

UNIVERSIDADE FEDERAL DO PARANÁ

LUIS GUSTAVO DE MATOS DOS SANTOS

APLICANDO INTELIGÊNCIA ARTIFICIAL EM IMAGENS DE RESSONÂNCIA MAGNÉTICA:
UM ESTUDO DE SEGMENTAÇÃO DE IMAGENS PARA O DIAGNÓSTICO DE CÂNCER DE
PRÓSTATA COM O MODELO U-NET

CURITIBA, PR
2024

LUIS GUSTAVO DE MATOS DOS SANTOS

APLICANDO INTELIGÊNCIA ARTIFICIAL EM IMAGENS DE RESSONÂNCIA MAGNÉTICA:
UM ESTUDO DE SEGMENTAÇÃO DE IMAGENS PARA O DIAGNÓSTICO DE CÂNCER DE
PRÓSTATA COM O MODELO U-NET

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Especialista em Inteligência Artificial Aplicada no Programa de Pós-Graduação em Inteligência Artificial Aplicada, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná.

Orientador: Jaime Wojciechowski.

CURITIBA, PR
2024



MINISTÉRIO DA EDUCAÇÃO
SETOR DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
CURSO DE PÓS-GRADUAÇÃO INTELIGÊNCIA ARTIFICIAL
APLICADA - 40001016348E1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação INTELIGÊNCIA ARTIFICIAL APLICADA da Universidade Federal do Paraná foram convocados para realizar a arguição da Monografia de Especialização de **LUIS GUSTAVO DE MATOS DOS SANTOS** intitulada: **APLICANDO INTELIGÊNCIA ARTIFICIAL EM IMAGENS DE RESSONÂNCIA MAGNÉTICA: UM ESTUDO DE SEGMENTAÇÃO DE IMAGENS PARA O DIAGNÓSTICO DE CÂNCER DE PRÓSTATA COM O MODELO U-NET**, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua aprovação no rito de defesa.

A outorga do título de especialista está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 10 de Junho de 2024.

JAIME WOJCIECHOWSKI

Presidente da Banca Examinadora

RAZER ANTHOM NIZER ROJAS MONTAÑO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Aplicando Inteligência Artificial em Imagens de Ressonância Magnética: Um Estudo de Segmentação de Imagens para o Diagnóstico de Câncer de Próstata com o Modelo U-Net

Luis Gustavo de Matos dos Santos
Setor de Educação Profissional e Tecnológica
Universidade Federal do Paraná
Curitiba, Brasil
gustavomatos@ufpr.br

Jaime Wojciechowski
Setor de Educação Profissional e Tecnológica
Universidade Federal do Paraná
Curitiba, Brasil
jaimewo@ufpr.br

Resumo—A ressonância magnética é uma ferramenta essencial no diagnóstico do câncer de próstata, fornecendo imagens detalhadas da glândula prostática e auxiliando na detecção de lesões. A segmentação de imagens desempenha um papel crucial na medicina, permitindo a identificação precisa de regiões suspeitas. Este estudo técnico investiga a aplicação de uma arquitetura *U-Net* aprimorada com unidades residuais para a segmentação automática de lesões prostáticas em imagens de ressonância magnética. Utilizando dados do desafio *PI-CAI*, o modelo *U-Net* foi treinado em um conjunto de dados abrangente, combinando anotações de especialistas humanos e anotações geradas por algoritmos de IA, e avaliado para diferentes *batch sizes* (8, 16 e 32). Os resultados mostram que o *batch size* de 16 proporcionou o melhor desempenho, com uma AUROC média de 0,704, AP média de 0,330 e ranking médio de 0,517, destacando o equilíbrio entre estabilidade de treinamento e capacidade de generalização. No entanto, a variabilidade observada nas métricas de desempenho (AUROC $\sigma = 0,111$, AP $\sigma = 0,137$, ranking $\sigma = 0,109$) ressalta os desafios inerentes à segmentação de imagens médicas, como a heterogeneidade dos dados e possíveis subjetividades nas anotações.

Palavras-chave—Câncer de Próstata, Segmentação de Imagem, Imagem de Ressonância Magnética, U-Net, Inteligência Artificial

Resumo—Magnetic resonance imaging (MRI) is an essential tool in diagnosing prostate cancer, providing detailed images of the prostate gland and aiding in lesion detection. Image segmentation plays a crucial role in medicine, allowing precise identification of suspicious regions. This technical study investigates the application of an improved *U-Net* architecture with residual units for automatic segmentation of prostate lesions in MRI images. Using data from the *PI-CAI* challenge, the *U-Net* model was trained on a comprehensive dataset, combining annotations from human experts and annotations generated by AI algorithms, and evaluated for different batch sizes (8, 16, and 32). The results show that a batch size of 16 provided the best performance, with an average AUROC of 0.704, average AP of 0.330, and average ranking of 0.517, highlighting the balance between training stability and generalization capability. However, the variability observed in performance metrics (AUROC $\sigma = 0.111$, AP $\sigma = 0.137$, ranking $\sigma = 0.109$) underscores the inherent challenges of medical image segmentation, such as data heterogeneity and possible subjectivity in annotations.

Index Terms—Prostate Cancer, Image Segmentation, Magnetic

Resonance Image, U-Net, Artificial Intelligence

I. DESENVOLVIMENTO

De acordo com Bray et al. [1], o câncer de próstata é uma preocupação significativa de saúde global entre os homens, sendo a principal causa de morte em 52 países. Em nível mundial, esse tipo de câncer é o segundo mais comum entre homens (apenas atrás do câncer de pulmão) e o quinto em termos de mortalidade. No Brasil, Instituto Nacional de Câncer [2] estima que, para os anos entre 2023 e 2025, o câncer de próstata seja o segundo mais frequente na população geral e o mais comum na população masculina.

O diagnóstico precoce do câncer de próstata é crucial para melhorar a evolução dos pacientes e reduzir as taxas de mortalidade. Nesse tipo de tumor, são utilizados diversos métodos de rastreamento como o antígeno prostático específico (PSA) desempenham um papel significativo na detecção precoce do câncer de próstata [3] [4]. O exame de toque retal também é uma importante ferramenta no diagnóstico precoce do câncer de próstata. Além disso, são realizados exames de imagem, como a ultrassonografia transretal e a **ressonância magnética (RM)** multiparamétrica da próstata.

A RM multiparamétrica transformou o diagnóstico do câncer de próstata ao proporcionar imagens detalhadas que são essenciais para a detecção, localização e estadiamento da doença [5] [6]. Estudos demonstram que esta modalidade de imagem possui uma precisão superior na identificação de tumores clinicamente significativos, possibilitando biópsias guiadas mais precisas e aprimorando a eficácia do diagnóstico [7].

A RM é uma técnica de imagem médica poderosa que utiliza campos magnéticos e ondas de rádio para criar imagens detalhadas dos órgãos e tecidos internos do corpo humano. Diferentemente dos raios X e da tomografia computadorizada, a RM não utiliza radiação ionizante, tornando-a um método seguro para visualizar estruturas anatômicas complexas [8].

Observe o processo de obtenção das imagens descrito a seguir e ilustrado na Figura 1:

- **Natureza da molécula de água:** Cada átomo de hidrogênio, presente aos pares em moléculas de água, possui um próton que gira em torno de seu eixo (*Spin*). Vemos na Figura 1 em (A).
- **Dispersão atômica do corpo:** O corpo humano é composto principalmente de água, possui uma distribuição aleatória nos Spins de cada hidrogênio. Vemos na Figura 1 em (B).
- **Alinhamento dos Prótons:** Quando o paciente é colocado dentro do aparelho de RM, um forte campo magnético faz com que esses prótons se alinhem na mesma direção. Vemos na Figura 1 em (C).
- **Excitação dos Prótons:** Ondas de rádio são então emitidas pelo aparelho, fornecendo energia aos prótons e fazendo com que eles se inclinem temporariamente de sua posição original. Vemos na Figura 1 em (D).
- **Relaxamento e Emissão de Sinais:** Quando as ondas de rádio são desligadas, os prótons retornam à sua posição original, liberando a energia absorvida na forma de sinais de rádio. Esses sinais são captados por antenas dentro do aparelho de RM. Vemos na Figura 1 em (F).

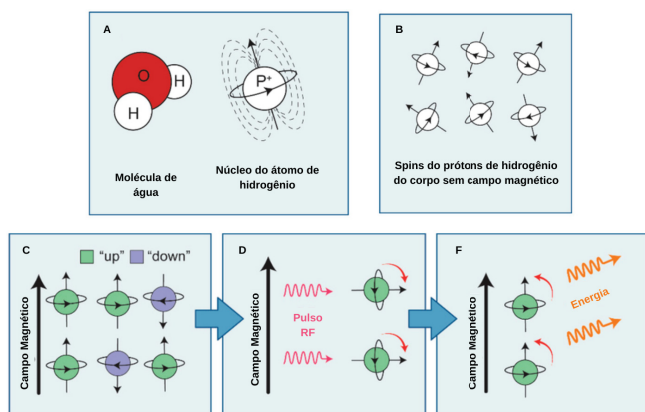


Figura 1. Mecanismo de ressonância magnética.
Fonte: Modificado de [9]

Os sinais capturados pelos sensores resulta na excitação e detecção de sinais de moleculares presentes na região de interesse [10]. Diferentes técnicas de RM, como a relaxometria¹, imagens ponderadas por difusão e de susceptibilidade magnética, permitem avaliar propriedades específicas do organismo [10].

Assim, é possível obter imagens de diversos parâmetros relevantes, destacando imagens com *T2-weighted* (*T2W*), *Apparent Diffusion Coefficient* (*ADC*), *High B-Value* (*HBV*) sendo fundamentais nesse processo. Com o parâmetro *T2W* obtêm-se sequências de RM que produzem imagens de alto contraste baseadas nas diferenças nos tempos de relaxamento *T2* dos tecidos da glândula prostática, permitindo a detecção de

anormalidades como tumores ou inflamação [11]. A imagem *ADC*, derivada das imagens de difusão ponderada (*DWI*), fornece informações sobre a mobilidade das moléculas de água nos tecidos e é útil na detecção e estadiamento do câncer de próstata [12].

No contexto do câncer prostático, as imagens geradas em altas intensidades de campo, como em ressonâncias magnéticas de 3,0 T, oferecem alta resolução para a detecção precoce de pequenos focos de câncer [13]. O valor "b" na imagem por difusão representa a força e a duração dos gradientes de difusão aplicados durante a ressonância magnética, sendo um parâmetro crítico na supressão da sinalização do tecido normal e no destaque de áreas com difusão restrita, como tumores [14]. Ao usar valores elevados de "b" na técnica de *HBV*, a sinalização do tecido normal é suprimida, enquanto as áreas com difusão restrita, como tumores, são destacadas [15].

Como exemplo, observe na Figura 2, a ressonância magnética fornece informações detalhadas sobre a estrutura do tecido prostático. Em (a) é apresentada a imagem ponderada em *T2*, que mostra um contraste claro entre os diferentes tipos de tecidos. Em (b), exibe-se a imagem de *ADC*, útil para avaliar a restrição de difusão nas regiões suspeitas. Por fim, (c) destaca a imagem com valores elevados de "b", que realça ainda mais a área de potencialidade de câncer prostático. A região circutada em todas as imagens indica a área de interesse para o diagnóstico, demonstrando a importância de múltiplas técnicas de imagem para uma avaliação abrangente.

Estudos demonstraram que a adição de *DWI* à *T2W* melhora significativamente a precisão na medição do volume tumoral da próstata, sendo sugerido que a combinação de *DWI* com outras componentes da RM multiparamétrica é a mais precisa na predição do volume tumoral da próstata [17]. Portanto, essas combinações desempenham um papel fundamental na avaliação pré-biopsia, orientação de biópsias e no planejamento de tratamentos após a recorrência do câncer de próstata [18] [19].

Estudos demonstraram que a adição de *DWI* à *T2W* melhora significativamente a precisão na medição do volume tumoral da próstata, sendo sugerido que a combinação de *DWI* com outras componentes da RM multiparamétrica é a mais precisa na predição do volume tumoral da próstata [17]. Portanto, essas combinações desempenham um papel fundamental na avaliação pré-biopsia, orientação de biópsias e no planejamento de tratamentos após a recorrência do câncer de próstata [18] [19].

A interpretação de imagens de RM da próstata é crucial para o diagnóstico preciso do câncer. Diretrizes como o Sistema de Relato e Dados de Imagem da Próstata (*PI-RADS*) orientam essa interpretação, enfatizando a importância das imagens ponderadas em *T2* (*T2W*) e de difusão (*DWI*) na detecção do câncer, enquanto as imagens de Realce Dinâmico por Contraste (*DCE*) desempenham um papel secundário [18]. O *PI-RADS* fornece critérios objetivos para classificar lesões em uma escala de 1 a 5, estratificando o risco de câncer clinicamente significativo com base em cada sequência de imagem, representando um avanço na padronização e

¹Investigação das propriedades de relaxamento em Imagens de Ressonância Magnética.

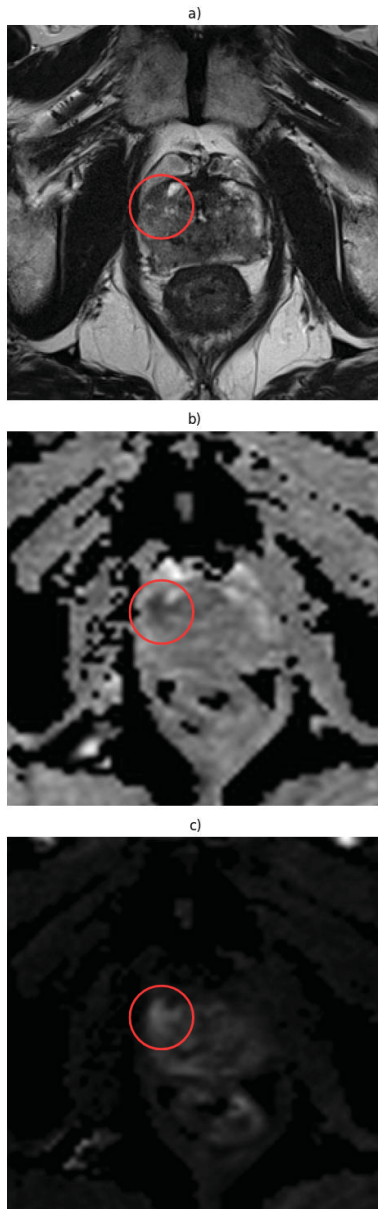


Figura 2. Imagens de ressonância magnética de uma seção: (a) T2W, (b) ADC, e (c) HBV. A região circulada indica a potencialidade de câncer prostático.

Fonte: Modificado de um estudo aleatório da base de dados do Desafio PI-CAI [16].

otimização do uso da RM multiparamétrica na detecção precoce do câncer de próstata [20] [21].

A segmentação de imagens, por sua vez, tem se consolidado como uma ferramenta valiosa no auxílio ao diagnóstico e tratamento do câncer de próstata. Métodos computacionais têm sido empregados na identificação de unidades glandulares prostáticas em imagens histopatológicas, extraindo características quantitativas que diferenciam o tecido canceroso do benigno [22]. Além disso, técnicas de segmentação têm sido amplamente aplicadas em imagens de RM multiparamétrica para delinear

as regiões de interesse dentro da próstata, permitindo uma avaliação mais precisa da extensão e localização do tumor. Estudos como o de Comelli et al. demonstram a eficácia de redes neurais profundas, como *U-Net*, *Efficient Neural Network* e *Efficient Residual Factorized Network*, na segmentação automatizada em tempo real da glândula prostática em imagens ponderadas em T2 [23].

A automatização da segmentação por meio de ferramentas computacionais é essencial para lidar com a variabilidade inerente às imagens da próstata, como diferenças em forma, tamanho e intensidade ao redor da glândula. Redes neurais profundas têm sido exploradas para a segmentação semântica da próstata em imagens de RM, evidenciando a importância de métodos robustos de segmentação para o diagnóstico e tratamento eficazes do câncer [24]. Algoritmos como o agrupamento *K-Means*, redes grafo-convolucionais e métodos baseados em *supervoxel* também têm apresentado resultados promissores na segmentação automática de imagens da próstata, contribuindo significativamente para o diagnóstico preciso [25] [26] [27].

Dentre os parâmetros de treinamento, a escolha do *batch size* é um fator crítico no treinamento de redes neurais, especialmente no domínio da segmentação de imagens médicas, onde os conjuntos de dados geralmente são complexos e desequilibrados. O tamanho do *batch*, que representa o número de amostras de treinamento usadas em uma iteração, influencia significativamente a eficiência e a eficácia do treinamento. *Batch sizes* maiores podem acelerar o treinamento devido ao paralelismo aprimorado e à eficiência computacional [28]. No entanto, *batch sizes* muito grandes podem resultar em desempenho de generalização inferior, enquanto *batch sizes* menores geralmente oferecem melhores capacidades de generalização [?].

Além disso, o impacto do *batch size* no treinamento pode ser influenciado pela arquitetura específica da rede neural em uso. Por exemplo, em arquiteturas como a *U-Net*, comumente empregadas na segmentação de imagens biomédicas, o sucesso das redes convolucionais foi historicamente limitado por fatores como o tamanho do conjunto de treinamento e a arquitetura da rede [?]. Portanto, ao treinar redes neurais como a *U-Net*, o *batch size* deve ser selecionado meticulosamente para equilibrar a eficiência do treinamento e a generalização do modelo.

Neste estudo técnico, será investigada a segmentação do câncer de próstata em imagens de RM disponibilizadas publicamente no desafio PI-CAI. Será empregada a arquitetura *U-Net*, implementada com o *framework Monai*, especializado em imagens médicas, com o objetivo de avaliar seu desempenho de acordo com os parâmetros estabelecidos pelos autores do desafio. Serão analisados a metodologia utilizada, os resultados obtidos e sua relevância para a prática clínica e para futuras pesquisas na área de diagnóstico por imagem do câncer de próstata. Além disso, diferentes parâmetros de *batch size* serão avaliados para determinar o impacto no desempenho do modelo.

A. Descrição dos Dados

O conjunto de dados utilizado neste estudo de segmentação de imagens 3D de RM de pacientes com suspeita de câncer de próstata é proveniente do desafio *PI-CAI* (Inteligência Artificial & Radiologistas na Detecção do Câncer de Próstata em RM). Este abrangente repositório público contém mais de 10.000 exames de ressonância magnética da próstata anonimizados, oriundos de 1.500 casos de estudos [29] [30], todos disponibilizados sob a licença *CC BY-NC 4.0*. O desafio tem como objetivo aprimorar a detecção, localização e estratificação de risco do câncer de próstata por meio da inteligência artificial e da análise de imagens médicas.

Os exames foram adquiridos em quatro centros de diagnóstico distintos:

- Radboud University Medical Center (RUMC)² - Holanda
- Ziekenhuis Groep Twente (ZGT)³ - Holanda
- Prostaat Centrum Noord-Nederland (PCNN)⁴ - Holanda
- St. Olav's Hospital, Trondheim University Hospital (STOH)⁵ - Noruega

De acordo com [29], as imagens foram adquiridas utilizando scanners da *Siemens Healthineers* ou *Philips Medical Systems*, ambos equipados com bobinas de superfície. O conjunto de dados inclui sequências ponderadas em *T2* nos planos axial, sagital e coronal, além de imagens axiais ponderadas em difusão de alto valor *b* (*HBV*) ($\geq 1000 \text{ s/mm}^2$) e seus respectivos mapas de coeficiente de difusão aparente (*ADC*).

Essa diversidade de centros e equipamentos, embora enriquecedora para a representatividade do conjunto de dados, também introduz uma heterogeneidade considerável em termos de qualidade de imagem, artefatos e protocolos de aquisição. Essa variabilidade intrínseca torna o desafio da segmentação ainda mais complexo, exigindo o desenvolvimento de algoritmos robustos e capazes de generalizar para diferentes condições de aquisição.

1) *Anotações*: A segmentação de imagens vai além da simples localização de objetos em uma cena. Ela requer um detalhamento preciso dos limites e da forma do objeto de interesse, tornando as anotações ou rótulos um componente crucial para o treinamento eficaz de modelos de aprendizado profundo. Essas anotações fornecem informações essenciais sobre a forma e os contornos dos objetos a serem segmentados, permitindo que o modelo aprenda a distinguir entre o objeto de interesse e o fundo com alta precisão. A qualidade das anotações afeta diretamente a performance do modelo, pois elas fornecem os exemplos necessários para que o algoritmo aprenda a reconhecer padrões e características específicas, resultando em uma segmentação mais precisa e robusta.

No contexto deste estudo, a heterogeneidade do conjunto de dados ressalta a importância de anotações de alta qualidade para garantir que o modelo de segmentação seja capaz de lidar com a variabilidade presente nas imagens. Além disso, como

apontado por Zhang et al. [31], o enriquecimento das anotações com informações adicionais, como características texturais ou funcionais, pode aprimorar ainda mais o desempenho dos modelos de segmentação.

No caso do desafio *PI-CAI*, as anotações, divulgadas e mantidas no repositório *picai_labels*⁶ no GitHub, consistem em máscaras binárias que delimitam as regiões da próstata em cada fatia das imagens de *RM*. Essas máscaras foram geradas por especialistas em radiologia ou por abordagens automáticas de segmentação por IA [32], e servem como referência para o treinamento e avaliação dos algoritmos de segmentação. A qualidade e a precisão dessas anotações são fundamentais para garantir que o modelo aprenda a segmentar a próstata de forma consistente e confiável, mesmo diante da variabilidade inerente ao conjunto de dados.

a) Anotações Derivadas de Especialistas Humanos:

Dez pesquisadores treinados ou um residente de radiologia dos institutos mencionados anteriormente foram responsáveis pela rotulagem das lesões de câncer de próstata clinicamente significativo (*csPCa*), sob a supervisão de três radiologistas especialistas [29]. Cada anotação foi realizada com meticulosidade, utilizando todos os exames de RM disponíveis do paciente, laudos radiológicos e patológicos, e, quando aplicável, espécimes de prostatectomia total. Os contornos das lesões foram delineados utilizando o *ITK-SNAP* v3.8.0, um software de código aberto desenvolvido pela Universidade da Carolina do Norte, amplamente utilizado para segmentação de imagens biomédicas [33].

No conjunto de dados de treinamento público, 1.075 casos foram identificados como tecido benigno ou câncer de próstata indolente, enquanto os 425 restantes foram diagnosticados com lesões *csPCa*. Dentre esses casos positivos, apenas 220 foram anotados por especialistas, resultando em 205 casos sem anotação. Consequentemente, apenas 17% das anotações derivadas por especialistas humanos apresentam lesões *csPCa* [29].

As anotações de lesões *csPCa* realizadas por especialistas no conjunto de dados de treinamento e desenvolvimento público estão disponíveis em dois formatos:

- **Anotações Originais**: Neste formato, todas as sequências axiais de *RM* (*T2W*, *DWI*, *ADC*) são utilizadas para identificar e delimitar as lesões de *csPCa*. As anotações são cuidadosamente mapeadas para corresponder às áreas nas imagens *DWI/ADC*, garantindo a precisão da localização das lesões.
- **Anotações Reamostradas**: Neste formato, as anotações originais são convertidas e fornecidas nas mesmas dimensões e resolução espacial das imagens *T2W* correspondentes. Essa padronização facilita a utilização das anotações por modelos de inteligência artificial, simplificando o processo de treinamento e avaliação.

A disponibilidade de anotações em dois formatos distintos oferece flexibilidade para a escolha do formato mais adequado

²<https://www.diagnijmegen.nl/>

³<https://www.zgt.nl/>

⁴<https://prostaacentrumnoordnederland.nl/>

⁵<https://www.stolav.no/en>

⁶https://github.com/DIAGNijmegen/picai_labels

para cada aplicação, seja para fins de pesquisa ou desenvolvimento de ferramentas de diagnóstico assistido por computador.

b) *Anotações Derivadas de IA*: No *RUMC*, a equipe implementou uma abordagem inovadora de aprendizagem semi-supervisionada para gerar anotações para todos os 1.500 casos de treinamento [32]. Esta metodologia resultou em um conjunto de anotações abrangente e de alta qualidade, permitindo um treinamento e avaliação mais eficazes dos modelos de segmentação.

B. Métodos

Diversos estudos têm investigado a aplicação de modelos de aprendizado profundo, especialmente a arquitetura *U-Net* e suas variações, na segmentação de órgãos em diferentes modalidades de imagem, como ressonância magnética cardíaca e ecocardiografia. Wu et al. [34] empregaram a *U-Net* para a segmentação automática do ventrículo esquerdo em ressonância magnética cardíaca, demonstrando a eficácia do aprendizado profundo na delimitação precisa desse órgão, essencial para a análise de imagens cardíacas.

No contexto deste trabalho, Dai et al. [35] utilizaram modelos *U-Net* 2D e 3D para segmentar a glândula prostática e lesões prostáticas em imagens de ressonância magnética multiparamétricas, evidenciando a versatilidade e o desempenho das arquiteturas *U-Net* em diferentes domínios de aplicação. A escolha dessa arquitetura se justifica por sua capacidade de lidar com dados médicos, especialmente em tarefas de segmentação.

1) *Arquitetura*: A *U-Net* é uma arquitetura de rede neural convolucional que se destaca por seu formato em "U" e suas conexões de salto (*skip connections*) [36]. Essa estrutura é composta por um caminho de contração (*encoder*) e um caminho de expansão (*decoder*), simétricos e interligados por camadas de convolução, permitindo a captura tanto do contexto global quanto dos detalhes locais das imagens.

O *encoder*, responsável por extrair características da imagem em diferentes níveis de abstração, e o *decoder*, encarregado da segmentação da imagem [37], trabalham em conjunto para atingir o objetivo da rede. As conexões de salto desempenham um papel crucial ao transmitir informações de baixo nível diretamente para camadas mais profundas, preservando detalhes importantes durante o processo de decodificação [36].

Antes de aprofundar nas particularidades da arquitetura *U-Net*, é importante revisar os conceitos fundamentais de redes neurais. As operações de convolução, por exemplo, são o cerne da extração de características em redes neurais convolucionais [38]. Elas consistem em deslizar um filtro (*kernel*⁷) sobre a imagem de entrada, realizando multiplicações elemento a elemento e somando os resultados. Cada filtro é especializado em detectar um tipo específico de característica, como bordas, cantos ou texturas, gerando um mapa de características que realça a presença e a intensidade dessas características.

O *pooling*, também chamado de *downsampling* ou *subsampling*, é uma operação que reduz a resolução espacial dos

mapas de características [39]. Aplicado após as camadas de convolução, ele diminui a quantidade de informações a serem processadas, tornando a rede mais eficiente e menos suscetível ao *overfitting*⁸. Dentre os tipos de *pooling*, destacam-se o *max pooling*, que seleciona o valor máximo em uma região do mapa de características, e o *average pooling*, que calcula a média dos valores.

A Figura 3 ilustra a operação de *max pooling* em um exemplo tridimensional. Cada número representa uma característica extraída da imagem original. Ao aplicar o *max pooling* com uma janela 2x2x2, o volume resultante é reduzido, preservando apenas a característica de maior valor em cada janela. Esse processo não apenas diminui a dimensionalidade dos dados, mas também realça as características mais relevantes, o que é fundamental para o aprendizado da rede neural.

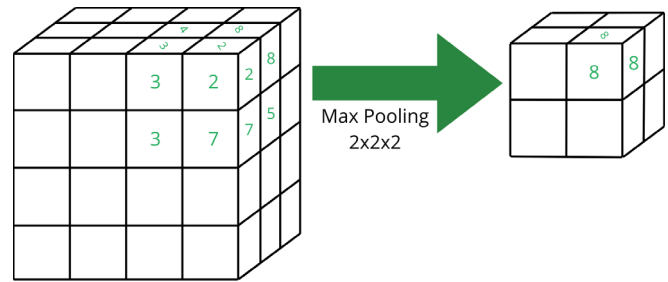


Figura 3. Operação de *Max Pooling* em uma imagem 3D
Fonte: O autor (2024)

Após a aplicação do *max pooling* e de outras operações, como as convoluções, é crucial introduzir não linearidade na rede para permitir o aprendizado de relações complexas entre as características. Uma forma de alcançar essa não linearidade é através da função de ativação *ReLU* (Unidade Linear Retificada), amplamente utilizada em redes neurais, incluindo a *U-Net* [40]. Ela é definida como:

$$f(x) = \max(0, x) \quad (1)$$

Essa função retorna o valor de entrada se for positivo e zero caso contrário. A Figura 4 apresenta o gráfico gerado por essa operação matemática, ilustrando seu comportamento.

Tanto o *max pooling* quanto a função de ativação *ReLU* desempenham papéis essenciais no *encoder* da *U-Net*, o qual é responsável por extrair e processar as características da imagem de entrada. A combinação dessas operações permite a redução da dimensionalidade dos dados, o realce das características mais relevantes e a introdução de não linearidade, contribuindo para a capacidade da rede de aprender e generalizar a partir dos dados.

a) *Encoder*: O *encoder*, ou caminho de contração, é a primeira etapa da *U-Net* e tem a função primordial de extrair as características contextuais da imagem. Para isso, ele emprega uma série de blocos compostos por duas camadas de convolução

⁷Pequena matriz de números que desliza sobre a imagem, aplicando uma operação matemática para extrair características.

⁸Ocorre quando um modelo aprende os detalhes específicos dos dados de treinamento em vez dos padrões gerais, resultando em um desempenho ruim em novos dados.

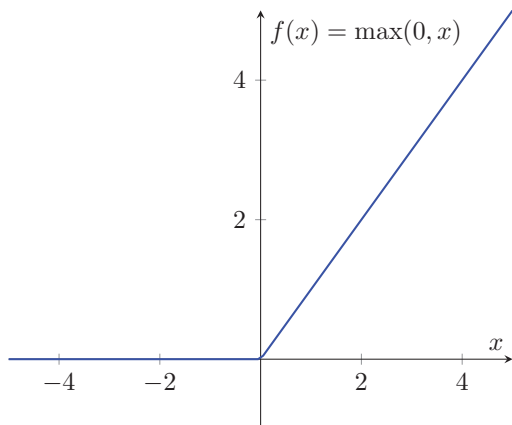


Figura 4. Gráfico da Função de Ativação *ReLU*
Fonte: O autor (2024)

de 3x3, um tamanho que se destaca por equilibrar a captura de detalhes locais com a eficiência computacional [41]. Em seguida, uma camada de *max pooling* de 2x2 com *stride* 2 reduz a resolução espacial pela metade, concentrando as informações mais relevantes [36]. A função de ativação *ReLU*, aplicada após cada convolução, introduz a não linearidade essencial para modelar relações complexas nos dados [36]. A cada bloco, a imagem é subamostrada e sua profundidade (número de canais) é duplicada, permitindo a extração progressiva de características mais abstratas e de alto nível.

b) *Decoder*: O *decoder*, ou caminho de expansão, é a segunda etapa da *U-Net* e atua na reconstrução da resolução espacial da imagem, buscando recuperar a informação de localização perdida durante a contração. Cada bloco do *decoder* inicia com uma camada de *upsampling*, que dobra a resolução espacial, seguida por uma convolução 2x2 para ajustar o número de canais. A inovação da *U-Net* reside nas conexões de salto (*skip connections*), que concatenam as características de alta resolução do *encoder* com a saída da camada de *upsampling*. Essa fusão de informações é então processada por duas camadas de convolução 3x3 e uma função de ativação *ReLU* [36]. O processo se repete até que a resolução original da imagem seja restaurada, culminando na segmentação final.

c) *Skip Connections*: As conexões de salto (*skip connections*) são um pilar fundamental da arquitetura *U-Net*, estabelecendo um elo crucial entre o *encoder* e o *decoder*. Elas permitem que a rede recupere informações de detalhes finos que poderiam ser perdidas durante as operações de *pooling* no *encoder*. Ao concatenar essas características de alta resolução com as camadas de *upsampling* no *decoder*, a *U-Net* combina eficientemente informações de contexto global com detalhes locais [36]. Essa estratégia se mostra particularmente eficaz em tarefas de segmentação, nas quais a localização precisa de objetos é fundamental, resultando em segmentações mais precisas e refinadas.

d) *Camada de Saída*: A camada final da *U-Net*, responsável por gerar a segmentação propriamente dita, consiste em uma convolução 1x1. Essa convolução reduz a profundidade dos

mapas de características para o número de classes desejadas, produzindo um mapa de probabilidade para cada classe em cada pixel da imagem. A escolha da função de ativação nesta camada é crucial e depende da natureza da tarefa de segmentação. Para problemas de segmentação multiclasse, a função *softmax* é frequentemente utilizada, pois normaliza as probabilidades de cada classe, garantindo que sua soma seja igual a um [36]. Já em problemas de segmentação binária, a função *sigmoid* é mais adequada, convertendo as saídas em probabilidades entre 0 e 1, representando a probabilidade de um pixel pertencer à classe de interesse [36].

A Figura 5 ilustra a arquitetura completa da *U-Net*, destacando o fluxo das informações através do *encoder*, do *decoder* e das conexões de salto. Cada caixa azul representa um mapa de características multicanal, com o número de canais indicado acima e o tamanho abaixo. As caixas brancas representam as cópias dos mapas de características do *encoder*, que são concatenadas no *decoder* através das conexões de salto. As setas indicam as diferentes operações realizadas em cada etapa, desde as convoluções e *pooling* no *encoder* até as convoluções e *upsampling* no *decoder*. Os retângulos tracejados delimitam o *encoder* e o *decoder*, evidenciando a simetria característica da arquitetura *U-Net*.

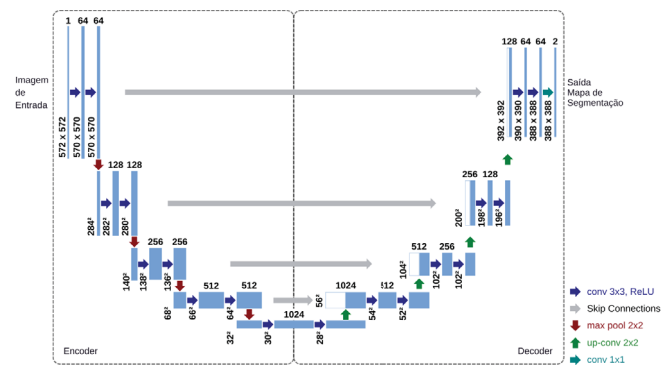


Figura 5. Arquitetura do Modelo de Rede Neural *U-Net*
Fonte: Adaptada de [36]

Em tarefas de segmentação volumétrica (3D), a *U-Net* pode ser adaptada utilizando sua estrutura central e substituindo operações em 2D por suas contrapartes volumétricas [42].

A arquitetura empregada neste trabalho consiste em uma versão otimizada da *U-Net* convencional. Essa implementação específica incorpora unidades residuais, introduzidas por He et al. [43], como solução para o problema do gradiente evanescente observado durante o treinamento de *U-Nets* profundas. Esse problema ocorre quando os gradientes de erro se tornam muito pequenos durante a retropropagação, dificultando o ajuste dos pesos das camadas iniciais da rede.

As unidades residuais são blocos de construção fundamentais em redes neurais profundas, permitindo o fluxo direto de informações através da rede e, consequentemente, mitigando o problema do gradiente evanescente. Em essência, uma unidade residual aprende a diferença entre sua entrada e a saída desejada,

em vez de aprender a mapear a entrada diretamente para a saída.

Em termos matemáticos, a saída de uma unidade residual $F(x_l)$ em uma determinada camada l de uma rede neural é definida por:

$$F(x_l) = H(x_l) + x_l \quad (2)$$

Onde:

- x_l representa a entrada da camada l .
- $H(x_l)$ representa a função de transformação aprendida pela rede (por exemplo, uma sequência de operações de convolução e funções de ativação).
- $F(x_l)$ representa a saída da unidade residual.

A função de ativação $H(x_l)$ geralmente é uma função não linear, como a *ReLU*. A presença da conexão de identidade x_l possibilita o fluxo direto do gradiente para as camadas anteriores durante o processo de treinamento, facilitando a otimização da rede. Essa característica é fundamental para mitigar o problema do gradiente evanescente, permitindo o treinamento de redes neurais mais profundas e complexas.

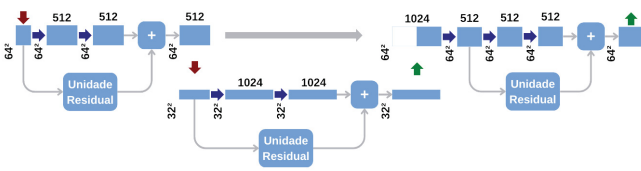


Figura 6. Ilustração do recorte da arquitetura *U-Net* aprimorada com blocos residuais
Fonte: O autor (2024)

A inclusão de blocos residuais na arquitetura *U-Net*, como ilustrado na Figura 6, tem se mostrado eficaz na mitigação do problema do gradiente evanescente e no aprimoramento do desempenho da rede em tarefas de segmentação. No entanto, a aplicação da *U-Net* em imagens de RM apresenta desafios adicionais devido à variabilidade na resolução espacial dos dados.

2) *Preparação dos Dados*: A variabilidade na resolução espacial das imagens de RM apresenta um desafio considerável para o treinamento da *U-Net*, uma vez que pode gerar discrepâncias entre as resoluções das imagens de treinamento e teste, impactando negativamente o desempenho da rede [44]. Essa variabilidade decorre de diversos fatores, como parâmetros de aquisição, equipamentos e protocolos clínicos distintos.

Para enfrentar esse desafio, diversas técnicas de re-amostragem podem ser empregadas. A interpolação, como a interpolação bilinear, bi-cúbica e Lanczos, é uma técnica comum para aumentar a resolução de imagens, mas pode introduzir suavização e perda de detalhes [45]. Técnicas de super-resolução, como redes neurais de super-resolução (*SRCNNs*) e redes adversárias generativas (*GANs*), têm sido utilizadas para aprimorar a resolução de imagens de RM, preservando detalhes e reduzindo artefatos [44]. Em muitos casos, a combinação de diferentes técnicas de re-amostragem

pode ser benéfica, maximizando a qualidade das imagens e minimizando os efeitos indesejados [44].

A re-amostragem pode ser isotrópica, tornando a resolução uniforme em todas as direções, ou anisotrópica, ajustando a resolução em direções específicas. Essa etapa é crucial para garantir a consistência e a compatibilidade nas etapas subsequentes de processamento, ajustando a resolução ou o tamanho das imagens a uma grade ou escala comum [46]. Ao re-amostrar as imagens para uma mesma grade, as variações na resolução são mitigadas, facilitando a obtenção de resultados de segmentação mais precisos [46].

Além da re-amostragem, o recorte de dados também desempenha um papel crucial na otimização do processo de segmentação. Técnicas de recorte são frequentemente utilizadas na segmentação de imagens biomédicas para direcionar o foco para regiões específicas de interesse dentro de uma imagem. Essa abordagem permite a extração de informações relevantes, reduzindo simultaneamente a complexidade computacional e os requisitos de memória [47]. A extração de *patches* (pequenas sub-regiões da imagem original) em diferentes escalas e sua combinação com *patches* de escala única possibilitam um processamento eficiente em termos de memória, sem comprometer a precisão da segmentação [47].

O treinamento foi realizado utilizando a biblioteca *picai_baseline*, que oferece ferramentas especializadas para o tratamento de dados do *PI-CAI* e aplica nativamente o aumento de dados. Essa biblioteca foi empregada para executar as seguintes etapas no preparo de dados para treinamento semi-supervisionado:

Para o pré-processamento dos dados, foi utilizada a biblioteca *picai_baseline*, que aplica nativamente o aumento de dados. Essa biblioteca oferece ferramentas especializadas para o tratamento de dados do *PI-CAI*. A função de preparo de dados para treinamento semi-supervisionado foi empregada para executar as seguintes etapas:

- 1) **Combinação de Anotações**: Para criar um conjunto de dados de treinamento mais abrangente e robusto, anotações de especialistas humanos foram combinadas com anotações geradas por algoritmos de inteligência artificial.
- 2) **Binarização de Anotações**: As anotações, que originalmente representavam diferentes graus de suspeita de câncer, foram simplificadas para indicar apenas a presença (valor 1) ou ausência (valor 0) de lesão. Essa binarização facilita o processo de treinamento, permitindo que o modelo se concentre em distinguir entre tecido tumoral e tecido saudável.
- 3) **Re-amostragem Anisotrópica**: As imagens de RM foram re-amostradas para um espaçamento anisotrópico de 3,0 mm no eixo Z (axial) e 0,5 mm nos eixos X e Y (sagital e coronal), resultando em um tamanho de matriz de 20 x 256 x 256 *voxels*⁹.
- 4) **Conversão de Formato**: As imagens e máscaras binarizadas foram convertidas para o formato *nnUNet*, um

⁹Equivalente tridimensional de um pixel em uma imagem bidimensional.

framework amplamente utilizado para segmentação de imagens médicas.

5) **Aumento de Dados:** A biblioteca *batchgenerators*, dependência da *picai_baseline*, aplica diversas transformações e técnicas de aumento de dados, incluindo:

- a) Transformações Espaciais:
 - i) Rotação: A imagem é rotacionada em torno dos eixos X, Y e Z por ângulos aleatórios dentro de intervalos predefinidos.
 - ii) Escalonamento: Escala a imagem por um fator aleatório dentro do intervalo de 0,7 e 1,4.
- b) Transformações de Intensidade:
 - i) Normalização de Intensidade: O escore z é normalizado por instância em todas as sequências de ressonância magnética, utilizando os percentis 0,5 e 99,5 para limitar os valores, de forma independente [29].
 - ii) *GaussianNoiseTransform*: Ruído gaussiano é adicionado à imagem com uma probabilidade de 0,1 por amostra.
 - iii) *GaussianBlurTransform*: Um filtro de desfoque gaussiano é aplicado com desvio padrão aleatório entre 0,5 e 1,0, com valores diferentes para cada canal e uma probabilidade de 0,2 por amostra.
 - iv) *BrightnessMultiplicativeTransform*: Os valores dos *voxels* são multiplicados por um fator aleatório entre 0,75 e 1,25, com uma probabilidade de 0,15 por amostra.
 - v) *ContrastAugmentationTransform*: O contraste da imagem é ajustado com uma probabilidade de 0,15 por amostra.
 - vi) *SimulateLowResolutionTransform*: Uma imagem de baixa resolução é simulada, reduzindo a escala e, em seguida, aumentando-a usando interpolação. A escala de redução é escolhida aleatoriamente entre 0,5 e 1,0.
 - vii) *GammaTransform*: Uma transformação gamma invertida é aplicada com uma probabilidade de 0,1 por amostra, e uma transformação gamma padrão com uma probabilidade de 0,3.
- c) Espelhamento:
 - i) *MirrorTransform*: A imagem é espelhada em torno de um ou mais eixos X, Y e Z.
- d) Outras Transformações:
 - i) *NumpyToTensor*: As imagens e máscaras de segmentação são convertidas de *arrays* NumPy para tensores PyTorch.

6) **Validação Cruzada:** Os 1.500 casos dos dados públicos do desafio *PI-CAI* foram divididos em 5 conjuntos de validação cruzada, garantindo que os dados de treino e validação não se sobreponham [29].

Vale destacar a importância do *framework nnUNet* nesse processo. O *nnUNet* se destaca pela capacidade de configurar

automaticamente *pipelines* de segmentação para diferentes conjuntos de dados, agilizando o desenvolvimento e a experimentação. No entanto, para que o *nnUNet* funcione corretamente, a organização e a formatação dos dados de entrada devem seguir rigorosamente a estrutura de diretórios específica do *framework*.

Conforme a documentação do *nnUNet* [48], cada conjunto de dados deve ser armazenado em um diretório raiz nomeado como "TaskXXX_YYY", onde XXX é um número sequencial e YYY é o nome do conjunto de dados. Dentro desse diretório raiz, devem existir as subpastas "imagesTr", "labelsTr", "imagesTs" e "labelsTs", que armazenam, respectivamente, as imagens e as máscaras de segmentação (*labels*) dos conjuntos de treinamento e teste.

Além da estrutura de diretórios, o formato dos arquivos também é fundamental. O *nnUNet* requer que as imagens e as máscaras de segmentação estejam no formato *NIfTI* (.nii.gz). Originalmente, as imagens fornecidas pelo desafio *PI-CAI* estavam no formato *MHA* (*MetaImage Header*), necessitando de conversão. Cada imagem e sua respectiva máscara devem ter o mesmo nome, diferenciando-se apenas pela extensão do arquivo (.nii.gz para imagens e .nii.gz para máscaras). Adicionalmente, o *nnUNet* espera que as máscaras de segmentação sejam codificadas por números inteiros, onde cada valor representa uma classe diferente, justificando a etapa de binarização das anotações realizada no pré-processamento.

Com a estrutura de dados devidamente organizada e formatada de acordo com os requisitos do *nnUNet* e do desafio *PI-CAI*, estabelecem-se bases sólidas para uma avaliação rigorosa dos modelos de segmentação, garantindo a confiabilidade e a reprodutibilidade dos resultados.

3) **Parâmetros de Treinamento:** Para a aplicação da arquitetura *U-Net* aprimorada, detalhada anteriormente, utilizou-se o *framework MONAI* (*Medical Open Network for AI*). Essa estrutura de código aberto, que estende o *PyTorch*, oferece suporte especializado para dados médicos, com foco em imagens. O *MONAI* fornece arquiteturas de modelos de IA específicas para fins médicos, além de transformações e utilitários que simplificam o desenvolvimento e a implantação de modelos de IA na área médica [49] [50] [51].

O ajuste fino dos hiperparâmetros é crucial para otimizar o desempenho do modelo e garantir sua capacidade de generalização. Pesquisadores têm explorado diferentes abordagens para alcançar essa otimização. A técnica de Otimização de Busca Gravitacional, por exemplo, foi aplicada para determinar valores ótimos de hiperparâmetros para arquiteturas como *DenseNet121* [52].

a) **Otimizador Adam:** No campo da imagem médica, as técnicas de otimização do aprendizado são fundamentais para alcançar alto desempenho. Estudos demonstram a eficácia do otimizador *Adam* nesse contexto, com modelos selecionados com base em métricas de validação, como o coeficiente de similaridade de *Dice* (*DSC*) [53]. Além disso, técnicas como Campos Aleatórios Condicionais (*CRF*) e Aumento de Tempo de Teste (*TTA*) têm sido exploradas para aprimorar os resultados

de modelos como o *ResUNet++* [54], evidenciando a busca constante por otimização.

Proposto por Kingma e Ba [55], o otimizador *Adam* (*Adaptive Moment Estimation*) baseia-se em estimativas adaptativas de momentos de baixa ordem para calcular taxas de aprendizado individuais para cada parâmetro da rede. Essa abordagem, quando combinada com o ajuste adequado de hiperparâmetros, pode melhorar significativamente o desempenho de modelos de aprendizagem profunda. Um exemplo prático é a detecção de câncer de próstata em ressonância magnética multiparamétrica, onde o otimizador *Adam* foi aplicado com sucesso em modelos *CNN* leves para identificar com precisão locais de câncer [27], comprovando sua efetividade em cenários médicos complexos.

b) *Função de Perda Focal*: A função *Focal Loss*, projetada para lidar com problemas de desequilíbrio de classes em tarefas de segmentação, ganhou destaque especialmente em tarefas de segmentação binária, oferecendo vantagens significativas sobre funções de perda tradicionais, como a entropia cruzada binária. Em cenários onde uma classe domina em frequência, modelos treinados com entropia cruzada binária tendem a ser enviesados para a classe majoritária, prejudicando o desempenho geral. A *Focal Loss* mitiga esse problema ao atribuir pesos maiores aos exemplos mal classificados, particularmente àqueles da classe minoritária, equilibrando o aprendizado e melhorando a precisão da segmentação.

A utilidade da *Focal Loss* em aprimorar os resultados da segmentação em imagens médicas foi demonstrada por Karagöz et al. [30]. Em seu estudo, utilizaram o *framework nnU-Net* para a segmentação de lesões na próstata e substituíram a combinação padrão de perda *Dice* e entropia cruzada por *Focal Loss* e entropia cruzada, obtendo resultados superiores. Matematicamente, a *Focal Loss* é definida como:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3)$$

onde:

- p_t é a probabilidade estimada pelo modelo para a classe alvo (após aplicação da função sigmoide).
- α_t é um fator de ponderação que equilibra a importância das classes, geralmente atribuindo maior peso à classe minoritária.
- γ (gama) é um fator de focalização que ajusta o peso atribuído aos exemplos mal classificados. Valores maiores de γ resultam em maior peso para exemplos mal classificados.

A função $(1 - p_t)^\gamma$ atua como um modulador da perda Entropia Cruzada Binária. Ela reduz o peso de exemplos bem classificados (quando p_t se aproxima de 1) e aumenta o peso de exemplos mal classificados (quando p_t se aproxima de 0). O parâmetro α controla o balanceamento entre as classes, enquanto o parâmetro γ ajusta o foco nos exemplos difíceis, permitindo uma personalização da função de perda para diferentes cenários de desequilíbrio de classes.

A função sigmoide, também conhecida como função logística, é uma função matemática amplamente utilizada em modelos de aprendizado de máquina, especialmente em tarefas

de classificação binária. Ela recebe um valor de entrada real (x) e o mapeia para um valor de saída entre 0 e 1, representando a probabilidade de que o exemplo pertença à classe positiva. A Figura 7 ilustra o gráfico da função sigmoide, destacando sua forma característica de "S".

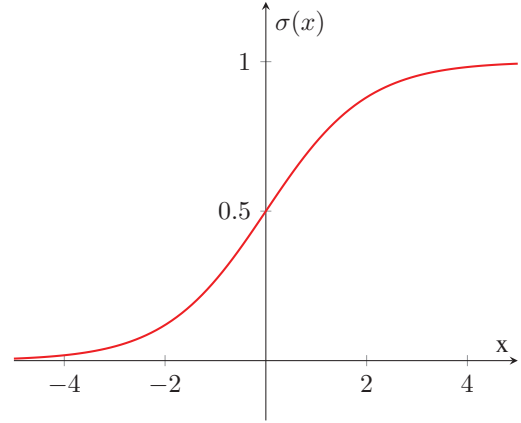


Figura 7. Gráfico da Função Sigmoide
Fonte: O autor (2024)

Matematicamente a função sigmoide é dada por:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

Essa função possui diversas propriedades desejáveis para o aprendizado de máquina. Ela é contínua e diferenciável, o que facilita o cálculo de gradientes durante o treinamento de modelos. Além disso, a sigmoide comprime os valores de entrada em um intervalo limitado, o que ajuda a estabilizar o aprendizado e evitar problemas de gradientes explosivos. Em redes neurais, a função sigmoide é frequentemente utilizada como função de ativação em neurônios, permitindo que a rede modele relações não lineares entre as variáveis de entrada e saída.

c) *Canais de Características (Features)*: O número de canais de características (*features*) em cada camada da *U-Net* determina a capacidade da rede de aprender representações complexas. Um número maior de *features* pode aumentar a capacidade de aprendizado da rede, permitindo que ela modele relações mais intrincadas nos dados. No entanto, um número excessivo de *features* pode levar ao *overfitting*, onde o modelo se ajusta em demasia aos dados de treinamento e apresenta um desempenho insatisfatório em novos dados. A escolha do número ideal de *features* deve considerar a quantidade de dados de treinamento disponíveis e a complexidade da tarefa de segmentação, buscando um equilíbrio entre capacidade de aprendizado e generalização [36].

d) *Passo da Convolução (Stride)*: O passo da convolução (*stride*) define o deslocamento do *kernel* (filtro) ao percorrer a imagem. Um *stride* de 1 significa que o *kernel* se move um pixel por vez, enquanto um *stride* maior, como 2, faz com que o *kernel* pule *pixels* a cada passo.

A escolha do *stride* influencia diretamente a resolução espacial dos mapas de características (*feature maps*) gerados pela rede. *Strides* maiores reduzem a resolução espacial, resultando em *feature maps* menores e mais abstratos, o que pode ser útil para capturar informações contextuais em larga escala [36]. Por outro lado, *strides* menores preservam mais detalhes espaciais, o que pode ser importante para a segmentação precisa de objetos pequenos ou com bordas complexas. No entanto, *strides* menores também podem aumentar o custo computacional e a suscetibilidade ao *overfitting* [56].

Na *U-Net* tradicional, a prática comum é utilizar *strides* de 2 nas camadas de *downsampling* (codificador) para reduzir a resolução espacial e aumentar o campo de visão dos neurônios, permitindo que a rede capture informações contextuais mais amplas [36]. Nas camadas de *upsampling* (decodificador), *strides* de 1 são frequentemente utilizados para manter a resolução espacial e preservar detalhes finos, essenciais para a delimitação precisa de bordas e contornos [36].

e) *Batch Size*: Em aprendizado profundo, o tamanho do lote (*batch size*) refere-se ao número de amostras de imagem processadas simultaneamente durante cada iteração do treinamento de uma rede neural. Essa escolha, apesar de parecer simples, pode impactar significativamente a velocidade, o uso de memória, a estabilidade e, consequentemente, o desempenho final do modelo, como demonstrado em estudos recentes [57] [58].

Durante o treinamento de uma *U-Net*, as imagens são divididas em lotes (*batches*), que são alimentados à rede para calcular as previsões e compará-las com os rótulos reais. O erro resultante é utilizado para atualizar os pesos da rede por meio de um algoritmo de otimização. Esse processo se repete para cada lote até que todas as imagens do conjunto de treinamento sejam processadas, completando uma época de treinamento.

A escolha do tamanho de lote ideal é um desafio, pois envolve um compromisso entre diversos fatores:

- **Velocidade de Treinamento:** Lotes maiores tendem a acelerar o treinamento devido ao processamento paralelo de várias imagens. No entanto, essa abordagem exige mais memória da *GPU* [59].
- **Uso de Memória:** Lotes maiores consomem mais memória da *GPU*, o que pode ser um fator limitante em hardwares com recursos restritos [59].
- **Estabilidade do Treinamento:** Lotes menores podem levar a um treinamento mais ruidoso, com maiores oscilações nos valores do gradiente. Por outro lado, lotes muito grandes podem levar a uma convergência prematura para mínimos locais [60].
- **Generalização:** Lotes menores podem favorecer a generalização do modelo, pois o expõem a uma maior variedade de amostras em cada época [61].

A configuração de treinamento para este trabalho, indicada na Tabela I, foi otimizada limitando o número de *threads* utilizados, melhorando o uso dos recursos da CPU e evitando a sobrecarga do sistema. A validação é realizada a cada época, permitindo o monitoramento contínuo do progresso do treinamento, iniciando desde a primeira época para garantir a avaliação do modelo em

Tabela I
TABELA DE PARÂMETROS DE CONFIGURAÇÃO DE TREINAMENTO

Parâmetro	Valor
max_threads	12
validate_n_epochs	1
validate_min_epoch	0
image_shape	20, 256, 256
num_channels	3
num_classes	2
num_epochs	150
base_lr	0.001
focal_loss_gamma	1.0
model_strides	(2, 2, 2), (1, 2, 2), (1, 2, 2), (1, 2, 2), (2, 2, 2)
model_features	32, 64, 128, 256, 512, 1024

Fonte: O autor (2024)

dados de validação desde o início. As dimensões das imagens de entrada (fatias, altura, largura) foram definidas de acordo com as características do conjunto de dados do desafio *PI-CAI*. O número de canais das imagens de entrada, bem como o número de classes a serem segmentadas, foi especificado, incluindo o fundo e o objeto de interesse. O número total de épocas de treinamento foi determinado juntamente com a taxa de aprendizado inicial. O parâmetro da função de perda *focal* foi configurado, com o valor de α calculado automaticamente como o inverso do equilíbrio de classe de nível de caso para a divisão de treinamento [29]. O passo da convolução foi definido para cada camada da *U-Net*, assim como o número de características em cada camada.

Além disso, o número de amostras processadas em cada iteração (*batch size*) foi ajustado para 8, 16 e 32, a fim de otimizar esse parâmetro e identificar o melhor desempenho do modelo.

4) *Métricas de Avaliação*: Para garantir a confiabilidade do método de segmentação utilizado no desafio *PI-CAI*, diversas métricas de avaliação são empregadas para analisar o desempenho dos modelos, cada uma com suas características e aplicações específicas. Taha & Hanbury [62] introduziram uma ferramenta de avaliação para segmentação de imagens médicas 3D, fornecendo uma análise detalhada de métricas e diretrizes para a seleção de critérios de avaliação apropriados, adaptados a tarefas específicas de segmentação.

a) *Curva de Precisão-Sensibilidade (Precision-Recall Curve)*: Uma ferramenta essencial no processo de avaliação de modelos de segmentação de imagens biomédicas é a curva de precisão-sensibilidade. Essa curva ilustra a relação entre a precisão (P) e a sensibilidade (S) em diferentes limiares de confiança, sendo que:

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$S = \frac{TP}{TP + FN} \quad (6)$$

Onde:

- (TP): Verdadeiros positivos (detecções corretas de anomalias)
- (FP): Falsos positivos (detecções incorretas de anomalias)
- (FN): Falsos negativos (anomalias não detectadas)

Em imagens biomédicas, a curva de precisão-sensibilidade é particularmente útil para avaliar o desempenho de modelos em cenários com grande desequilíbrio entre classes, onde a presença de anomalias é frequentemente muito menor do que a de tecidos saudáveis [63]. A curva permite visualizar a compensação entre precisão e sensibilidade, auxiliando na escolha do limiar de confiança mais adequado para a aplicação em questão.

b) Precisão Média (Average Precision - AP): A Precisão Média (AP) é uma métrica fundamental em tarefas de recuperação de informação, especialmente em sistemas de busca e classificação [64], onde o objetivo é identificar e classificar itens relevantes em um grande conjunto de dados. No contexto da segmentação de imagens, a AP é particularmente significativa, pois permite avaliar a eficiência dos modelos na recuperação de regiões de interesse, como lesões ou tumores [65].

A AP é definida como a área sob a curva precisão-sensibilidade (*precision-recall curve*), capturando o equilíbrio entre precisão e sensibilidade em diferentes limiares de confiança. Matematicamente, a AP é representada como:

$$AP = \sum_n (S_n - S_{n-1}) \cdot P_n \quad (7)$$

Onde:

- P_n : Precisão no n -ésimo limiar
- S_n : Sensibilidade no n -ésimo limiar

A AP oferece uma medida única e informativa do desempenho do modelo, levando em consideração tanto a precisão quanto a sensibilidade em diferentes pontos de operação. Um valor de AP mais alto indica um modelo mais eficiente na recuperação de informações relevantes, minimizando tanto os falsos positivos quanto os falsos negativos.

c) Curva de Características Operacionais do Receptor (Receiver Operating Characteristic - ROC): A curva ROC (*Receiver Operating Characteristic*) é uma ferramenta gráfica amplamente utilizada para avaliar o desempenho de classificadores binários, como os modelos de segmentação de imagens médicas. Ela representa a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (1 - especificidade) em diferentes limiares de decisão. A especificidade (E) é definida como:

$$E = \frac{FP}{FP + TN} \quad (8)$$

Onde:

- (TN): Verdadeiros negativos (detecções corretas de não-anomalias)

A curva ROC (*Receiver Operating Characteristic*) oferece uma visão abrangente do desempenho de um classificador,

permitindo analisar o compromisso entre sensibilidade e especificidade em diferentes pontos de operação. Um classificador ideal apresenta uma curva ROC que se aproxima do canto superior esquerdo do gráfico, indicando alta sensibilidade e alta especificidade. A área sob a curva ROC (*AUC-ROC*) é uma métrica numérica que resume o desempenho geral do classificador, sendo 1 o valor ideal. Essa curva é amplamente utilizada em diversas aplicações para avaliar modelos de classificação binária, incluindo a eficácia de testes diagnósticos na medicina, onde sensibilidade e especificidade são métricas cruciais para determinar a capacidade de um teste em detectar doenças [66].

d) Área Sobre a Curva ROC (Area Under the ROC Curve - AUROC): A AUROC (*Area Under the Receiver Operating Characteristic Curve*) também pode ser expressa como uma soma discreta, aproximando a área sob a curva ROC por meio de retângulos. Essa formulação é útil em situações práticas, onde a curva ROC é construída a partir de um conjunto finito de pontos. A AUROC como soma é dada por:

$$AUROC = \sum_n (E_n - E_{n-1}) \cdot S_n \quad (9)$$

Onde:

- S_n : Sensibilidade no n -ésimo limiar
- E_n : Especificidade no n -ésimo limiar

A AUROC é amplamente utilizada em diversas áreas, incluindo a medicina, para avaliar a precisão de testes diagnósticos e modelos de classificação [66]. Ela oferece uma medida única e agregada do desempenho do classificador, levando em consideração todos os possíveis pontos de operação na curva ROC.

Provost e Fawcett [67] destacam a robustez da AUROC como uma métrica comparativa, especialmente em cenários onde as classes estão desbalanceadas, como é comum em aplicações médicas, onde a prevalência de doenças pode ser baixa. A AUROC é menos sensível a variações na distribuição das classes do que outras métricas, como a acurácia, tornando-a uma escolha mais confiável para a comparação de diferentes modelos.

e) Pontuação Geral de Ranking: No desafio PI-CAI, a pontuação geral de ranking é calculada como a média da AUROC e da AP (*Precisão Média*), representando o desempenho geral do algoritmo de IA tanto na tarefa de classificação (presença ou ausência de lesão) quanto na tarefa de detecção e localização de lesões [16]. A fórmula para calcular a pontuação de ranking é:

$$\text{Ranking} = \frac{AP + AUROC}{2} \quad (10)$$

Essa métrica combinada oferece uma avaliação mais completa do desempenho do modelo, levando em consideração sua capacidade de discriminar entre lesões e não lesões, bem como sua habilidade em localizar e quantificar as lesões presentes na imagem.

5) Tecnologias: Diante dos desafios computacionais inerentes à segmentação de imagens médicas e à complexidade

da arquitetura *U-Net*, a escolha de uma plataforma adequada para o treinamento se torna fundamental. O Google Colab, combinado com a GPU A100, oferece uma solução acessível e poderosa para acelerar o processo de treinamento, permitindo a exploração eficiente de grandes conjuntos de dados e a obtenção de resultados em tempo hábil.

A GPU NVIDIA A100 destaca-se como uma ferramenta poderosa para o processamento paralelo, projetada para acelerar cargas de trabalho de computação intensiva, como *deep learning* e análises de dados. Suas principais características incluem:

- **Tensor Cores:** Unidades de processamento especializadas que aceleram consideravelmente operações de matrizes, como multiplicação e convolução, essenciais para o treinamento de redes neurais profundas.
- **Núcleos CUDA:** A A100 possui um grande número de núcleos CUDA (6912), que são processadores paralelos projetados para executar eficientemente as operações da rede neural.
- **Alta Largura de Banda de Memória:** A A100 oferece alta largura de banda de memória (2 TB/s), permitindo acesso rápido aos dados armazenados na memória da GPU. Essa característica é crucial para o treinamento de modelos complexos, que demandam o carregamento e processamento de grandes volumes de dados.
- **Grande Capacidade de Memória:** Com até 80 GB de memória HBM2e (High Bandwidth Memory), a A100 pode armazenar grandes modelos e conjuntos de dados, reduzindo a necessidade de transferir dados entre a GPU e a CPU, otimizando o processo de treinamento.

O desenvolvimento e treinamento do *framework* de segmentação de lesões prostáticas foram realizados utilizando um conjunto robusto de bibliotecas especializadas em processamento e análise de imagens de RM. A combinação estratégica dessas ferramentas permitiu alcançar um desempenho eficiente e preciso na segmentação. As principais bibliotecas e suas respectivas versões são:

- **MONAI (1.3.1):** *Framework* principal para a construção e treinamento dos modelos de *deep learning*, oferecendo funcionalidades avançadas para o processamento de imagens médicas.
- **Numpy (1.25.2):** Essencial para operações matemáticas e manipulação de *arrays* multidimensionais, facilitando o pré-processamento e a transformação dos dados.
- **Pytorch (2.3.0+cu121):** *Framework* fundamental para a implementação de redes neurais, aproveitando a aceleração de GPU para um treinamento eficiente dos modelos.
- **Nibabel (4.0.2):** Crucial para o acesso e manipulação de formatos de arquivo de neuroimagem, como *NIfTI*, comumente utilizados em imagens de RM.
- **scikit-image (0.19.3):** Utilizada para aplicar técnicas de processamento de imagens, como filtragem e transformação, auxiliando na preparação das imagens para a segmentação.
- **scipy (1.11.4):** Complementa as capacidades do *NumPy*,

forneendo ferramentas adicionais de processamento de imagem e operações matemáticas avançadas.

- **TorchVision (0.18.0+cu121):** Biblioteca auxiliar ao *PyTorch*, contendo implementações de modelos de visão computacional e transformações de imagem.
- **pandas (2.0.3):** Utilizada para a manipulação e análise de dados tabulares, facilitando a organização dos resultados e métricas dos experimentos.
- **pikai_baseline (0.8.5):** Pacote específico para os métodos de *baseline* no contexto do desafio *PI-CAI*, fornecendo arquiteturas e funções padronizadas.
- **pikai_eval (1.4.6):** Ferramenta para avaliação de desempenho dos modelos de segmentação, permitindo a análise detalhada dos resultados e comparações com *benchmarks*.
- **pikai_prep (2.1.6):** Biblioteca destinada ao pré-processamento dos dados de entrada, assegurando a consistência e qualidade das imagens utilizadas no treinamento.
- **SimpleITK (2.3.1):** Utilizada para o processamento e análise de imagens médicas, oferecendo uma interface simples para manipulação e transformação das imagens.
- **batchgenerators (0.25):** Biblioteca para geração de *batches* de dados de treinamento, incluindo técnicas de *data augmentation* para melhorar a robustez do modelo.

II. RESULTADOS E DISCUSSÕES

Considerando a heterogeneidade inerente aos dados médicos, a performance do modelo U-Net com unidades residuais foi submetida a uma avaliação rigorosa através de validação cruzada em cinco etapas com diferentes *batch sizes*. As métricas de AUROC, AP e ranking foram monitoradas ao longo das épocas de treinamento, conforme detalhado na Tabela II.

Os dados da Tabela II destacam a variação nas métricas de desempenho com diferentes tamanhos de *batch*. Observa-se que o *batch size* de 16 proporcionou a melhor média de AUROC (0.704) e ranking (0.517), indicando um equilíbrio ideal entre a estabilidade do treinamento e a capacidade de generalização do modelo.

A média da perda de treino (μ) aumentou conforme o *batch size* foi ampliado, com valores de 15313.523 para *batch size* 8, 26004.638 para *batch size* 16 e 52813.746 para *batch size* 32. No entanto, a variabilidade da perda de treino (σ) também cresceu significativamente, refletindo a maior instabilidade em *batches* maiores, conforme mostrado pelos desvios padrão de 57527.105, 119658.471 e 264437.809, respectivamente.

As métricas de AUROC (μ) mostraram um desempenho superior para *batch sizes* intermediários, com um desvio padrão (σ) relativamente menor (0.111 para *batch size* 16) em comparação com *batches* menores (0.124) e maiores (0.088). Este padrão foi similar para a métrica de AP, onde o *batch size* de 16 alcançou a maior média (0.330) e um desvio padrão moderado (0.137), sugerindo um desempenho mais consistente em relação aos outros tamanhos de *batch*.

Por fim, a métrica de ranking (μ) também indicou um melhor desempenho com *batch size* de 16, alcançando a média mais

Tabela II
TABELA DE RESUMO DAS MÉTRICAS POR *Batch Size*

Batch Size	Perda Treino μ	Perda Treino σ	AUROC μ	AUROC σ	AP μ	AP σ	Ranking μ	Ranking σ
8	15313.523	57527.105	0.656	0.124	0.306	0.134	0.481	0.115
16	26004.638	119658.471	0.704	0.111	0.330	0.137	0.517	0.109
32	52813.746	264437.809	0.687	0.088	0.313	0.176	0.500	0.117

Fonte: O autor (2024)

alta (0.517) e um desvio padrão (σ) de 0.109, confirmando a eficácia desse tamanho de *batch* na classificação das predições.

Essas observações ressaltam a importância de um tamanho de *batch* adequado para otimizar o desempenho do modelo U-Net, balanceando a precisão e a estabilidade durante o treinamento. A escolha do *batch size* de 16 demonstrou ser a mais eficiente em termos de métricas de desempenho, destacando-se como a melhor opção entre os tamanhos avaliados.

A Figura 8 apresenta a média dos rankings ao longo das épocas para os diferentes *batch sizes* (8, 16 e 32). Esse gráfico ilustra como cada *batch size* influenciou a evolução do desempenho do modelo ao longo do treinamento.

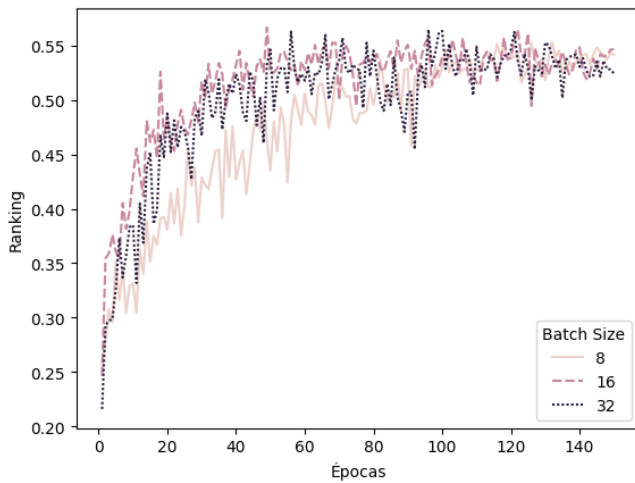


Figura 8. Desempenho de Ranking ao Longo das Épocas para Diferentes Tamanhos de Batch
Fonte: O Autor (2024)

Analisando o gráfico na Figura 8, observa-se que o *batch size* de 16 apresenta uma curva de crescimento mais consistente e atinge um patamar de desempenho superior mais rapidamente em comparação com os *batch sizes* de 8 e 32. Nas primeiras épocas, todos os *batch sizes* mostraram um aumento rápido na média dos rankings, indicando uma fase inicial de aprendizado acelerado.

Contudo, à medida que o treinamento avança, o *batch size* de 16 mantém uma trajetória mais estável e contínua, enquanto os *batch sizes* de 8 e 32 apresentam mais variabilidade, refletida nas flutuações mais intensas das suas curvas. Este comportamento sugere que o *batch size* intermediário (16) não apenas facilita um aprendizado inicial eficiente, mas também

promove uma estabilidade ao longo das épocas que é crucial para a robustez do modelo.

A Figura 9 apresenta a perda de treino ao longo das épocas para os diferentes *batch sizes* (8, 16 e 32). Este gráfico ajuda a visualizar como a perda de treino diminui com o aumento das épocas, indicando a eficiência do processo de aprendizado do modelo.

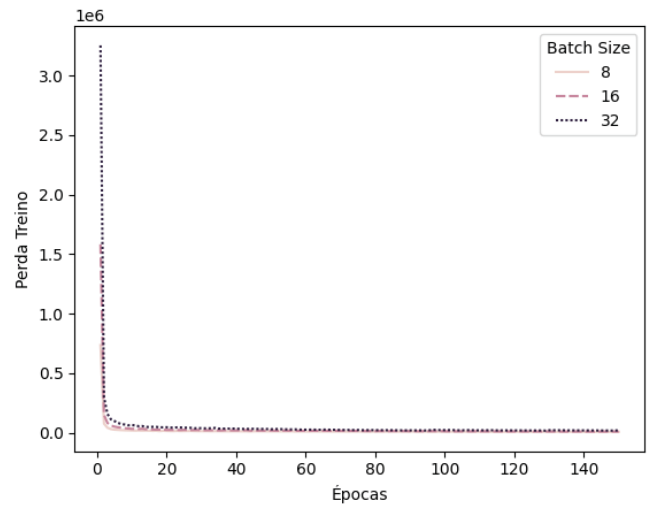


Figura 9. Desempenho de Perda de Treino ao Longo das Épocas para Diferentes Tamanhos de Batch
Fonte: O Autor (2024)

O gráfico contido na Figura 9 mostra que todos os *batch sizes* experimentaram uma rápida diminuição na perda de treino nas primeiras épocas, estabilizando-se à medida que o treinamento progredia. Este comportamento é típico de um treinamento eficaz, onde o modelo rapidamente ajusta seus parâmetros para minimizar o erro.

A Figura 10 apresenta um *boxplot* da métrica de ranking para diferentes *batch sizes*. Este gráfico ilustra a distribuição dos rankings, destacando a mediana e os quartis (excluindo possíveis outliers).

Analisando o Gráfico 10, observa-se que:

- **Batch Size 8:** Apresenta uma maior variabilidade nos rankings, conforme indicado pela faixa interquartil ampla. A mediana está próxima de 0.48, mas a dispersão sugere que os resultados são menos consistentes.
- **Batch Size 16:** Mostra uma menor variabilidade nos rankings, com uma faixa interquartil mais estreita. A

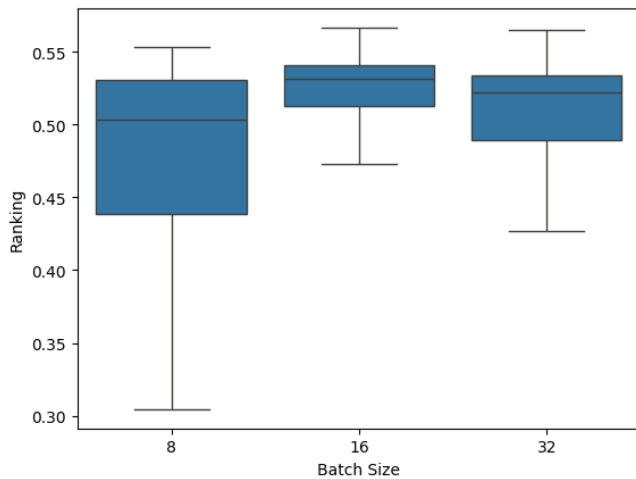


Figura 10. Desempenho de Perda de Treino ao Longo das Épocas para Diferentes Tamanhos de Batch
Fonte: O Autor (2024)

mediana mais alta, em torno de 0.51, e a menor dispersão indicam que este tamanho de *batch* proporciona resultados mais consistentes e estáveis.

- *Batch Size* 32: Tem uma variabilidade nos rankings comparável ao *batch size* 16, mas com uma mediana ligeiramente inferior, em torno de 0.50.

Em suma, o boxplot reforça a análise anterior de que o *batch size* de 16 oferece um equilíbrio ideal entre a precisão e a consistência dos resultados. A menor variabilidade e a mediana mais alta indicam que esse *batch size* é o mais eficaz para o modelo U-Net aplicado ao diagnóstico de câncer de próstata.

A. Considerações Finais

Este estudo técnico investigou a aplicação de uma arquitetura U-Net aprimorada com unidades residuais para a segmentação automática de lesões prostáticas em imagens de ressonância magnética. Utilizando dados do desafio *PI-CAI*, foram avaliados diferentes tamanhos de *batch* (8, 16 e 32) para determinar o impacto no desempenho do modelo. Os resultados indicaram que o *batch size* de 16 proporcionou o melhor desempenho, com uma média de *AUROC* de 0,704, *AP* média de 0,330 e *ranking* médio de 0,517, destacando o equilíbrio entre a estabilidade do treinamento e a capacidade de generalização do modelo. Comparando com o desempenho médio de outros modelos no desafio *PI-CAI*, o nosso modelo apresentou um *ranking* médio máximo de 0,517, enquanto o *leaderboard* do desafio *PI-CAI* [29] alcançou um *ranking* de 0,813.

A principal contribuição deste estudo reside na análise detalhada dos efeitos de diferentes tamanhos de *batch* no desempenho do modelo e fornece insights valiosos sobre a escolha do *batch size* ideal, balanceando eficiência de treinamento e capacidade de generalização. A utilização de um conjunto de dados abrangente e variado como o do desafio

PI-CAI permitiu uma avaliação robusta do modelo, destacando sua aplicabilidade em cenários clínicos reais.

Apesar dos resultados promissores, o estudo apresenta algumas limitações. A heterogeneidade dos dados provenientes de diferentes centros de diagnóstico e equipamentos pode introduzir variações na qualidade das imagens, afetando o desempenho do modelo. Futuras pesquisas poderiam focar em técnicas de normalização mais avançadas para lidar com essa variabilidade. Além disso, as anotações utilizadas para o treinamento, embora combinadas entre especialistas humanos e algoritmos de *IA*, podem conter vieses ou inconsistências que afetam a precisão da segmentação. Investigações adicionais sobre métodos de anotações mais uniformes e precisas são necessárias. A implementação em ambientes clínicos requer uma validação extensiva e possíveis ajustes na arquitetura do modelo para garantir desempenho consistente em larga escala. Estudos futuros devem explorar a escalabilidade do modelo em diferentes ambientes clínicos.

Para futuras direções de pesquisa, é recomendável investigar a integração de outras arquiteturas de redes neurais profundas, como *EfficientNet* ou modelos baseados em atenção, para potencialmente melhorar a precisão e a eficiência da segmentação. Explorar técnicas de enriquecimento de dados, como a geração de dados sintéticos de alta qualidade usando redes generativas adversárias (*GANs*), pode aumentar a diversidade e a representatividade do conjunto de dados de treinamento. Implementar métodos de normalização mais avançados que possam reduzir a variabilidade entre diferentes conjuntos de dados melhorará a consistência do desempenho do modelo. Realizar estudos de validação clínica em colaboração com instituições médicas é crucial para avaliar o desempenho do modelo em ambientes reais e identificar possíveis melhorias na sua aplicação prática.

Em suma, este estudo sublinha a importância da escolha adequada dos hiperparâmetros e da implementação de técnicas avançadas de *data augmentation* e otimização no treinamento de redes neurais para a segmentação de imagens médicas. Futuras pesquisas podem focar na integração de técnicas de *fine-tuning* e aumento do conjunto de dados de treinamento para alcançar resultados ainda melhores, alinhando-se com os padrões de excelência observados nos modelos líderes do *PI-CAI* challenge.

REFERÊNCIAS

- [1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 3, pp. 229–263, 2024.
- [2] I. N. de Câncer (INCA), "Estimativa 2023: Incidência de câncer no Brasil," 2022.
- [3] B. Holmström, M. Johansson, A. Bergh, U.-H. Stenman, G. Hallmans, and P. Stattin, "Prostate specific antigen for early detection of prostate cancer: longitudinal study," *BMJ*, vol. 339, 2009.
- [4] R. Damião, R. T. Figueiredo, M. C. Dornas, D. S. Lima, and M. Koschorke, "Câncer de próstata," *Revista HUPE*, vol. 14, pp. 80–86, 2015.
- [5] J. O. Barentsz, J. Richenberg, R. Clements, P. L. Choyke, S. Verma, G. Villeirs, O. Rouvière, V. Løgager, and J. J. Fütterer, "Esur prostate mr guidelines 2012," *European Radiology*, vol. 22, pp. 746–757, 2012.

- [6] N. Huebner, S. Korn, I. Resch, B. Grubmüller, T. Gross, R. Gale, G. Kramer, N. Poetsch, P. Clauser, A. Haitel, H. Fajković, S. Shariat, and P. Baltzer, "Visibility of significant prostate cancer on multiparametric magnetic resonance imaging (mri)—do we still need contrast media?," *European Radiology*, vol. 31, pp. 3754–3764, 2020.
- [7] J. S. Jue, D. Mikhail, J. González, and M. Alameddine, "Value of mri to improve deep learning model that identifies high-grade prostate cancer. comment on gentile et al. optimized identification of high-grade prostate cancer by combining different psa molecular forms and psa density in a deep learning model. diagnostics 2021, 11, 335," *Diagnostics*, vol. 11, 2021.
- [8] N. I. of Biomedical Imaging and Bioengineering, "Magnetic resonance imaging (mri)," n.d. Accessed: 2024-05-27.
- [9] A. Fordham, C. Hachert, N. Patel, K. Jones, B. Myers, M. Abraham, and J. Gendreau, "Differentiating glioblastomas from solitary brain metastases: an update on the current literature of advanced imaging modalities," *Cancers*, vol. 13, p. 2960, 2021.
- [10] G. Castellano and C. Salmon, "Técnicas quantitativas de imagens por ressonância magnética," *Revista Brasileira De Física Médica*, vol. 13, p. 57, 2019.
- [11] A. Algohary, E. I. Zacharakis, A. L. Breto, M. Alhusseini, V. Wallaengen, I. R. Xu, S. M. Gaston, S. Punnen, P. Castillo, P. M. Pattany, O. N. Kryvenko, B. Spieler, M. C. Abramowitz, A. D. Pra, J. C. Ford, A. Pollack, and R. Stoyanova, "Uncovering prostate cancer aggressiveness signal in t2-weighted mri through a three-reference tissues normalization technique," *NMR in Biomedicine*, vol. 37, no. 3, p. e5069, 2024.
- [12] C.-C. Lai, H.-K. Wang, F.-N. Wang, Y.-C. Peng, T.-P. Lin, H.-H. Peng, and S.-H. Shen, "Autosegmentation of prostate zones and cancer regions from biparametric magnetic resonance images by using deep-learning-based neural networks," *Sensors*, vol. 21, p. 2709, 2021.
- [13] H. Chen, J. Sutedjo, L. Wang, and X. Yin, "Prostate cancer magnetic resonance spectroscopy imaging at 1.5 and 3.0 t: A meta-analysis," *Technology in Cancer Research & Treatment*, vol. 15, pp. 625–631, 2016.
- [14] D. Le Bihan and H. Johansen-Berg, "Diffusion mri at 25: Exploring brain tissue structure and function," *NeuroImage*, vol. 61, no. 2, pp. 324–341, 2012. NEUROIMAGING: THEN, NOW AND THE FUTURE.
- [15] T. Ichikawa, S. M. Erturk, U. Motosugi, H. Sou, H. Iino, T. Araki, and H. Fujii, "High-b-value diffusion-weighted mri in colorectal cancer," *American Journal of Roentgenology*, vol. 187, pp. 181–184, 2006.
- [16] J. S. B. B. v. G. D. Y. M. E. J. V. J. F. M. d. R. H. H. Anindo Saha, Jasper J. Twilt, "Artificial intelligence and radiologists at prostate cancer detection in mri: The pi-cai challenge (study protocol)," 2022.
- [17] L. Johnson, B. Türkbey, W. Figg, and P. Choyke, "Multiparametric mri in prostate cancer management," *Nature Reviews Clinical Oncology*, vol. 11, pp. 346–353, 2014.
- [18] J. Mason, E. Adiotomre, P. Bowles, B. Carey, and A. Henry, "Importance of dynamic contrast enhanced magnetic resonance imaging for targeting biopsy and salvage treatments after prostate cancer recurrence," *Journal of Contemporary Brachytherapy*, vol. 10, no. 6, pp. 570–572, 2018.
- [19] B. Min, D. Kim, B. Cho, Y. Kim, S. Yun, S. Lee, and W. Kim, "Usefulness of a combined approach of t1-weighted, t2-weighted, dynamic contrast-enhanced, and diffusion-weighted imaging in prostate cancer," *Korean Journal of Urology*, vol. 53, p. 830, 2012.
- [20] A. B. Rosenkrantz, L. A. Ginocchio, D. Cornfeld, A. T. Froemming, R. T. Gupta, B. Türkbey, A. C. Westphalen, J. S. Babb, and D. Margolis, "Interobserver reproducibility of the pi-rads version 2 lexicon: a multicenter study of six experienced prostate radiologists," *Radiology*, vol. 280, pp. 793–804, 2016.
- [21] J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura, D. Margolis, M. D. Schnall, F. Shtern, C. M. Tempany, H. C. Thoeny, and S. Verma, "Pi-rads prostate imaging – reporting and data system: 2015, version 2," *European Urology*, vol. 69, no. 1, pp. 16–40, 2016.
- [22] Y. Peng, Y. Jiang, L. Eisengart, M. A. Healy, F. H. Straus, and X. J. Yang, "Computer-aided identification of prostatic adenocarcinoma: Segmentation of glandular structures," *Journal of Pathology Informatics*, vol. 2, no. 1, p. 33, 2011.
- [23] A. Comelli, N. Dahiya, A. Stefano, F. Vernuccio, M. Portoghese, G. Cutaia, A. Bruno, G. Salvaggio, and A. Yezzi, "Deep learning-based methods for prostate segmentation in magnetic resonance imaging," *Applied Sciences*, vol. 11, p. 782, 2021.
- [24] Z. Khan, N. Yahya, K. Alsaihi, S. S. A. Ali, and F. Meriaudeau, "Evaluation of deep neural networks for semantic segmentation of prostate in t2w mri," *Sensors*, vol. 20, p. 3183, 2020.
- [25] R. Sammouda and A. El-Zaart, "An optimized approach for prostate image segmentation using k-means clustering algorithm with elbow method," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–13, 2021.
- [26] Y. Qian, "Prosegnet: A new network of prostate segmentation based on mr images," *IEEE Access*, vol. 9, pp. 106293–106302, 2021.
- [27] M. Soni, I. Khan, K. Babu, S. Nasrullah, A. Madduri, and S. Rahin, "Light weighted healthcare cnn model to detect prostate cancer on multiparametric mri," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–11, 2022.
- [28] D. Xiang, S. Xu, Y. Li, A. Wang, M. Saripan, L. Li, X. Zhang, and L. Lu, "Multi-view secondary input collaborative deep learning for lung nodule 3d segmentation," *Cancer Imaging*, vol. 20, 2020.
- [29] G. Challenge, "The pi-cai challenge."
- [30] A. Karagoz, M. E. Seker, M. Yergin, T. A. Kan, M. S. Kartal, E. Karaarslan, D. Alis, and I. Oksuz, "Prostate lesion estimation using prostate masks from biparametric mri," 2023.
- [31] Y. Zhang, M. T. C. Ying, and D. Z. Chen, "Decompose-and-integrate learning for multi-class segmentation in medical images," 2019.
- [32] M. H. I. S. M. d. R. Joeran S. Bosma, Anindo Saha and H. Huisman, "Semisupervised learning with report-guided pseudo labels for deep learning-based prostate cancer detection using biparametric mri," *Radiology: Artificial Intelligence*, p. e230031, 2023.
- [33] "ITK-SNAP homepage." <http://www.itksnap.org/pmwiki/pmwiki.php>. Acessado em: 21/05/2024.
- [34] B. Wu, Y. Fang, and X. Lai, "Left ventricle automatic segmentation in cardiac mri using a combined cnn and u-net approach," *Computerized Medical Imaging and Graphics*, vol. 82, p. 101719, 2020.
- [35] Z. Dai, E. Carver, C. Liu, J. Lee, A. Feldman, W. Zong, M. Pantelic, M. Elshaikh, and N. Wen, "Segmentation of the prostatic gland and the intraprostatic lesions on multiparametric magnetic resonance imaging using mask region-based convolutional neural networks," *Advances in Radiation Oncology*, vol. 5, no. 3, pp. 473–481, 2020.
- [36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.
- [37] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-unet: A dimension-fusion u shape network for chronic stroke lesion segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 3, pp. 940–950, 2021.
- [38] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- [39] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," *Artificial Neural Networks – ICANN 2010*, pp. 92–101, 2010.
- [40] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.
- [41] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," 2013.
- [42] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.
- [43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [44] P. Kaur, A. K. Sao, and C. K. Ahuja, "Super resolution of magnetic resonance images," *Journal of Imaging*, vol. 7, p. 101, 2021.
- [45] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," 2020.
- [46] M. Brudfors, Y. Balbastre, J. Ashburner, G. Rees, P. Nachev, S. Ourselin, and M. J. Cardoso, "Fitting segmentation networks on varying image resolutions using splatting," 2022.
- [47] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
- [48] MIC-DKFZ, "nnunet documentation." https://github.com/MIC-DKFZ/nnUNet/blob/master/documentation/dataset_format.md, 2023. Acesso em: 25 de maio de 2024.
- [49] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, V. Nath, Y. He, Z. Xu, A. Hatamizadeh, A. Myronenko, W. Zhu, Y. Liu, M. Zheng, Y. Tang, I. Yang, M. Zephyr, B. Hashemian, S. Alle, M. Z. Darestani, C. Budd,

- M. Modat, T. Vercauteren, G. Wang, Y. Li, Y. Hu, Y. Fu, B. Gorman, H. Johnson, B. Genereaux, B. S. Erdal, V. Gupta, A. Diaz-Pinto, A. Dourson, L. Maier-Hein, P. F. Jaeger, M. Baumgartner, J. Kalpathy-Cramer, M. Flores, J. Kirby, L. A. D. Cooper, H. R. Roth, D. Xu, D. Bericat, R. Floca, S. K. Zhou, H. Shuaib, K. Farahani, K. H. Maier-Hein, S. Aylward, P. Dogra, S. Ourselin, and A. Feng, "Monai: An open-source framework for deep learning in healthcare," 2022.
- [50] D. Ushizima, Z. Sordo, P. Andeer, J. Sethian, and T. Northen, "Rhizonet: Image segmentation for plant root in hydroponic ecosystem," *bioRxiv*, 2023.
- [51] A. P. Apte, E. LoCastro, A. Iyer, J. Jiang, J. H. Oh, H. Veeraraghavan, A. Shukla-Dave, and J. O. Deasy, "Technical note: Stratis: A cloud-enabled software toolbox for radiotherapy and imaging analysis," *bioRxiv*, 2022.
- [52] D. Ezzat, A. E. Hassanien, and H. A. Ella, "An optimized deep learning architecture for the diagnosis of covid-19 disease based on gravitational search optimization," *Applied Soft Computing*, vol. 98, p. 106742, 2021.
- [53] M. Rozak, J. Mester, A. Attarpour, A. Dorr, M. Goubran, and B. Stefanovic, "A deep learning pipeline for mapping in situ network-level neurovascular coupling in multi-photon fluorescence microscopy," *bioRxiv*, 2024.
- [54] D. Jha, P. H. Smedsrud, D. Johansen, T. de Lange, H. D. Johansen, P. Halvorsen, and M. A. Riegler, "A comprehensive study on colorectal polyp segmentation with resnet++, conditional random field and test-time augmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 2029–2040, 2021.
- [55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.
- [56] Özgün Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," 2016.
- [57] L. N. Smith, "Cyclical learning rates for training neural networks," 2017.
- [58] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," 2018.
- [59] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," 2012.
- [60] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," 2017.
- [61] E. Hoffer, I. Hubara, and D. Soudry, "Train longer, generalize better: closing the generalization gap in large batch training of neural networks," 2018.
- [62] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, 2015.
- [63] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," p. 233–240, 2006.
- [64] C. Buckley and E. M. Voorhees, "Retrieval evaluation with incomplete information," p. 25–32, 2004.
- [65] A. Brown, W. Xie, V. Kalogeiton, and A. Zisserman, "Smooth-ap: Smoothing the path towards large-scale image retrieval," 2020.
- [66] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve.," *Radiology*, vol. 143, pp. 29–36, 1982.
- [67] F. Provost and T. Fawcett, "Robust classification for imprecise environments," *Machine Learning*, vol. 42, pp. 203–231, 2001.