

Ensemble Ponderado de Deep Learning para Segmentação de Lesões Mamárias em Ultrassonografia

**Neilson P. Ribeiro^{1,2}, Celso L. S. Soares Filho², Felipe R. S. Teles²,
Marcos R. A. Amorim², João O. B. Diniz^{1,2}, Anselmo C. de Paiva²,
Aristófanés C. Silva², Antonio O. C. Filho³**

¹Fábrica de Inovação – Instituto Federal do Maranhão (IFMA)
Maranhão – Brasil

²Núcleo de Computação Aplicada (NCA) – Universidade Federal do Maranhão (UFMA)
São Luís – Maranhão – Brasil

³Universidade Federal do Piauí –(UFPI)
Picos – PI – Brasil

neilson.ribeiro@ifma.edu.br

Abstract. *Breast cancer is one of the leading causes of mortality among women worldwide, making early diagnosis essential to improve survival rates. In this context, automatic lesion segmentation in ultrasound images remains a challenging task due to speckle noise, low contrast, and high morphological variability. This work proposes a weighted ensemble approach combining U-Net++, DeepLabV3+, and Swin-UNet architectures, leveraging complementary features to enhance prediction robustness. Evaluated on the BUS-BRA (INCA) dataset, the proposed method achieved an IoU of 85.62% and a Dice score of 93.17%, outperforming previously reported results in the literature. The findings demonstrate the potential of the proposed ensemble strategy as a clinical decision-support tool.*

Resumo. *O câncer de mama é uma das principais causas de mortalidade entre mulheres no mundo, tornando o diagnóstico precoce fundamental para o aumento das taxas de sobrevivência. Nesse contexto, a segmentação automática de lesões em imagens de ultrassonografia constitui uma tarefa desafiadora devido ao ruído speckle, baixo contraste e alta variabilidade morfológica. Este trabalho propõe uma abordagem baseada em ensemble ponderado das arquiteturas U-Net++, DeepLabV3+ e Swin-UNet, explorando características complementares para aumentar a robustez das predições. Avaliado na base BUS-BRA (INCA), o método alcançou IoU de 85,62% e Dice de 93,17%, superando resultados previamente reportados na literatura. Os resultados indicam que a estratégia de ensemble melhora a qualidade da segmentação, mostrando potencial como ferramenta de apoio clínico.*

1. Introdução

O câncer de mama (CM) é o tumor mais incidente entre mulheres e uma das principais causas de mortalidade feminina em todo o mundo. Em 2022, foram registrados aproximadamente 2,3 milhões de novos casos e 670 mil óbitos, segundo estimativas do

GLOBOCAN [Kim et al. 2025]. A detecção precoce é um fator decisivo para a redução da mortalidade, pois possibilita intervenções terapêuticas em estágios iniciais da doença, aumentando significativamente as taxas de sobrevivência [Diniz et al. 2024].

Entre os métodos de imagem utilizados no rastreamento e diagnóstico do CM, destacam-se a mamografia e a ultrassonografia (US). A US mamária é amplamente empregada como exame complementar, especialmente em pacientes com mamas densas, por ser um método não invasivo, livre de radiação e de menor custo operacional [Sood et al. 2019]. Entretanto, as imagens de US apresentam desafios significativos, como ruído *speckle*, baixo contraste e bordas pouco definidas, fatores que dificultam a delimitação precisa das lesões e aumentam a variabilidade interobservador [Litjens et al. 2017].

Nesse contexto, sistemas de diagnóstico assistido por computador (CAD) baseados em aprendizado profundo têm sido desenvolvidos para auxiliar radiologistas na interpretação das imagens, reduzindo a subjetividade e aumentando a reprodutibilidade dos resultados [Gonçalves et al. 2024]. Dentre as tarefas envolvidas, a segmentação automática das lesões desempenha papel central, pois permite a delimitação estruturada das regiões tumorais, fornecendo subsídios quantitativos relevantes para a análise clínica e para o desenvolvimento de modelos preditivos [Teles et al. 2025].

Considerando essa problemática e as evidências de que estratégias baseadas em *ensemble* podem melhorar o desempenho em tarefas envolvendo imagens de US [Diniz et al. 2024], o objetivo deste trabalho é desenvolver e avaliar um método automático para segmentação de lesões mamárias em imagens de US baseado em uma abordagem de segmentação por *ensemble* de aprendizado profundo. O método proposto apresenta as seguintes contribuições:

- Uma estratégia baseada na integração de múltiplas arquiteturas profundas (UNet++, DeepLabV3+ e Swin-UNet), explorando suas características complementares para aumentar a robustez da segmentação;
- Um método automático de segmentação incluindo etapa de refinamento morfológico para melhorar a coerência espacial das máscaras;

2. Trabalhos Relacionados

Esta seção apresenta os trabalhos mais relevantes e recentes relacionados à segmentação automática de lesões de câncer de mama em imagens de US. Todos os estudos discutidos foram conduzidos utilizando a base BUS-BRA [Gómez-Flores et al. 2024], um conjunto de dados brasileiro composto por exames coletados no contexto do sistema público de saúde. O uso dessa base garante que as características das imagens estejam alinhadas à realidade clínica nacional, fortalecendo a relevância e aplicabilidade dos métodos no contexto do Sistema Único de Saúde (SUS). São destacadas, a seguir, as principais técnicas empregadas e as métricas de desempenho reportadas na literatura.

Em [Zhang et al. 2024], é introduzido o modelo DDTransUNet, que combina camadas convolucionais profundas com módulos *Transformer*, visando capturar simultaneamente características locais e dependências globais. No contexto do BUS-BRA, os autores reportam acurácia de 98% e coeficiente *Dice* de 88,23%, evidenciando a eficácia da abordagem híbrida CNN e *Transformer* para a segmentação de lesões mamárias com contornos mais complexos.

O trabalho [Qiu et al. 2025] propõe a *Lightweight U-Net (LBA-Net)*, uma arquitetura convolucional otimizada com foco na redução da complexidade computacional. Avaliada na base de dados BUS-BRA, a abordagem apresentou coeficiente *Dice* de 88,88% e *IoU* de 81,26%, indicando desempenho consistente na delimitação das regiões lesionadas, mesmo com um modelo de menor porte e sem o uso explícito de estratégias de *ensemble* ou refinamento morfológico.

[Wei et al. 2025] apresentam uma abordagem baseada em *State Space Models*, utilizando as arquiteturas BS-Mamba e SAM-VMamba. Os resultados obtidos no BUS-BRA indicam coeficiente *Dice* de 90,25% e *IoU* de 82,54%, com destaque para a capacidade do modelo em capturar dependências de longo alcance e preservar a coerência estrutural das máscaras segmentadas em cenários de maior variabilidade morfológica.

O trabalho de [Guo et al. 2025] propõe a MSRA-Net, uma arquitetura convolucional *multi-scale* com mecanismos *region-aware*, desenvolvida para aprimorar a delimitação de regiões lesionadas em imagens de US mamária. Avaliada na base de dados BUS-BRA, a abordagem alcançou coeficiente *Dice* de 89,83% e *IoU* de 83,08%, demonstrando desempenho competitivo na segmentação de lesões com diferentes padrões morfológicos. Os resultados evidenciam a eficácia da exploração de informações em múltiplas escalas e do reforço de regiões relevantes, contribuindo para uma segmentação mais precisa e consistente nesse conjunto de dados.

A análise da literatura indica que, apesar do uso crescente da base BUS-BRA em tarefas de segmentação, apenas poucos trabalhos reportam métricas quantitativas completas, como *IoU* e *Dice*, de forma explícita em experimentos conduzidos exclusivamente com esse conjunto de dados. Esse cenário está relacionado ao fato de o BUS-BRA ser uma base recente, o que ainda limita o número de estudos comparativos consolidados.

Dessa forma, observa-se que os trabalhos na base BUS-BRA exploram predominantemente arquiteturas individuais, sem integração entre modelos. Em contraste, este estudo propõe um *ensemble* ponderado que explora a complementaridade entre U-Net++, DeepLabV3+ e Swin-UNet, integrando mecanismos de extração local de características (CNN) e modelagem de dependências globais (*Transformer*). Essa estratégia visa aumentar a robustez, a estabilidade e a consistência espacial das máscaras segmentadas.

3. Materiais e Métodos

Nesta seção, é descrito o método proposto. De forma geral, o método é composto por cinco etapas, conforme ilustrado na Figura 1. A seguir, cada etapa é apresentada detalhadamente.

3.1. Materiais

Para a avaliação do método proposto, foi utilizada a base de dados pública BUS-BRA [Gómez-Flores et al. 2024], disponibilizada pelo Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). O conjunto é composto por 1.875 imagens de ultrassonografia mamária, provenientes de 1.064 pacientes, todos com lesões confirmadas por exame histopatológico, divididas em 1.268 imagens benignas e 607 imagens malignas. Além disso, inclui máscaras de segmentação manual das regiões tumorais, cujas marcações foram realizadas por um especialista em US. A Figura 2. apresenta exemplos representativos das imagens e suas respectivas máscaras.

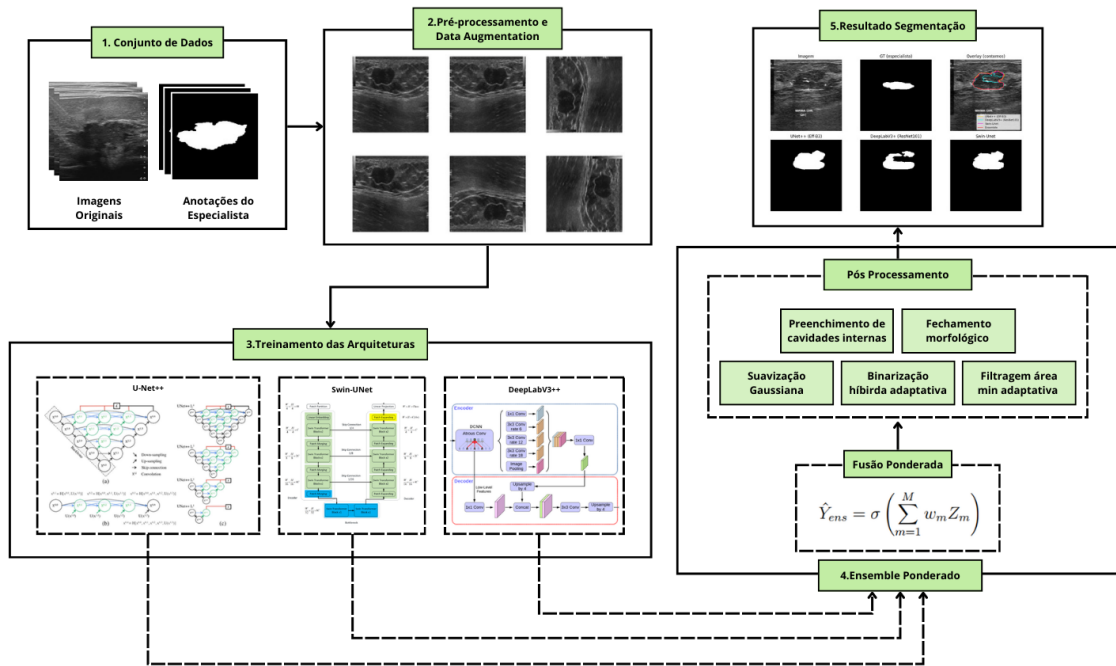


Figura 1. Fluxograma do método proposto.

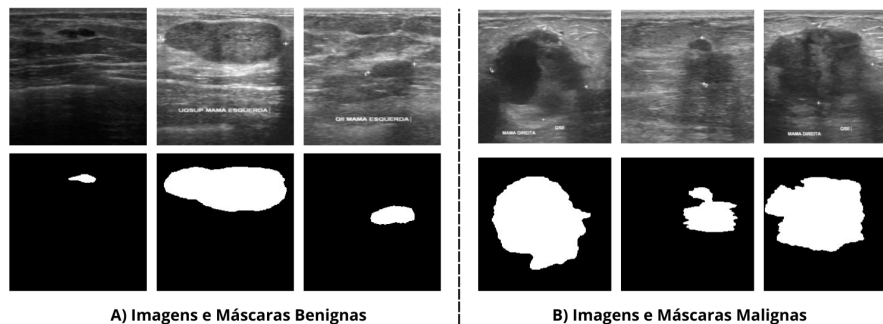


Figura 2. Exemplos de imagens e máscaras de segmentação da base BUS-BRA.

3.2. Pré-processamento e *Data Augmentation*

O pré-processamento foi realizado para padronizar as imagens de entrada e garantir compatibilidade com as arquiteturas de segmentação. As imagens, originalmente em escala de cinza e com resoluções variadas, foram redimensionadas para 224×224 pixels. Essa resolução foi adotada por ser a configuração padrão utilizada no treinamento de modelos pré-treinados na base *ImageNet*.

As máscaras de referência foram binarizadas, caracterizando o problema como uma tarefa de segmentação binária, na qual o objetivo consiste em distinguir a região da lesão do fundo. Essa etapa assegura o alinhamento entre a saída dos modelos e o formato das anotações manuais realizadas por especialistas, permitindo uma avaliação coerente do desempenho das arquiteturas.

O aumento de dados foi empregado exclusivamente durante o treinamento com o objetivo de ampliar a variabilidade amostral e reduzir o risco de sobreajuste. Inicialmente, cada imagem foi submetida a seis transformações geométricas fixas: identidade,

espelhamento horizontal, espelhamento vertical e rotações de 90°, 180° e 270°.

Além das transformações geométricas, foram aplicadas transformações probabilísticas exclusivamente no conjunto de treinamento com o objetivo de simular variações comuns em imagens de US. Essas operações incluíram ajustes de brilho e contraste, equalização adaptativa de histograma (*CLAHE*), adição de ruído Gaussiano, desfoque de movimento e deformações elásticas. Estratégias desse tipo são amplamente utilizadas em tarefas de segmentação médica para ampliar a variabilidade dos dados e melhorar a capacidade de generalização dos modelos [Ronneberger et al. 2015].

3.3. Treinamento das Arquiteturas

Foram empregadas três arquiteturas de segmentação: U-Net++ [Zhou et al. 2018], DeepLabV3+ [Chen et al. 2018] e Swin-UNet [Cao et al. 2022], descritas em detalhes nas subseções seguintes.

Os modelos de segmentação foram treinados por 50 épocas, com tamanho de lote igual a 16 amostras, utilizando o otimizador *AdamW*, com taxa de aprendizado inicial de 1×10^{-4} e regularização L2 de mesma magnitude. A escolha do *AdamW* deve-se à sua capacidade de desacoplar a regularização do processo de atualização dos pesos, contribuindo para uma melhor estabilidade durante o treinamento. A taxa de aprendizado foi ajustada dinamicamente por meio do *Cosine Annealing Scheduler*, promovendo uma redução gradual ao longo das épocas e evitando oscilações abruptas no gradiente.

A função de perda adotada foi a *Focal Tversky Loss*, uma métrica especialmente adequada para cenários de segmentação com forte desbalanceamento entre as classes [Abraham and Khan 2019].

3.3.1. Arquiteturas de Segmentação Utilizadas

A escolha das arquiteturas de segmentação que compõem o *ensemble* foi orientada pela necessidade de capturar, de forma complementar, diferentes características presentes em imagens de US mamária, como bordas pouco definidas, variações intensas de contraste, ruído e padrões morfológicos distribuídos em múltiplas escalas espaciais.

A primeira arquitetura utilizada no *ensemble* é a U-Net++, proposta por [Zhou et al. 2018], com *encoder EfficientNet-B3* foi selecionada devido à sua capacidade de refinamento progressivo das representações espaciais por meio de conexões densas e supervisão profunda. Essa arquitetura favorece a preservação de detalhes locais e a delimitação precisa das bordas da lesão, sendo particularmente eficaz na segmentação de estruturas pequenas ou com contornos irregulares, comuns em imagens de US.

A segunda arquitetura integrante do *ensemble* é a DeepLabV3+, proposta por [Chen et al. 2018], com backbone ResNet-101 foi incorporada ao *ensemble* por sua habilidade em capturar contexto multiescalar através do uso de convoluções dilatadas no módulo *ASPP (Atrous Spatial Pyramid Pooling)*. Esse mecanismo permite ao modelo integrar informações de diferentes campos receptivos sem perda de resolução espacial, tornando-o adequado para a segmentação de regiões extensas e para a preservação da forma global da lesão, mesmo em cenários de baixo contraste.

O terceira arquitetura do *ensemble* é o Swin-UNet, proposto pelo [Cao et al. 2022], por sua vez, introduz mecanismos de atenção hierárquica baseados em janelas deslocadas, permitindo a modelagem eficiente de dependências espaciais de longo alcance. Essa característica é particularmente relevante em imagens de US, nas quais a distinção entre tecido saudável e lesão pode depender de padrões distribuídos globalmente e não apenas de informações locais.

Essa diversidade arquitetural favorece o aprendizado de representações complementares, reduzindo vieses individuais dos modelos e contribuindo para a obtenção de segmentações mais estáveis, consistentes e alinhadas a marcação manual realizada por especialistas clínicos.

3.4. Ensemble Ponderado

Para tornar a segmentação mais estável e menos sensível às limitações de um único modelo, adotou-se uma estratégia de *ensemble* baseada na combinação de diferentes arquiteturas de segmentação treinadas de forma independente. Cada modelo tende a responder melhor a determinados padrões presentes nas imagens de US, como variações locais de borda, estruturas em diferentes escalas ou relações espaciais mais amplas. A integração dessas predições permite explorar essas complementaridades, resultando em máscaras finais mais consistentes.

A fusão das predições é realizada no espaço dos *logits*, ou seja, utilizando as saídas diretas dos modelos antes da conversão em probabilidades. Os pesos atribuídos a cada arquitetura foram calculados a partir do melhor *Dice Score* obtido no conjunto de validação, sendo normalizados pela soma dos *Dice Scores* dos três modelos. Dessa forma, arquiteturas com melhor desempenho de validação contribuirão proporcionalmente mais para o *ensemble*. Após a combinação ponderada dos *logits*, aplica-se a função sigmoide para converter o resultado em probabilidades no intervalo $[0, 1]$.

As saídas são combinadas conforme:

$$\hat{Y}_{ens} = \sigma \left(\sum_{m=1}^M w_m Z_m \right),$$

em que Z_m representa o mapa de *logits* do modelo m , w_m é o peso normalizado obtido a partir do melhor *Dice Score* de validação do respectivo modelo e $\sigma(\cdot)$ é a função sigmoide.

3.4.1. Pós processamento

Embora a estratégia de *ensemble* aumente a estabilidade das segmentações, as máscaras resultantes ainda podem apresentar ruídos e irregularidades típicas das imagens de US. Para reduzir essas imperfeições e garantir maior coerência morfológica, foi desenvolvido um módulo de pós-processamento morfológico adaptativo, cujo objetivo é regularizar a máscara final e assegurar maior coerência estrutural.

O processo é composto por cinco etapas sequenciais baseadas em técnicas clássicas de processamento digital de imagens [Gonzalez and Woods 2000]: Suavização

Gaussiana: atenua ruídos de alta frequência e suaviza as bordas da lesão; Binarização híbrida adaptativa: combina o limiar global de *Otsu* com um limiar base ajustável; Filtragem por área mínima adaptativa: remove regiões espúrias por meio da análise de componentes conectados; Preenchimento de cavidades internas: elimina vazios artificiais através de *flood fill*; Fechamento morfológico elíptico: garante continuidade e regularidade ao contorno da lesão.

Os parâmetros do pós-processamento foram definidos por meio de uma varredura em grade (*grid search*) no conjunto de validação.

4. Resultados e Discussão

Nesta seção, são apresentados e discutidos os resultados obtidos com a aplicação do método proposto para segmentação de lesões mamárias na base BUS-BRA. Inicialmente, descreve-se o ambiente de treinamento adotado. Em seguida, são detalhadas as métricas alcançadas pelas arquiteturas individuais, pela estratégia de *ensemble* e pelo módulo de pós-processamento.

Os dados foram divididos em 80% para treinamento, 10% para validação e 10% para teste, por meio de estratificação no nível do paciente, garantindo que imagens do mesmo indivíduo não fossem distribuídas entre os subconjuntos e evitando vazamento de dados. Optou-se por uma divisão fixa, em vez de validação cruzada (*k-fold*), devido ao elevado custo computacional associado ao treinamento de um *ensemble* composto por três arquiteturas distintas.

Para validação do método, foram empregadas as métricas Acurácia, Precisão, Sensibilidade, *IoU* e *Dice*, que avaliam o desempenho global e a qualidade da sobreposição entre as máscaras preditas e o *ground truth*. Posteriormente, realiza-se a comparação com trabalhos da literatura que utilizam a mesma base de dados. Por fim, são analisados estudos de caso qualitativos, destacando os pontos fortes e as limitações da abordagem proposta.

4.1. Ambiente de Treinamento

O ambiente de experimentação foi configurado com recursos computacionais e bibliotecas específicas para suportar a implementação e validação do método proposto, foram utilizados os seguintes recursos de *hardware*: processador Intel® Core™ i7 CPU 2.30GHz, GPU Nvidia® RTX-3070 com 8GB, 16GB de RAM e o sistema operacional Windows 11 Pro. A implementação foi realizada em *Python*, utilizando as bibliotecas *PyTorch*, *NumPy*, *OpenCV*, *Matplotlib* e *Pandas* para modelagem, processamento, análise e visualização dos dados.

4.2. Métricas sem Aumento de Dados e sem Pós-processamento

Na primeira avaliação, sem aumento de dados e sem pós-processamento, as três arquiteturas apresentaram desempenho consistente na segmentação da base BUS-BRA. Entre os modelos individuais, o Swin-UNet obteve os melhores resultados, com *IoU* de 76,80% e *Dice* de 86,95%, seguido pelo DeepLabV3+ e pelo U-Net++.

A estratégia de *ensemble* superou todas as redes individuais, alcançando *IoU* de 78,56% e *Dice* de 87,28%, indicando que a combinação das predições contribui para maior estabilidade e melhor delimitação das regiões lesionadas, mesmo na ausência de técnicas adicionais de refinamento.

Rede	Acurácia (%)	Precisão (%)	Sensibilidade (%)	IoU (%)	Dice (%)
Swin-Unet	97,48	86,55	86,10	76,80	86,95
DeepLabV3+ (ResNet101)	97,28	84,90	86,12	75,32	85,50
U-Net++ (EffB3)	97,15	83,80	85,45	74,25	84,88
Ensemble	97,60	87,12	87,45	78,56	87,28

Tabela 1. Métricas sem aumento de dados e sem pós-processamento na base BUS-BRA.

4.2.1. Métricas com Aumento de Dados

Com a inclusão do aumento de dados durante o treinamento, foram aplicadas transformações geométricas e probabilísticas. Essas operações foram aplicadas exclusivamente ao conjunto de treinamento, visando ampliar a variabilidade amostral e reduzir o risco de sobreajuste.

Rede	Acurácia (%)	Precisão (%)	Sensibilidade (%)	IoU (%)	Dice (%)
Swin-Unet	98,21	89,45	90,12	81,15	89,82
DeepLabV3+ (ResNet101)	98,02	87,55	89,40	79,85	88,46
U-Net++ (EffB3)	97,92	86,90	89,15	79,12	88,05
Ensemble	98,32	90,05	90,56	82,14	90,30

Tabela 2. Métricas com aumento de dados na base BUS-BRA.

Os resultados da Tabela 2 evidenciam melhora consistente em todas as arquiteturas quando comparadas ao Cenário 1. O Swin-Unet apresentou *IoU* de 81,15% e *Dice* de 89,82%, enquanto o ensemble alcançou *IoU* de 82,14% e *Dice* de 90,30%, superando os modelos individuais. Esse comportamento indica que o aumento de dados contribuiu para maior robustez do modelo frente às variações espaciais e estruturais das lesões presentes na base BUS-BRA.

4.2.2. Métricas com Pós-processamento

Por fim, foi aplicada uma etapa de pós-processamento morfológico adaptativo às máscaras geradas pelas redes e pelo *ensemble*. Mesmo após a fusão ponderada, as predições podem apresentar ruídos residuais, bordas irregulares e pequenas discontinuidades estruturais, características comuns em imagens de US. Para mitigar essas imperfeições, foi implementado um módulo de pós-processamento conforme descrito na Seção 3.4.1.

Rede	Acurácia (%)	Precisão (%)	Sensibilidade (%)	IoU (%)	Dice (%)
Swin-Unet	98,68	91,95	92,85	84,70	92,45
DeepLabV3+ (ResNet101)	98,48	90,30	92,25	83,15	91,26
U-Net++ (EffB3)	98,40	89,82	92,05	82,75	90,92
Ensemble (Final)	98,78	92,50	93,85	85,62	93,17

Tabela 3. Métricas com pós-processamento morfológico na base BUS-BRA.

Os resultados da Tabela 3 demonstram melhora consistente em todas as arquiteturas após o refinamento morfológico. O Swin-Unet alcançou *IoU* de 84,70% e *Dice*

de 92,45%, enquanto o *ensemble* final atingiu *IoU* de 85,62% e *Dice* de 93,17%, superando os cenários anteriores. Esses ganhos indicam que o pós-processamento contribuiu para regularizar os contornos das lesões, reduzir regiões espúrias e aumentar a coerência estrutural das máscaras segmentadas.

4.2.3. Comparação com Trabalhos Relacionados

A comparação com os trabalhos relacionados evidencia que o método proposto apresenta desempenho superior na base BUS-BRA, especialmente em termos de *IoU* e *Dice*. Enquanto o DDTransUNet [Zhang et al. 2024] combina CNNs e Transformers para capturar dependências locais e globais, e o LBA-Net [Qiu et al. 2025] prioriza eficiência computacional, ambos operam como arquiteturas individuais, sem estratégias explícitas de fusão ou refinamento morfológico. De forma semelhante, as abordagens baseadas em *State Space Models* [Wei et al. 2025] e a MSRA-Net [Guo et al. 2025] exploram mecanismos avançados de modelagem espacial e multiescala, porém ainda dependem de uma única arquitetura para geração da máscara final.

Conforme apresentado na Tabela 4, o método proposto alcançou *IoU* de 85,62% e *Dice* de 93,17%, superando os melhores resultados previamente reportados na literatura para a BUS-BRA. Em comparação à MSRA-Net [Guo et al. 2025], que obteve 83,08% de *IoU* e 89,83% de *Dice*, observa-se melhora consistente nas duas métricas de sobreposição. De maneira similar, o *Dice* obtido também supera o maior valor anteriormente reportado em [Wei et al. 2025].

Trabalho	Técnica	Acurácia (%)	Precisão (%)	Sensibilidade (%)	IoU (%)	Dice (%)
[Zhang et al. 2024]	DDTransUNet (CNN + Transformer)	98,00	–	–	80,75	88,23
[Qiu et al. 2025]	Lightweight U-Net (LBA-Net)	–	–	–	81,26	88,88
[Wei et al. 2025]	BS-Mamba / SAM-VMamba	98,58	96,19	85,43	82,54	90,25
[Guo et al. 2025]	CNN multi-scale region-aware	98,50	–	–	83,08	89,83
Método Proposto	Ensemble ponderado + pós-processamento	98,78	92,50	93,85	85,62	93,17

Tabela 4. Comparação de desempenho na base BUS-BRA.

O diferencial do método proposto, ilustrado na Figura 1, está na integração estruturada de múltiplas arquiteturas complementares (U-Net++, DeepLabV3+ e Swin-Unet) por meio de um *ensemble* ponderado em nível probabilístico, seguido de um módulo de pós-processamento morfológico adaptativo. Essa estratégia combina representações convolucionais profundas e multiescala com maior capacidade de modelagem global, reduzindo a variabilidade entre predições individuais e aumentando a robustez do sistema.

Além disso, o pós-processamento contribui para refinar as máscaras ao atenuar ruídos residuais, remover regiões espúrias, preencher cavidades internas e regularizar contornos, promovendo maior coerência estrutural e melhor alinhamento com o *ground truth* (GT). Assim, os resultados indicam que a combinação de *ensemble* e refinamento morfológico constitui uma abordagem mais robusta do que o uso isolado de arquiteturas individuais na base BUS-BRA.

4.3. Estudos de Casos

A Figura 3 apresenta uma lesão benigna com bordas relativamente bem definidas no (GT). As três arquiteturas demonstram alta concordância estrutural, com pequenas variações

periféricas, especialmente em regiões laterais e inferiores. Enquanto U-Net++ e DeepLabV3+ apresentam discretas variações de borda, a Swin-Unet preserva melhor a forma global. O ensemble proposto, integra as predições e aplica refinamento morfológico, resultando em melhor alinhamento com o GT.

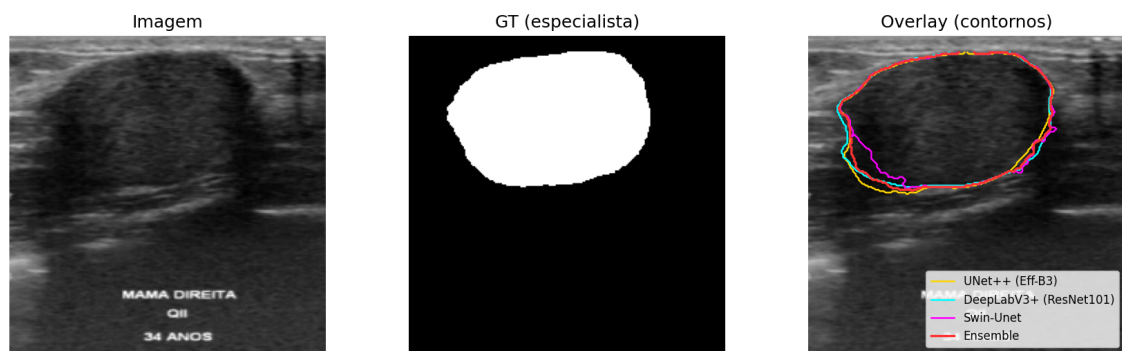


Figura 3. Estudo de Caso 1: segmentações U-Net++, DeepLabV3+, Swin-Unet e *ensemble* proposto

Conforme mostrado na Figura 4, o Estudo de Caso 2 apresenta uma lesão maligna com bordas pouco definidas, tornando a segmentação mais desafiadora. Observa-se maior divergência entre os modelos individuais. Embora o Swin-Unet apresente melhores métricas gerais, neste caso houve fragmentação da máscara. O U-Net++ e o DeepLabV3+ capturaram a estrutura principal com maior continuidade, porém com pequenas imprecisões. O *ensemble* ponderado integrou as predições e produziu uma segmentação mais próxima da anotação do especialista.

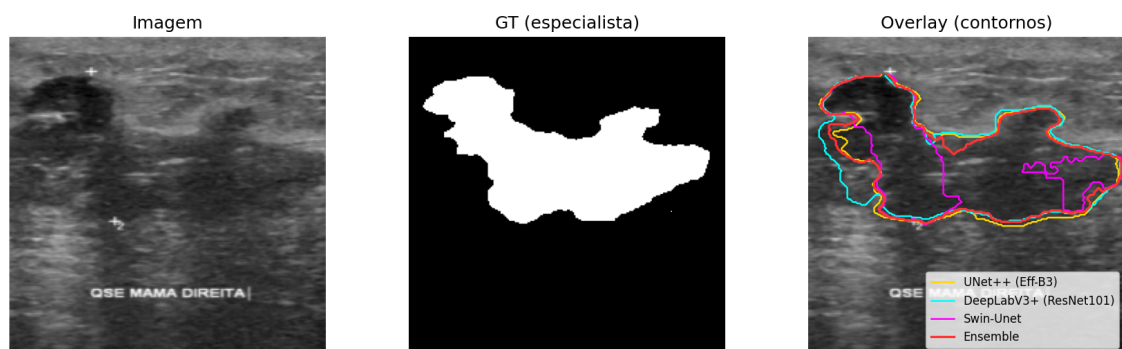


Figura 4. Estudo de Caso 2: Variabilidade de segmentação em lesão maligna com padrão estrutural complexo

A análise qualitativa dos estudos de caso, contemplando lesões benignas e malignas com diferentes graus de complexidade morfológica, demonstra que os modelos individuais apresentam comportamentos complementares, com variações na delimitação de bordas e na preservação estrutural das máscaras. A estratégia de *ensemble* ponderado mostrou-se mais estável e consistente, reduzindo fragmentações, supersegmentações e perdas regionais observadas nas arquiteturas isoladas. Esses resultados qualitativos corroboram os achados quantitativos apresentados anteriormente, evidenciando maior robustez do método proposto frente à variabilidade estrutural presente na base BUS-BRA.

5. Conclusão e Trabalhos Futuros

A segmentação automática de lesões mamárias em imagens de US permanece um desafio relevante devido à presença de ruído, baixo contraste, sombra acústica posterior e elevada variabilidade morfológica das lesões. Neste trabalho, foi proposto um método completo para segmentação automática na base BUS-BRA, combinando pré-processamento, múltiplas arquiteturas de aprendizado profundo (U-Net++, DeepLabV3+ e Swin-UNet), estratégia de *ensemble* ponderado e pós-processamento morfológico adaptativo. A integração dessas etapas permitiu reduzir inconsistências individuais dos modelos, melhorar a aderência espacial das máscaras ao GT e aumentar a estabilidade das segmentações. Os resultados quantitativos e qualitativos evidenciam o potencial da abordagem como ferramenta de apoio ao especialista na análise de exames de ultrassonografia mamária.

Como trabalhos futuros, pretende-se avaliar o método em outras bases de dados, investigando sua capacidade de generalização em exames provenientes de diferentes equipamentos e protocolos de aquisição. Além disso, propõe-se explorar estratégias de ponderação adaptativa no *ensemble*, considerando critérios dinâmicos baseados em incerteza preditiva ou desempenho local das segmentações, com o objetivo de tornar a combinação entre os modelos ainda mais eficiente e sensível às características específicas de cada imagem.

Agradecimentos

Este trabalho foi financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Brasil – Código de Financiamento 001, pela Fundação de Amparo à Pesquisa do Maranhão (FAPEMA), pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pela Empresa Brasileira de Serviços Hospitalares (Ebserh) – Brasil (Proc. 409593/2021-4). Ainda, reconhecemos o uso de LLM para verificação ortográfica, correção gramatical e assistência na tradução de termos específicos.

Referências

- Abraham, N. and Khan, N. M. (2019). A novel focal tvsky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 683–687. IEEE.
- 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.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.
- Diniz, J. O., Dias Jr, D. A., da Cruz, L. B., Marques, R. C., Gomes Jr, D. L., Cortês, O. A., de Carvalho Filho, A. O., and Quintanilha, D. B. (2024). Efficientensemble: Diagnóstico de câncer de mama em imagens de ultrassom utilizando processamento de imagens e ensemble de efficientnets. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 202–213. SBC.

- Gómez-Flores, W., Gregorio-Calas, M. J., and Coelho de Albuquerque Pereira, W. (2024). Bus-bra: a breast ultrasound dataset for assessing computer-aided diagnosis systems. *Medical Physics*, 51(4):3110–3123.
- Gonçalves, J. V., de Souza, D. V., dos Santos, C. I., do Nascimento, C. E., da Cruz, L. B., Junior, D. A., and Diniz, J. O. (2024). D. iagnóstica: Ferramenta cadx para diagnóstico de doenças pulmonares em imagens radiológicas. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 214–225. SBC.
- Gonzalez, R. and Woods, R. (2000). *Processamento de imagens digitais*. Edgard Blucher.
- Guo, Y., Qiang, Y., Chen, Q., Li, Q., and Sun, J. (2025). Msra-net: A multi-scale and region-aware network for breast cancer ultrasound image segmentation. *Digital Signal Processing*, page 105534.
- Kim, J., Harper, A., McCormack, V., Sung, H., Houssami, N., Morgan, E., Mutebi, M., Garvey, G., Soerjomataram, I., and Fidler-Benaoudia, M. M. (2025). Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nature Medicine*, pages 1–9.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.
- Qiu, W., Hamburg, E., Zhou, Y., and Salehani, Y. E. (2025). Lightweight u-net for breast ultrasound image segmentation.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Sood, R., Rositch, A. F., Shakoob, D., Ambinder, E., Pool, K.-L., Pollack, E., Mollura, D. J., Mullen, L. A., and Harvey, S. C. (2019). Ultrasound for breast cancer detection globally: a systematic review and meta-analysis. *Journal of global oncology*.
- Teles, F. R., Ribeiro, N. P., da Cruz, L. B., Júnior, G. B., de Paiva, A. C., Diniz, J. O., and Cortes, O. A. (2025). Atlas-guided u-net++ with efficientnetb5 for automatic pancreas segmentation in abdominal ct scans. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 104–115. SBC.
- Wei, W., Wu, J., and Shao, G. (2025). Research on breast tumor segmentation based on the mamba architecture. *Frontiers in Oncology*, 15:1672274.
- Zhang, C., Wang, L., Wei, G., Kong, Z., and Qiu, M. (2024). A dual-branch and dual attention transformer and cnn hybrid network for ultrasound image segmentation. *Frontiers in Physiology*, 15:1432987.
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *International workshop on deep learning in medical image analysis*, pages 3–11. Springer.