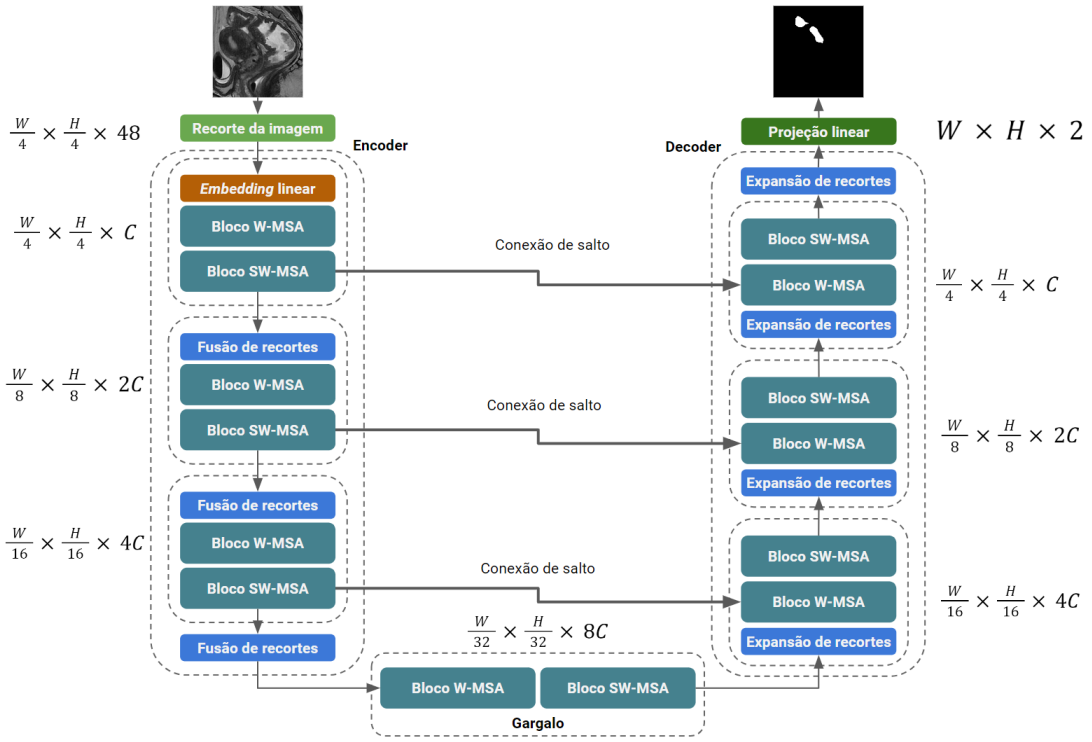


Figura 3. Arquitetura da Swin-Unet.

2.4. Eliminação de Falsos Positivos

A segmentação das lesões pode gerar falsos positivos, que são áreas cujos pixels foram marcados pela rede neural como apresentando lesão, quando na verdade não apresentam. Isso diminui a eficácia do modelo de segmentação como auxílio ao especialista, ao exibir mais áreas que este deverá analisar. Áreas de textura e cor semelhante às de lesão podem estar presentes em outras partes da imagem, levando o modelo ao erro. Assim, faz-se necessário uma etapa de eliminação de falsos positivos.

A carcterística das lesões é que elas podem se estender pelo volume da RM de um paciente em múltiplas imagens consecutivas, dado o protocolo de marcação da base pelo especialista. Assim, a área de lesão em uma imagem apresenta uma conectividade de seus pixels com áreas de lesão de imagens anteriores e posteriores. Isso é possível observar na Figura 4, onde a sobreposição das máscaras binárias das lesões demonstra a conectividade entre elas. Dessa forma, entende-se que segmentações realizadas pelo modelo em um mesmo paciente devam apresentar essa conectividade para serem consideradas corretas.

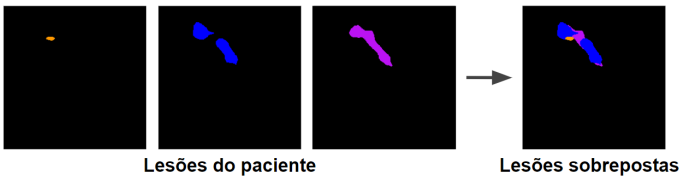


Figura 4. Máscaras binárias das imagens com lesão de um paciente e sua sobreposição.

Entretanto, deve-se ter cautela para não perder segmentações corretas de lesões dos pacientes. Um paciente que tenha muitas imagens apresentando áreas de lesão pode

ter, por exemplo, uma dessas áreas na sua imagem inicial conexa com a imagem seguinte, mas desconexa com as últimas. Além disso, o modelo treinado para a etapa de segmentação (Seção 2.3) considera apenas as informações da imagem que está processando, portanto pode segmentar corretamente regiões de lesão de forma pouco conexa entre as imagens do paciente.

Com base nessa análise, é considerada a conectividade entre segmentações de duas imagens diferentes quando há interseção de ao menos um pixel entre elas. A partir dessa definição foi estabelecido um critério de conectividade entre as segmentações para decidir quais delas manter e quais descartar. A conectividade entre as segmentações deve ocorrer através de, no mínimo, 50% das imagens do paciente, para que sejam consideradas corretas. Essas imagens não precisam ser consecutivas. O limite mínimo de 50% faz com que pacientes que possuam apenas uma ou duas imagens com lesão tenham todas as suas segmentações mantidas, já que nesses casos não há informação suficiente para determinar erros na segmentação. Além disso, todas as segmentações que sejam conexas com ao menos uma segmentação, de uma imagem adjacente à sua no volume, que adeque-se ao critério também são preservadas. As demais segmentações são descartadas.

A Figura 5 simula o comportamento da eliminação de falsos positivos em um paciente que possui 5 imagens. Os 7 objetos brancos numerados exemplificam as segmentações geradas pela rede neural. O caso A de segmentações conexas, entre a 3, 4 e 6, exemplifica segmentações que possuem conectividade com outras ao longo de no mínimo 50% das imagens do paciente, portanto são mantidas. No caso B, a segmentação 1 seria descartada por não passar no critério de conectividade, porém é conexa com a segmentação 3, na imagem imediatamente posterior à sua, portanto é mantida. As segmentações 5 e 7 são conexas com segmentações em menos de 50% das imagens, portanto são descartadas. A segmentação 2 não é conexa com nenhuma outra, então também não será mantida.

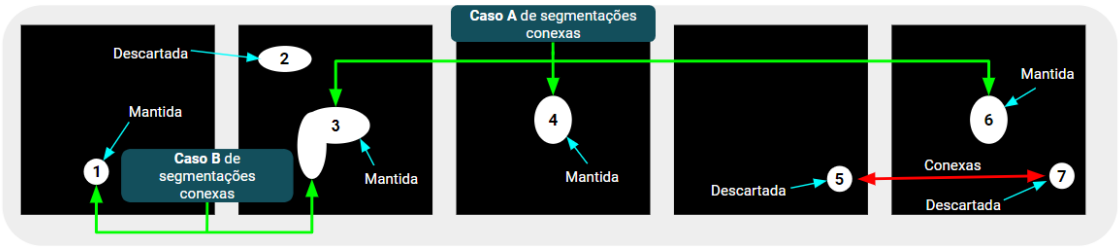


Figura 5. Simulação do comportamento da eliminação de falsos positivos. As formas brancas numeradas exemplificam segmentações da rede neural.

2.5. Métricas de Avaliação

Como resultado da segmentação, além dos falsos positivos (FP), obtemos falsos negativos (FN), que são pixels de lesão que não foram identificados pelo modelo, bem como os verdadeiros positivos (VP), que são os pixels de lesão corretamente identificados. Com esses valores, foram aferidas as seguintes métricas para avaliar a qualidade do modelo: a precisão, calculada por $\frac{VP}{VP+FP}$ e que determina a capacidade do modelo de segmentar lesões sem gerar muitos falsos positivos; a sensibilidade, calculada por $\frac{VP}{VP+FN}$ e que penaliza o modelo por não segmentar áreas de lesão; e o Índice Jaccard, calculado por

$\frac{VP}{VP+FP+FN}$ e que indica a similaridade em posição e formato das segmentações do modelo em comparação às lesões reais.

Em aplicações médicas, uma das métricas de maior importância é a sensibilidade. O risco maior ao paciente está em não encontrar áreas de lesão e a sensibilidade determina a eficácia do modelo em detectar a doença nas regiões em que ela realmente está presente. Porém, um modelo que gere muitos falsos positivos ainda pode ter uma sensibilidade alta, mas apresenta mais áreas para o especialista analisar, diminuindo a eficiência. Assim observa-se com destaque o Coeficiente Dice, calculado por $\frac{2*VP}{2*VP+FP+FN}$, uma média harmônica entre sensibilidade e precisão, portanto determinante de um modelo capaz de segmentar corretamente as lesões enquanto gera poucos falsos positivos.

3. Resultados e Discussão

Esta seção apresenta as configurações experimentais utilizadas neste trabalho e a análise dos resultados alcançados por cada etapa do método proposto.

3.1. Configurações Experimentais

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

Os hiperparâmetros de treinamento foram: tamanho de lote 4, otimizador SGD com taxa de aprendizado de 0,05 tamanho de recorte da imagem de 4x4 pixels. A função de perda utilizada foi uma combinação de Perda Cross-Entropy e Perda Dice, com 40% e 60% de peso, respectivamente, com a qual a Swin-Unet foi treinada por [Cao et al. 2022]. As imagens foram redimensionadas para 224x224 pixels, o tamanho de entrada exigido pela Swin-Unet. Foi utilizado aumento de dados *online*, ou seja, cada batch de imagens sofreu transformações aleatórias para aumentar a variabilidade do dataset, essencial para transformers [Dosovitskiy et al. 2020]. As transformações utilizadas foram rotação de até 20º, nos dois sentidos, chance de giro horizontal e vertical, até 20% de aumento ou diminuição da imagem, bem como até 20% de aumento ou diminuição de brilho. O treinamento foi realizado em uma GPU NVIDIA GeForce RTX 3060, no sistema operacional Ubuntu 20.04, utilizando o *framework* PyTorch.

3.2. Resultados da Extração da ROI do Reto e Sigmoide

A etapa de extração da ROI foi bem sucedida em manter uma área menor de 256x256 pixels para todo paciente, sem perder área das lesões. Na Figura 6 podemos observar um exemplo do resultado, com as ROIs obtidas das imagens da Figura 2.

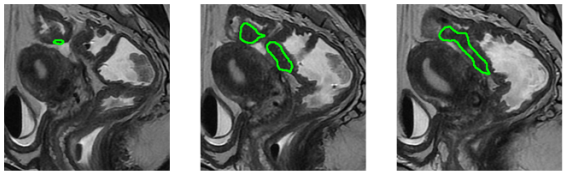


Figura 6. ROIs extraídas das imagens da Figura 2. As marcações das lesões pelo especialista estão em verde.

3.3. Resultados da Segmentação das Lesões

As métricas dos modelos treinados foram aferidas por imagem, ou seja, calculando-as para cada imagem de teste segmentada, então calculando a média de cada métrica entre todas as imagens. Para o resultado médio da *k-fold cross-validation* foram calculadas, entre as 5 *folds*, as médias das suas métricas médias por imagem. Obteve-se precisão de $46,6\% \pm 4,0\%$, sensibilidade de $57,0\% \pm 3,2\%$, dice de $46,0\% \pm 1,9\%$ e jaccard de $35,2\% \pm 1,8\%$. Esses resultados demonstram certo sucesso do modelo em segmentar as lesões da base de imagens, independentemente de quais imagens são utilizadas para treinamento e validação. Para a Swin-Unet a ser usada na etapa de segmentação de lesões, foi selecionado o melhor modelo da *fold* que alcançou o maior dice no teste, de $47,7\%$. Sua precisão foi $45,6\%$, sensibilidade $61,9\%$ e jaccard $36,2\%$.

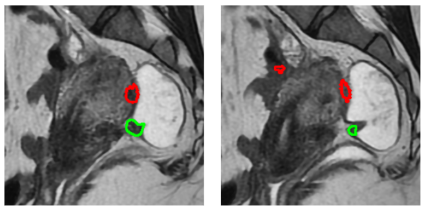
Pela característica global da análise por imagem, torna-se interessante uma análise dos resultados por paciente. Isto é feito obtendo-se as médias por imagem entre apenas as imagens do paciente, para cada paciente, e então aferindo a média das métricas entre os pacientes. A comparação dos resultados dessa perspectiva com a perspectiva por imagem são apresentados na Tabela 1. A pouca diferença em comparação com os resultados por imagem mostra que os erros e acertos do modelo estão bem distribuídos entre os pacientes, havendo para todos lesões mais difíceis e mais fáceis de segmentar.

Tabela 1. Resultado médio da predição do modelo sobre o conjunto de teste. A coluna "Análise" indica se as médias são por imagem ou por paciente.

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

É importante destacar que o modelo foi capaz de segmentar ao menos uma lesão em 17 dos 18 pacientes. Na Figura 7 é apresentado o Estudo de Caso 1, o único paciente no qual o modelo falhou em segmentar qualquer lesão. Como é possível observar, o paciente possui uma lesão pequena, visível em apenas duas imagens de seu exame, um caso extremamente desafiador para a segmentação.

O protocolo de marcação das lesões pelo especialista, descrito na Seção 2.1, é uma indicação de que se o modelo for capaz de segmentar com grande eficácia ao menos uma fatia do exame de MR do paciente, poderá tornar mais eficiente o trabalho do profissional, que irá poder encontrar a área total de lesão a partir da fatia segmentada. A eficiência obtida é maior pois o especialista não precisará buscar a lesão no volume inteiro, e poderá focar em áreas específicas das fatias indicadas. Sendo assim, é feita também a análise da melhor segmentação de cada paciente. Para todo paciente, é selecionada sua imagem



Estudo de Caso 1

Figura 7. Estudo de Caso 1. A marcação do especialista está em verde e a segmentação feita pelo modelo está em vermelho.

cuja segmentação obteve o maior dice e é calculada a média do dice apenas entre essas imagens selecionadas. Essa abordagem traz um dice médio de $72,6\%$ que estabelece uma boa eficácia do modelo em realizar uma marcação de qualidade em ao menos uma das fatias de cada paciente, a exceção sendo o Estudo de Caso 1 (Figura 7).

Para comprovar a vantagem da Swin-Unet em comparação a uma CNN para a etapa de segmentação das lesões, foram treinadas também U-Nets utilizando diferentes *backbones* comumente utilizados para segmentação: VGG-16, ResNet50, ResNet152 e Efficient Net B7. As implementações das CNNs utilizadas estão disponíveis na biblioteca *Segmentation Models (PyTorch)* [1]. Todas as U-Nets foram treinadas com os mesmos hiperparâmetros da Swin-Unet descritos na Seção 3.1, com a exceção de que as CNNs não exigiam que as imagens de entrada fossem redimensionadas para 224x224 pixels. Para cada rede neural treinada, foram aferidas a média e desvio padrão do resultado do teste do melhor modelo de cada *fold* sobre o conjunto de teste, analisando por imagem. A comparação dos resultados é apresentada na Tabela 2. Ela reforça a dificuldade do problema e demonstra que a Swin-Unet pré-treinada tem maior eficácia na tarefa, comparada a uma CNN convencional.

Tabela 2. Resultado médio entre as *folds* da *k-fold cross-validation* de cada modelo sobre o conjunto de teste. Os melhores resultados estão realçados.

Arquitetura	Precisão	Sensibilidade	Dice	Jaccard
Swin-Unet	$\mathbf{46,6\% \pm 4,0\%}$	$\mathbf{57,0\% \pm 3,2\%}$	$\mathbf{46,0\% \pm 1,9\%}$	$\mathbf{35,2\% \pm 1,8\%}$
VGG-16	$14,6\% \pm 13,1\%$	$16,1\% \pm 22,8\%$	$10,5\% \pm 11,2\%$	$6,9\% \pm 7,4\%$
Efficient Net B7	$10,6\% \pm 15,0\%$	$4,6\% \pm 6,5\%$	$5,2\% \pm 7,3\%$	$3,6\% \pm 5,1\%$
ResNet50	$1,3\% \pm 2,2\%$	$0,8\% \pm 1,2\%$	$0,7\% \pm 1,0\%$	$0,5\% \pm 0,7\%$
ResNet152	$2,9\% \pm 3,9\%$	$3,0\% \pm 5,6\%$	$1,9\% \pm 2,8\%$	$1,2\% \pm 1,8\%$

3.4. Resultados da Eliminação de Falsos Positivos

Ao aplicar a etapa de eliminação de falsos positivos nas segmentações obtidas, foram alcançados os resultados que podem ser encontrados na Tabela 3. A coluna "Sem FP" indica se foi realizada a eliminação de falsos positivos antes de aferir as métricas. Comparando os resultados com os da etapa de segmentação das lesões, é possível ver que há uma pequena melhora nas métricas, menos na sensibilidade que sofre uma leve piora. Porém, a melhora das demais métricas é de maior magnitude que a piora da sensibilidade. A média do dice das melhores imagens, conforme essa métrica, de cada paciente também aumentou 1%, de $72,6\%$ para $73,6\%$.

[1] `https://github.com/qubvel/segmentation_models.pytorch`

Tabela 3. Comparação de resultados da etapa de segmentação das lesõs com a de eliminação de falsos positivos, por tipo de análise. Os melhores resultados em cada análise estão realçados.

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

O comportamento da técnica de eliminação de falsos positivos (Seção 2.4) pode ser observado no Estudo de Caso 2 (Figura 8), que apresenta as segmentações em cinco imagens consecutivas de um paciente. Na linha "Com FP" estão os resultados da etapa da segmentação de lesões e na linha "Sem FP" estão as segmentações mantidas pela etapa de eliminação de falsos positivos. As segmentações estão numeradas, seus números indicados pelas setas azuis. As segmentações 1, 2, 3 e 8 são conexas entre si, adequando-se ao critério de conectividade em ao menos 50% das imagens. A segmentação 4 é conexa apenas com a 2, mas não é descartada pois sua imagem é adjacente à da 2 que passou no critério. As demais segmentações são consideradas falsos positivos e eliminadas, por não se enquadrarem em nenhum caso como os anteriores.

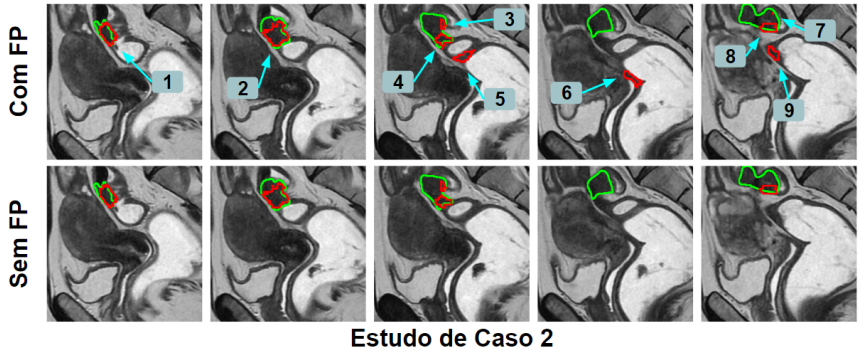


Figura 8. Estudo de Caso 2. A marcação do especialista está em verde e as segmentações estão em vermelho.

Observando os resultados da eliminação de falsos positivos em comparação com as segmentações iniciais das lesões, percebe-se que ainda é suave a eliminação de segmentações incorretas. Porém é visível que em muitos casos os falsos positivos mais desconexos e distantes da área de lesão foram removidos com sucesso. Propõe-se então uma análise dos erros e acertos por imagem, tomando como abordagem verificar se as áreas segmentadas pelo modelo têm alguma interseção com as áreas de lesão. Os resultados dessa análise podem ser encontrados na Tabela 4. Nessa tabela, as lesões que possuem interseção com as segmentações e as que não possuem são apresentadas nas colunas "Lesões Atingidas" e "Lesões Erradas", respectivamente. Além disso, na coluna "Falsos Positivos" é possível observar segmentações sem nenhuma interseção com qualquer lesão. Essa análise oferece uma perspectiva adicional do efeito da etapa de eliminação de falsos positivos.

É notável a pequena piora na sensibilidade, apesar que a grande maioria das segmentações corretas foram preservadas. As duas lesões perdidas pertencem a apenas

Tabela 4. Resultados de erro e acerto por imagem.

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

um dos pacientes, analisado no Estudo de Caso 3, apresentado na Figura 9. Nele, as lesões reais estão numeradas e indicadas pelas setas em laranja. A linha "Com FP" mostra as segmentações antes da eliminação de falsos positivos e a linha "Sem FP" exibe-as após esta etapa. É possível observar que foram perdidas as segmentações das lesões 1 e 3, devido à eliminação de falsos positivos. As segmentações geradas pelo modelo para essas lesões são muito pequenas, então não são conexas nem entre si nem com as das outras imagens. Entretanto, outras lesões do paciente foram segmentadas com maior sucesso.

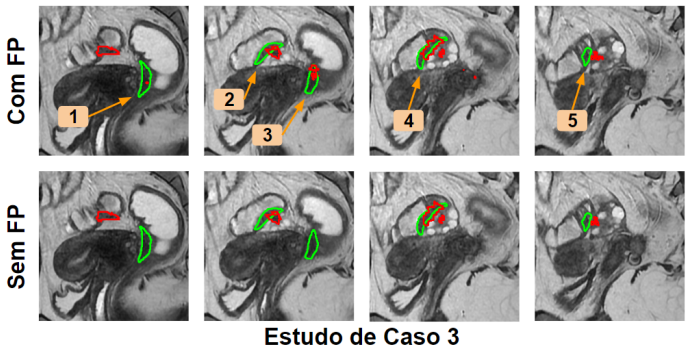


Figura 9. Estudo de Caso 3. A marcação do especialista está em verde e as segmentações estão em vermelho.

Com a etapa de eliminação de falsos positivos, muitas segmentações incorretas foram removidas. Isso deu-se ao custo de perder a segmentação de duas lesões de um paciente que teve outras lesões corretamente segmentadas. Assim, esse método de redução de falsos positivos mostra-se promissor como pós-processamento das segmentações do modelo. O resultado final do método proposto oferece mais precisão da área que deve ser avaliada pelo profissional em busca das lesões. Porém, se a etapa de segmentação das lesões gerar falhas nas segmentações, ou segmentações muito pequenas em comparação às lesões reais, estas poderão ser perdidas pela eliminação de falsos positivos.

4. Conclusão

Este trabalho aplicou a Swin-Unet para a segmentação da endometriose profunda no reto e sigmoide em imagens de RM que apresentam a doença. O método desenvolvido alcançou resultados promissores, que indicam o bom desempenho da transformer pré-treinada em tarefas mais difíceis para uma CNN convencional. A etapa de eliminação de falsos positivos promoveu uma pequena melhora deste resultado. O método completo foi capaz de segmentar com grande sucesso ao menos uma lesão em 17 dos 18 pacientes. Tendo em vista a dificuldade da segmentação manual das lesões esses resultados indicam o potencial para um sistema de auxílio ao diagnóstico.

A segmentação automática da endometriose em RM é uma tarefa desafiadora e ainda há uma escassez de soluções utilizando aprendizado profundo. Como trabalhos

futuros, pretende-se avaliar o desempenho do método em um cenário com imagens sem lesão inclusas. Também devem ser estudadas outras formas de aproveitar a correlação entre as lesões do paciente, fornecendo para a rede informações da imagem anterior e posterior à sendo processada. Além disso, é interessante avaliar a eficácia de alterações arquiteturais da Swin-Unet na tarefa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 , Fundação de Amparo à Pesquisa do Maranhão (FAPEMA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Empresa Brasileira de Serviços Hospitalares (Ebserh) Brazil (Proc. 409593/2021-4)

Referências

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M. (2022). Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Figueredo, W., Silva, I., Diniz, J., Silva, A., Paiva, A., Salomão, A., and Oliveira, M. (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, Porto Alegre, RS, Brasil. SBC.

Kimori, Y. (2011). Mathematical morphology-based approach to the enhancement of morphological features in medical images. *Journal of clinical bioinformatics*, 1:1–10.

Leibetseder, A., Schoeffmann, K., Keckstein, J., and Keckstein, S. (2022). Endometriosis detection and localization in laparoscopic gynecology. *Multimedia Tools and Applications*, 81(5):6191–6215.

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. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022.

Lundervold, A. S. and Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127.

Manganaro, L., Fierro, F., Tomei, A., Irimia, D., Lodise, P., Sergi, M., Vinci, V., Sollazzo, P., Porpora, M., Delfini, R., et al. (2012). Feasibility of 3.0 t pelvic mr imaging in the evaluation of endometriosis. *European journal of radiology*, 81(6):1381–1387.

Schneider, C., Oehmke, F., Tinneberg, H.-R., and Krombach, G. (2016). Mri technique for the preoperative evaluation of deep infiltrating endometriosis: current status and protocol recommendation. *Clinical Radiology*, 71(3):179–194.

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern recognition*, 48(9):2839–2846.