

DAVI SANTOS DE LIMA

**FERRAMENTA DE SEGMENTAÇÃO
AUTOMÁTICA E VISUALIZAÇÃO 3D DE
TUMORES CEREBRAIS UTILIZANDO
ITK-SNAP”**

São Paulo
2025

DAVI SANTOS DE LIMA

**FERRAMENTA DE SEGMENTAÇÃO
AUTOMÁTICA E VISUALIZAÇÃO 3D DE
TUMORES CEREBRAIS UTILIZANDO
ITK-SNAP”**

Trabalho apresentado à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Engenheiro Eletricista com
ênfase em Sistemas Eletrônicos.

São Paulo
2025

DAVI SANTOS DE LIMA

**FERRAMENTA DE SEGMENTAÇÃO
AUTOMÁTICA E VISUALIZAÇÃO 3D DE
TUMORES CEREBRAIS UTILIZANDO
ITK-SNAP”**

Trabalho apresentado à Escola Politécnica
da Universidade de São Paulo para obtenção
do Título de Engenheiro Eletricista com
ênfase em Sistemas Eletrônicos.

Orientador:
Hae Yong Kim

São Paulo
2025

À mulher que me ensinou o que é
coragem

AGRADECIMENTOS

Agradeço à minha mãe, que me ensinou que a vida é boa; aos meus irmãos, que me acompanham desde o início de tudo; aos amigos que fiz nesta jornada, que fizeram desta a melhor temporada dessa história; ao homem com quem compartilho a vida, que me ensina diariamente a confiar mais em mim mesmo; ao Argo, que me possibilitou conhecer um mundo incrível; aos meus professores, que contribuíram para o meu crescimento; e, por fim, ao meu pai, que me ensinou a maior lição que carrego sempre comigo.

“Benditas coisas que eu não sei”

-- Mart'nália & Zélia Duncan

RESUMO

Este trabalho apresenta uma abordagem automatizada para segmentação de tumores cerebrais em imagens de ressonância magnética (MRI), com foco em gliomas, utilizando técnicas de deep learning. O pipeline incluiu pré-processamento cuidadoso dos dados, análise exploratória de metadados e preparação de volumes 3D, garantindo qualidade e consistência nas entradas do modelo. A segmentação foi realizada com o nnU-Net 3D, alcançando resultados robustos e confiáveis na identificação das diferentes regiões tumorais. A avaliação quantitativa mostrou desempenho elevado em termos de precisão e concordância com segmentações manuais, demonstrando a capacidade do modelo de aprender representações significativas mesmo a partir de conjuntos de dados limitados. Os achados reforçam o potencial de métodos automatizados para apoiar decisões clínicas, planejamento terapêutico e pesquisas em neuroimagem, além de estabelecer uma base para futuras melhorias em arquiteturas e estratégias de aprendizado profundo.

Palavras-Chave – Segmentação de imagens médicas; Glioma; Deep Learning; nnU-Net; Ressonância magnética; Pré-processamento; Avaliação quantitativa.

ABSTRACT

This work presents an automated approach for brain tumor segmentation in magnetic resonance imaging (MRI), focusing on gliomas, using deep learning techniques. The pipeline included careful data preprocessing, exploratory analysis of metadata, and preparation of 3D volumes, ensuring quality and consistency of model inputs. Segmentation was performed using 3D nnU-Net, achieving robust and reliable results in identifying different tumor regions. Quantitative evaluation demonstrated high accuracy and strong agreement with manual segmentations, showing the model's ability to learn meaningful representations even from limited datasets. These findings highlight the potential of automated methods to support clinical decision-making, therapeutic planning, and neuroimaging research, providing a foundation for future improvements in architectures and deep learning strategies.

Keywords – Medical image segmentation; Glioma; Deep Learning; nnU-Net; Magnetic Resonance Imaging; Preprocessing; Quantitative evaluation.

SUMÁRIO

1	Introdução e Objetivos	10
1.1	Objetivo Geral	10
1.2	Objetivos Específicos	10
2	Relevância Clínica e Valor da Visualização 3D	12
3	Estado da Arte	15
3.1	Automated Brain Tumor Segmentation Using Multimodal Brain Scans: A Survey Based on Models Submitted to the BraTS 2012–2018 Challenges	15
3.1.1	Trends in Development of Novel Machine Learning Methods for the Identification of Gliomas in Datasets That Include Non-Glioma Images: A Systematic Review	16
3.1.2	Advancing Precision: A Comprehensive Review of MRI Segmentation Datasets from BraTS Challenges (2012–2025)	17
3.1.3	A Review on Brain Tumor Segmentation Based on Deep Learning Methods with Federated Learning Techniques	19
3.1.4	Advanced Deep Learning and Machine Learning Techniques for MRI Brain Tumor Analysis: A Review	20
4	Metodologia	21
4.1	Dataset	21
4.2	Pré-processamento e análise exploratória dos dados	23
4.3	Modelo e Treinamento	24
4.4	Inferência e Geração de Máscaras	25
4.5	Métricas de Avaliação e Análise Estatística	26
4.6	Questões Éticas, Privacidade e Uso Responsável	26

5 Plano de Experimentos	27
5.1 Análise Exploratória dos Dados e Extração de Subsets	28
5.2 Experimento 1: Baseline nnU-Net	30
5.2.1 Preparação dos dados experimentais	31
5.2.2 Preparação do ambiente de execução	32
5.2.3 Execução e Resultados	33
5.3 Experimento 2: PCA dos canais + nnU-Net	37
5.3.1 Implementação do PCA	37
5.3.2 Execução e Resultados	38
5.4 Experimento 3: Attention Gates + nnU-Net	41
5.4.1 Implementação	41
5.4.2 Resultados	45
6 Escolha do Modelo e Treinamento Final	47
6.1 Escolha do Modelo	48
6.2 Treinamento Final	49
6.3 Custos e Limitações	54
7 Integração com o ITK-SNAP	55
7.1 Preparação do ambiente no WSL	55
7.2 Desenvolvimento da interface gráfica (GUI)	55
7.3 Execução da inferência com o nnU-Net v2	56
7.4 Integração com o ITK-SNAP	56
7.5 Resultado final e benefícios do fluxo integrado	57
8 Melhorias e Próximos Passos	59
9 Conclusão	61

Referências	63
Apêndice A – Repositório de Códigos e Reprodutibilidade	65

1 INTRODUÇÃO E OBJETIVOS

A volumetria tumoral em 3D (tridimensional) é reconhecida como uma ferramenta mais informativa do que medições bidimensionais para avaliação de progressão, resposta a tratamento e planejamento cirúrgico em neuro-oncologia. Entretanto, a segmentação manual é demorada e sujeita à variabilidade inter e intraobservador. O avanço de métodos de deep learning para segmentação semântica, em particular arquiteturas do tipo U-Net (U-shaped Network) [1] e variantes auto-configuráveis como o nnU-Net (no-new-U-Net) [2], viabiliza a criação de pré-segmentações automáticas de boa qualidade, que podem ser revistas rapidamente por especialistas.

1.1 Objetivo Geral

Desenvolver e avaliar uma ferramenta que realize segmentação automática de tumores cerebrais em MRI (Magnetic Resonance Imaging, ou Ressonância Magnética) e integre as máscaras resultantes ao ITK-SNAP (Insight Segmentation and Registration Toolkit - SNAP) [3] para visualização tridimensional e análise volumétrica, visando reduzir o tempo de trabalho manual e aumentar a reproduzibilidade das análises.

1.2 Objetivos Específicos

1. Implementar um pipeline de pré-processamento de imagens de ressonância magnética (MRI, Magnetic Resonance Imaging), com etapas como conversão de arquivos DICOM (Digital Imaging and Communications in Medicine) para NIfTI (Neuroimaging Informatics Technology Initiative) [4], normalização e reamostragem.
2. Treinar um modelo baseline (nnU-Net ou U-Net) utilizando um conjunto de dados de referência (dataset do BraTS 2024 - Brain Tumor Segmentation Challenge) [5].
3. Construir um script de inferência que gere máscaras NIfTI (Neuroimaging Informatics Technology Initiative) compatíveis com o ITK-SNAP [3].

4. Integrar o fluxo com o ITK-SNAP (Insight Segmentation and Registration Toolkit - SNAP), permitindo revisão manual e cálculo de volume de forma simples.
5. Avaliar a qualidade das segmentações com métricas padronizadas, como Dice Score (DSC, Dice Similarity Coefficient) [6].
6. Documentar a solução e discutir limitações, riscos e trabalhos futuros.

2 RELEVÂNCIA CLÍNICA E VALOR DA VISUALIZAÇÃO 3D

A visualização do tumor em 3D fornece uma representação espacial intuitiva de forma e extensão, favorecendo o planejamento cirúrgico ao permitir compreender relações com estruturas críticas (feixes vasculares e áreas funcionais). Além disso, o cálculo volumétrico preciso é fundamental para acompanhar a evolução do paciente e avaliar a resposta terapêutica. Em contraste, medições lineares em 2D (ou estimativas simples) podem subestimar ou superestimar o volume em tumores de morfologia irregular.

A utilização de múltiplas aquisições de ressonância magnética (canais) é um elemento central não apenas para a segmentação automática, mas também para a compreensão clínica do tumor e de seus efeitos sobre o tecido cerebral. Cada sequência de ressonância enfatiza características biológicas distintas do cérebro e dos tumores: a imagem ponderada em T1 fornece uma referência anatômica detalhada; T1 com contraste (T1ce) realça regiões com barreira hematoencefálica comprometida; T2-weighted (T2) destaca o conteúdo de fluido e edema peritumoral; e a sequência FLAIR suprime o sinal do líquido cefalorraquidiano, melhor evidenciando lesões hiperintensas adjacentes e infiltrativas. Essa diversidade de informação complementar foi demonstrada empiricamente: modelos que utilizam quatro modalidades (T1, T1ce, T2 e FLAIR) alcançam melhor precisão de segmentação em comparação com modelos unicanal, e abordagens multimodais geralmente superam configurações com menos sequências, pois aproveitam a informação complementar entre os diferentes pesos de intensidade de cada sequência [2,5].

A importância clínica dessa diferenciação está diretamente ligada às sub-regiões tumorais definidas no desafio BraTS. A segmentação não busca apenas identificar a presença de uma massa anormal, mas discriminar componentes relevantes para prognóstico e planejamento terapêutico:

- **Tumor total (Whole Tumor – WT):** comprehende todas as anomalias associadas ao tumor, incluindo edema e tecido infiltrado, usualmente melhor capturado em T2

e FLAIR [5].

- **Tumor core (TC)**: agrupa a massa tumoral propriamente dita, incluindo áreas de realce e necrose [5].
- **Enhancing tumor (ET)**: regiões que exibem realce após administração de contraste, indicando angiogênese e atividade tumoral agressiva [5].
- **Necrotic and non-enhancing core (NCR/NET)**: representa o centro necrótico ou componentes sem contraste dentro da massa tumoral [7].

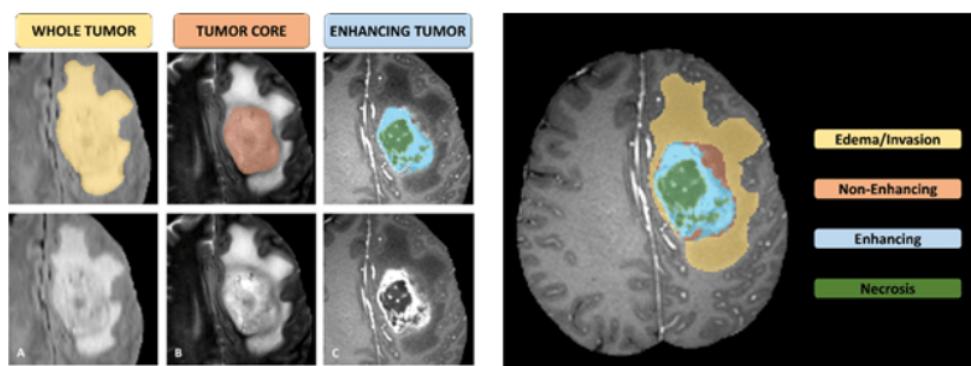


Figura 1: Glioma sub-regions. The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (A), the tumor core (red) visible in T2 (B), the active tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (C). The segmentations are combined to generate the final labels of the tumor sub-regions (D): ED (yellow), NET (red), NCR cores (green), AT (blue). Fonte: https://www.researchgate.net/figure/Glioma-sub-regions-The-image-patches-show-from-left-to-right-the-whole-tumor-yellow_fig2_332383804

Do ponto de vista clínico, a segmentação dessas regiões permite medidas volumétricas e espaciais detalhadas essenciais para planejamento cirúrgico, radioterápico e acompanhamento longitudinal. Por exemplo, a proporção de volume do ET pode refletir agressividade tumoral, enquanto a extensão do edema influencia decisões sobre descompressão ou terapia adjuvante. Terminações periféricas do edema são frequentemente locais de recidiva, tornando sua quantificação importante prognostica.

A interação entre diferentes canais de imagem e regiões tumorais também impacta modelos de aprendizado profundo: a combinação de T1ce com FLAIR frequentemente captura tanto a extensão do tumor ativo quanto a resposta de edema, equilibrando precisão e menor redundância de informação [8].

Quando integrada a uma ferramenta de visualização como o ITK-SNAP, a segmentação automatizada de múltiplas sub-regiões torna-se uma ferramenta de apoio ao especialista,

permitindo inspeção qualitativa e revisão humana [3]. A combinação de imagens multimedias com máscaras segmentadas em 3D oferece uma visão integrada da patologia, apoiando decisões clínicas que vão além de simples medidas lineares bidimensionais.



Figura 2: Logo ITK-SNAP. Fonte: <https://www.itksnap.org/pmwiki/pmwiki.php>

3 ESTADO DA ARTE

A segmentação de tumores cerebrais em imagens de ressonância magnética (MRI) evoluiu rapidamente ao longo da última década. As competições BraTS impulsionaram avanços na curadoria de dados, protocolos de pré-processamento e validação, enquanto revisões sistemáticas destacam a migração de pipelines clássicos para arquiteturas 2D/3D baseadas em U-Net, V-Net, conexões densas e atenção, além de técnicas mais recentes como aprendizado federado, aprendizado auto-supervisionado e estratégias de robustez (ensembles, incerteza, test-time augmentation). A seguir, estão sintetizadas as principais inovações relatadas em cinco artigos de revisão selecionados, relacionando cada inovação às fontes correspondentes.

3.1 Automated Brain Tumor Segmentation Using Multimodal Brain Scans: A Survey Based on Models Submitted to the BraTS 2012–2018 Challenges

Este artigo de revisão, publicado em 2020 na *IEEE Reviews in Biomedical Engineering*, oferece uma análise abrangente da evolução dos métodos automatizados de segmentação de tumores cerebrais baseada em imagens multimodais de ressonância magnética, comparando-os pelo benchmark BraTS entre 2012 e 2018 [8].

Os autores destacam a transição dos tradicionais métodos baseados em modelagem estatística clássica — como Markov Random Fields (MRF), representações esparsas e segmentação via grafos — para arquiteturas modernas de aprendizado profundo, especialmente CNNs 2D e 3D que incorporam diretamente informações volumétricas. Essa mudança refletiu a necessidade de modelos capazes de lidar com a heterogeneidade dos tumores em forma, localização e intensidade de sinal das imagens.

Além da adoção de redes convolucionais, observaram-se tendências inovadoras como a segmentação em cascata, onde primeiro se identifica o tumor inteiro (“whole tumor”) e depois suas sub-regiões mais específicas, como o núcleo tumoral (“core”) e a área com con-

traste (“enhancing”), melhorando a precisão em cada fase. O uso de redes 3D patch-based, exemplificado por modelos como o DeepMedic, permitiu explorar o contexto tridimensional sem exigir excesso de memória computacional.

As arquiteturas U-Net e V-Net foram consolidadas com saltos de conexão (“skip-connections”) e *loss functions* adaptadas para lidar com o desbalanceamento entre regiões tumorais — como Dice, Generalized Dice e Focal Loss. O pipeline desses métodos também passou a incluir pós-processamento probabilístico ou morfológico, usando técnicas como CRF (Conditional Random Field) 3D ou análise de componentes conectadas para reduzir os falsos positivos. Por fim, práticas de *ensemble* — combinando múltiplos modelos ou execuções de treino — emergiram como estratégia eficaz para aumentar a robustez e estabilidade dos resultados.

No plano prático, essa evolução metodológica resultou em melhorias consistentes nas métricas de desempenho, especialmente no índice de Sobreposição Dice para as sub-regiões tumorais ao longo dos anos 2012 a 2018. A padronização das modalidades de imagem usadas e dos procedimentos de pré-processamento facilitou a comparabilidade entre métodos, impulsionando o progresso na área.

3.1.1 Trends in Development of Novel Machine Learning Methods for the Identification of Gliomas in Datasets That Include Non-Glioma Images: A Systematic Review

Esta revisão sistemática publicada em 2021 na *Frontiers in Oncology* investigou como técnicas de machine learning têm sido usadas para identificar gliomas em conjuntos de dados que incluem imagens normais ou de outras patologias — um desenho mais próximo da prática clínica. A busca, conduzida por bibliotecários, abrangeu Ovid Embase, Ovid MEDLINE, Cochrane CENTRAL e Web of Science, foi registrada no PROSPERO (International Prospective Register of Systematic Reviews; CRD42020209938) e seguiu as diretrizes PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), o que reforça a transparência do processo.

Do total de 11.727 registros inicialmente encontrados, apenas 12 estudos preencheram os critérios: sete focavam distinguir exames normais de anormais em coleções que continham gliomas e cinco buscavam separar imagens de glioma de não-glioma. Os conjuntos de dados mais frequentes eram de instituição única (cinco artigos, quatro deles do Harvard Medical School), seguidos por BRATS (três), multicêntricos (dois) e TCIA (dois); a amostra mediana relatada foi de 280 pacientes. Nos BRATS e TCIA, o ground truth foi

definido por patologia; já no dataset de Harvard e em parte dos multicêntricos o método de referência não estava claro.

Metodologicamente, os estudos foram heterogêneos. Redes neurais predominaram (10/12), mas as estratégias de avaliação variaram: cinco usaram five-fold cross-validation e sete separaram conjuntos de treino e teste dentro do mesmo dataset — desenhos que, em geral, ainda não testam de forma robusta a generalização entre centros. As acurárias reportadas variaram de 0,75 a 1,00 (mediana 0,96), mas a qualidade do relato foi apenas moderada: a aderência média ao TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) foi de 0,50 (DP 0,14; 0,37–0,85).

A mensagem central é de cautela: apesar de resultados numéricos animadores, persistem limitações que dificultam a aplicação clínica — bases pouco heterogêneas, estratégias de treino/teste pouco generalizáveis e relatos incompletos. Os autores recomendam priorizar validação externa em bancos independentes e maior adesão a padrões de relato para que os modelos realmente se sustentem fora do ambiente controlado de pesquisa.

Em resumo, embora centrado em identificação/diagnóstico, o trabalho ressalta que segmentadores robustos devem ser validados em cenários que incluem casos negativos e variações de aquisição, com protocolos de teste multicêntricos e reportes padronizados.

3.1.2 Advancing Precision: A Comprehensive Review of MRI Segmentation Datasets from BraTS Challenges (2012–2025)

O artigo “Advancing Precision: A Comprehensive Review of MRI Segmentation Datasets from BraTS Challenges (2012–2025)” (Sensors, 2025) traça a trajetória evolutiva dos conjuntos de dados do BraTS e como mudanças na curadoria e nos protocolos implicaram avanços práticos na pesquisa de segmentação de tumores cerebrais. Ao longo dos anos, os organizadores ampliaram progressivamente o tamanho das bases de treinamento e passaram a incorporar dados multi-institucionais, o que aumentou a variabilidade de aquisição e melhorou a generalização dos modelos quando bem tratados. Paralelamente, houve um refinamento sistemático dos protocolos de anotação — com múltiplos anotadores e mecanismos de consenso — e esforços de harmonização cross-site para reduzir a variabilidade interobservador.

No pré-processamento, práticas padronizadas como registro espacial, normalização de intensidade, correção de bias-field e padronização do spacing foram consolidadas, permitindo treinos 3D mais estáveis e comparáveis entre estudos.

No contexto de imagens de ressonância magnética (MRI), o bias-field se refere a variações lentas e não uniformes na intensidade do sinal da imagem, que podem ocorrer devido a imperfeições no campo magnético do equipamento ou na sensibilidade do receptor. Essas variações não correspondem à anatomia real e podem prejudicar algoritmos de segmentação, pois regiões semelhantes podem apresentar intensidades diferentes. A correção de bias-field é, portanto, um procedimento de pré-processamento que visa ajustar a imagem, uniformizando essas intensidades e tornando os dados mais consistentes para análise computacional.

Já a padronização do spacing diz respeito ao espaçamento físico entre os voxels da imagem 3D. Como diferentes scanners e protocolos podem gerar imagens com resoluções distintas (por exemplo, voxels mais alongados em uma direção do que em outra), essa padronização transforma todas as imagens para um mesmo espaçamento isotrópico, garantindo que cada voxel represente o mesmo volume real. Isso é fundamental para treinar modelos de deep learning de forma estável, comparar volumes e métricas entre pacientes, e aplicar algoritmos 3D que assumem uma geometria uniforme do espaço.

Além disso, o BraTS evoluiu ao oferecer tarefas auxiliares (por exemplo, predição de sobrevivência, quantificação de incerteza, detecção de recorrência) que estimulam arquiteturas multi-tarefa e aplicações clínicas mais ricas.

Essas mudanças na curadoria e nos desafios se refletiram diretamente nas escolhas de arquitetura e nas práticas de treinamento utilizadas pela comunidade. Modelos migraram para soluções 3D mais eficientes — frequentemente patch-based ou baseadas em estratégias de sliding window e downsampling agressivo — para conciliar o contexto volumétrico com limitações de memória; técnicas de mixed precision e janelas deslizantes tornaram-se comuns em implementações de produção e pesquisa. Do lado das perdas e métricas, há maior uso de funções compostas (por exemplo, combinações envolvendo Dice) e métricas alinhadas ao objetivo clínico (como Hausdorff-95) para capturar tanto sobreposição quanto erro de contorno. Finalmente, a disponibilização contínua de conjuntos de validação, leaderboards públicos e avaliações multicêntricas fomentou reproduzibilidade e benchmarking transparente, ao mesmo tempo em que deixou explícito que a verdadeira robustez exige testes multicêntricos independentes e relatórios metodológicos padronizados.

3.1.3 A Review on Brain Tumor Segmentation Based on Deep Learning Methods with Federated Learning Techniques

A revisão publicada em 2024 na Biocybernetics and Biomedical Engineering aborda o estado atual da segmentação de tumores cerebrais utilizando técnicas de aprendizado profundo (deep learning, DL) e a integração do aprendizado federado (federated learning, FL) para preservar a privacidade dos dados e permitir colaboração entre instituições.

O aprendizado federado permite que modelos sejam treinados de forma colaborativa em múltiplas instituições sem a necessidade de compartilhar dados sensíveis. Isso é alcançado através da agregação de atualizações de modelos locais, preservando a privacidade dos dados originais. Técnicas como o algoritmo FedAvg (Federated Averaging) e suas variações são comumente utilizadas nesse contexto.

Entretanto, a heterogeneidade dos dados entre diferentes instituições, conhecida como distribuição não independente e identicamente distribuída (non-IID), representa um desafio significativo. Para mitigar esse problema, estratégias como ponderação por tamanho ou qualidade dos dados de cada cliente, personalização de camadas específicas do modelo e agregação robusta são empregadas.

Além disso, a privacidade e segurança dos dados são preocupações centrais. Técnicas como agregação segura, ruído diferencial e auditoria de gradientes são implementadas para reduzir o risco de reconstrução de dados sensíveis a partir das atualizações do modelo.

A integração do aprendizado federado com arquiteturas de DL, como U-Net, V-Net e nnU-Net, permite treinos colaborativos entre diferentes centros, utilizando normalização adaptativa e sincronização de estatísticas, como Batch Normalization e Instance Normalization, para melhorar a performance do modelo.

Porém, desafios como o desbalanceamento severo entre os dados de diferentes clientes e o desvio de domínio (domain shift) devido às variações nos scanners utilizados exigem a implementação de normalização robusta e técnicas de adaptação de domínio federado.

Para avaliar efetivamente os ganhos do aprendizado federado, é essencial o uso de métricas padronizadas, como o coeficiente de Dice por região, especialmente em validações externas.

Em resumo, a combinação de aprendizado profundo com aprendizado federado oferece uma abordagem promissora para a segmentação de tumores cerebrais, permitindo a colaboração entre instituições e a preservação da privacidade dos dados, ao mesmo tempo em que enfrenta desafios técnicos relacionados à heterogeneidade dos dados e à segurança.

3.1.4 Advanced Deep Learning and Machine Learning Techniques for MRI Brain Tumor Analysis: A Review

A revisão publicada em 2025 na revista Sensors aborda os avanços recentes em técnicas de aprendizado profundo (deep learning, DL) e aprendizado de máquina (machine learning, ML) aplicados à análise de tumores cerebrais em imagens de ressonância magnética (magnetic resonance imaging, MRI). O estudo destaca o uso de arquiteturas avançadas, como redes neurais convolucionais tridimensionais (3D) combinadas com módulos de atenção, Transformers e blocos Axial/Windowed, para capturar contextos de longo alcance nas imagens. Além disso, enfatiza a adoção de aprendizado auto-supervisionado por meio de tarefas pretextar e técnicas de aumento de dados no momento do teste (test-time augmentation, TTA) para melhorar a robustez e a calibração da incerteza dos modelos.

Os pipelines de análise foram aprimorados para integrar informações multimodais, incluindo diferentes sequências de MRI (T1: ponderação em tempo de relaxamento longitudinal, T1ce: T1 com contraste, T2: ponderação em tempo de relaxamento transversal, FLAIR: fluid-attenuated inversion recovery) e dados clínicos ou metabólicos, quando disponíveis. Estratégias de aprendizado sensível ao custo (cost-sensitive learning) e o uso de funções de perda compostas, como combinações de Dice, entropia cruzada (cross-entropy) e métricas de contorno como Hausdorff, foram implementadas para melhorar a precisão na delimitação dos tumores.

Além disso, a explicabilidade dos modelos foi abordada por meio de técnicas como Grad-CAM (gradient-weighted class activation mapping), mapas de saliência (saliency maps) e SHAP (SHapley Additive exPlanations), visando aumentar a confiança na adoção clínica dos modelos. Observou-se uma tendência crescente para frameworks auto-configuráveis, como o nnU-Net (no-new-U-Net), que servem como base sólida para a integração de módulos de atenção e Transformers. Por fim, a revisão destaca a importância da validação multicêntrica, calibração e monitoramento pós-implantação (drift) para garantir a eficácia e a segurança dos modelos em ambientes clínicos reais.

4 METODOLOGIA

4.1 Dataset

O dataset do BraTS consiste em imagens de ressonância magnética multiparamétrica (mpMRI) de pacientes com glioma, coletadas rotineiramente após o tratamento em sete diferentes centros médicos acadêmicos [5, 9].



Figura 3: Logo Brain Tumor Segmentation. Fonte: <http://braintumorsegmentation.org/>

A Ressonância Magnética Multiparamétrica refere-se a uma técnica avançada de ressonância magnética que combina diferentes sequências ou parâmetros de aquisição em uma única análise, permitindo uma avaliação mais completa e detalhada do tecido cerebral. Na prática clínica e em pesquisas, o mpMRI normalmente inclui sequências como T1 ponderada, que destaca a anatomia e o contraste com agentes de contraste; T1ce, que corresponde à T1 ponderada com realce de contraste e evidencia áreas de barreira hematoencefálica comprometida, típicas de tumores ativos; T2, sensível a alterações de água, edema e lesões hiperintensas; FLAIR (Fluid-Attenuated Inversion Recovery), que suprime o sinal do líquido cefalorraquidiano e destaca lesões adjacentes a ventrículos ou áreas de edema; DWI (Diffusion Weighted Imaging), que avalia a difusão da água e é útil para caracterizar tecido tumoral e áreas de necrose; e imagens de perfusão, geralmente chamadas PWI (Perfusion Weighted Imaging), que medem o fluxo sanguíneo local, auxiliando na avaliação da vascularização tumoral [9].

O principal objetivo do mpMRI é fornecer informações complementares sobre morfologia, função, perfusão e microestrutura dos tecidos, aumentando a acurácia no diagnóstico, na caracterização tumoral e no planejamento terapêutico, especialmente em

neuro-oncologia [5, 9]. Essa abordagem multiparamétrica permite aos especialistas obter uma visão integrada do tumor e das estruturas adjacentes, fornecendo uma base mais sólida para decisões clínicas e pesquisas.

Para cada paciente, as sub-regiões do tumor foram anotadas e validadas por neuroradiologistas especializados, garantindo uma referência confiável (ground truth) para avaliar quantitativamente a acurácia das segmentações previstas pelos algoritmos [2, 9].

São 1809 imagens no total, sendo 1621 imagens com suas respectivas máscaras segmentadas e 188 imagens sem anotação. As 1621 imagens que possuem segmentação são aquisições de 731 pacientes diferentes [5, 9].

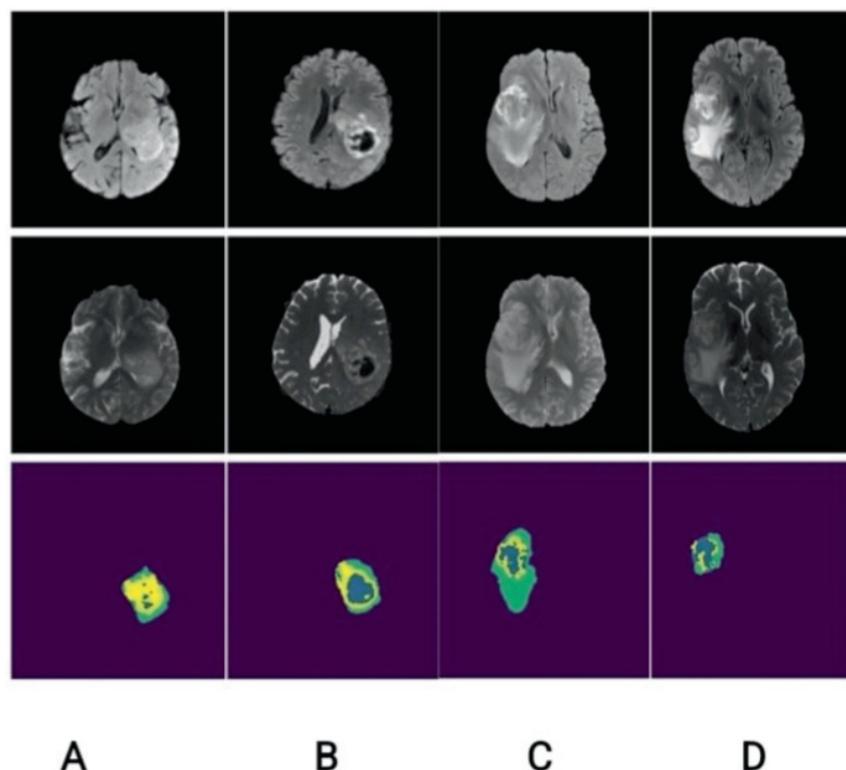


Figura 4: Exemplos do dataset. Fonte: <https://www.synapse.org/Synapse:syn53708126/wiki/626320>

Em resumo, é um conjunto multimodal, multi-institucional e anotado por especialistas, usado para treinar, validar e testar modelos de segmentação de tumores cerebrais [5, 9].

4.2 Pré-processamento e análise exploratória dos dados

O pré-processamento das imagens médicas foi realizado utilizando a pipeline automatizada do nnU-Net [2], que segue um conjunto de etapas específicas para preparar os dados de forma otimizada para tarefas de segmentação. Esse processo de pré-processamento foi fundamental para garantir a consistência e qualidade dos dados antes de serem alimentados na rede neural.

Inicialmente, as imagens foram convertidas do formato DICOM para o formato NIfTI, o que é uma prática comum em neuroimagem [5], pois o formato NIfTI é amplamente compatível com ferramentas de análise de imagem médica e facilita o armazenamento de volumes tridimensionais. Essa conversão assegura a preservação da estrutura espacial e das informações importantes de cada exame.

A seguir, foi aplicada a reamostragem isotrópica, na qual todas as imagens foram ajustadas para um espaçamento uniforme nos três eixos (X, Y e Z) [9]. O nnU-Net automaticamente determina a resolução ideal com base nas características do conjunto de dados e realiza o ajuste, garantindo que as imagens possuam uma resolução consistente, o que é crucial para que o modelo aprenda padrões espaciais de forma eficaz.

Além disso, a normalização das intensidades foi realizada para minimizar as variações de intensidade entre diferentes scanners e protocolos de aquisição [9]. Essa normalização assegura que as diferenças de intensidade entre imagens adquiridas com equipamentos distintos não interfiram no aprendizado da rede, permitindo que o modelo se concentre nas características anatômicas e patológicas.

Quando o conjunto de dados contém múltiplas modalidades de imagem (por exemplo, imagens de diferentes sequências de ressonância magnética), o alinhamento (registration) entre as modalidades foi automaticamente realizado [9], garantindo que todas as imagens e seus respectivos rótulos estivessem corretamente registrados no mesmo espaço físico. Esse alinhamento é essencial para garantir que as informações de diferentes modalidades correspondam entre si de maneira precisa, o que é um requisito para a segmentação multi-modal.

Para otimizar o processamento e garantir que o modelo receba entradas de tamanho adequado, o nnU-Net realiza corte (cropping) das imagens, removendo áreas não relevantes e ajustando as dimensões para um tamanho de entrada específico [2]. Essa etapa é importante para evitar sobrecarga computacional e também para focar nas regiões anatômicas

de interesse.

Por fim, durante o treinamento, a técnica de data augmentation foi aplicada de maneira online [2], realizando transformações como rotação, translação, deformações elásticas (elastic transforms) e variações de intensidade. Essas transformações aumentam a diversidade do conjunto de dados e ajudam a melhorar a robustez do modelo, permitindo que ele se generalize melhor para dados não vistos durante o treinamento.

Além dessas etapas de pré-processamento, foi realizada uma análise exploratória dos metadados dos exames [10], com foco em informações demográficas e clínicas dos pacientes. A análise estatística dos metadados incluiu a avaliação de variáveis como o fabricante do equipamento de imagem (Manufacturer), idade dos pacientes (Patient's Age), sexo dos pacientes (Patient's Sex), tipo de glioma (Glioma type) e local de aquisição. Essa análise permitiu entender a distribuição dos dados e identificar possíveis fontes de variabilidade associadas a fatores como o equipamento utilizado e a localização da aquisição, fornecendo informações valiosas para a interpretação dos resultados e a validação dos modelos treinados.

4.3 Modelo e Treinamento

Neste trabalho, a segmentação automática de gliomas foi investigada a partir de uma abordagem baseada em aprendizado profundo, adotando o nnU-Net 3D como modelo baseline [2], em função de sua capacidade de auto-configuração e do desempenho competitivo amplamente reportado na literatura, especialmente no contexto do desafio BraTS [5]. Como ponto de partida, o nnU-Net foi utilizado em sua configuração padrão, empregando as quatro modalidades de ressonância magnética como canais de entrada independentes. A partir desse baseline, foram propostas variações experimentais com o objetivo de analisar o impacto de diferentes estratégias de representação dos dados e de modificações arquiteturais no desempenho da segmentação.

Especificamente, foram conduzidos três experimentos: (i) o modelo baseline com nnU-Net 3D, proposto por Fabian Isensee em 2021 [2]; (ii) uma variação baseada na redução de dimensionalidade das modalidades por meio da Análise de Componentes Principais (PCA) [11], na qual as quatro modalidades originais foram projetadas em três componentes principais utilizados como canais de entrada; e (iii) uma abordagem com mecanismos de atenção (attention gates) incorporados à arquitetura, como proposto por Ozan Oktay em 2018 [12], visando melhorar a capacidade do modelo de focar em regiões relevantes do

tumor. Todos os experimentos seguiram um protocolo de treinamento consistente, incluindo divisão dos dados em conjuntos de treino, validação e teste, bem como amostragem que respeitou as distribuições das imagens nos diferentes grupos de metadados analisados — fabricante do equipamento de imagem (Manufacturer), idade dos pacientes (Patient's Age), sexo dos pacientes (Patient's Sex), tipo de glioma (Glioma type) e local de aquisição (site). O monitoramento das métricas de desempenho e o controle da reproduzibilidade garantiram uma comparação justa e sistemática entre as abordagens avaliadas.

Com base nos resultados obtidos nesses experimentos, foi selecionado o modelo com melhor desempenho global para a etapa final de treinamento. Essa etapa foi realizada utilizando os créditos gratuitos iniciais do Google Cloud Platform (GCP) [13], possibilitando o uso de recursos computacionais adequados para o treinamento do modelo escolhido. Os experimentos realizados e os critérios de seleção do modelo final são melhor detalhados nas seções subsequentes deste trabalho.

4.4 Inferência e Geração de Máscaras

Após o treinamento dos modelos, foi implementado um fluxo específico para a inferência e geração das máscaras de segmentação, com o objetivo de facilitar a aplicação dos modelos treinados a novos dados. Esse fluxo foi operacionalizado por meio de uma interface gráfica (GUI), desenvolvida para executar o processo de inferência de acordo com os parâmetros fornecidos pelo usuário, como a escolha do modelo e os dados de entrada. A partir dessas configurações, a GUI aciona automaticamente o pipeline de inferência [2], resultando na geração das máscaras de segmentação correspondentes às imagens analisadas.

As máscaras produzidas são salvas preservando a correspondência espacial com as imagens originais e, em seguida, a interface realiza a abertura conjunta da imagem e de sua respectiva segmentação no software ITK-SNAP [14]. Essa integração permite a visualização imediata e a análise qualitativa dos resultados, facilitando a inspeção das regiões segmentadas e a comparação entre diferentes modelos ou configurações experimentais. Os detalhes de implementação da interface e do processo de inferência são apresentados nas seções subsequentes deste trabalho.

4.5 Métricas de Avaliação e Análise Estatística

A avaliação do desempenho dos modelos de segmentação foi realizada por meio do coeficiente de Dice (Dice Similarity Coefficient – DSC), uma métrica amplamente utilizada em tarefas de segmentação de imagens médicas [15]. O coeficiente de Dice mede o grau de sobreposição entre a máscara predita pelo modelo e a segmentação de referência, sendo particularmente adequado para cenários em que há desbalanceamento entre as classes, como é comum na segmentação de tumores cerebrais.

Formalmente, o coeficiente de Dice é definido como:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

em que A representa o conjunto de voxels da segmentação predita e B o conjunto de voxels da segmentação de referência. Valores do coeficiente de Dice variam entre 0 e 1, sendo que valores mais próximos de 1 indicam maior similaridade entre as segmentações. Essa métrica tem sido amplamente adotada em desafios de segmentação médica, incluindo o BraTS, e é frequentemente utilizada como critério principal de comparação entre diferentes métodos, conforme reportado na literatura especializada.

4.6 Questões Éticas, Privacidade e Uso Responsável

O projeto não se destina a uso clínico direto sem aprovação institucional. Em caso de uso de dados reais, serão seguidos os procedimentos de consentimento e anonimização, com aprovação de comitê de ética quando necessário. O sistema é apresentado como ferramenta de apoio, mantendo o especialista no centro da decisão.

5 PLANO DE EXPERIMENTOS

A metodologia apresentada até este ponto teve como objetivo fornecer uma visão geral das etapas envolvidas no desenvolvimento do trabalho, abrangendo a descrição do conjunto de dados, o pipeline de pré-processamento, os modelos empregados, o processo de inferência e as métricas de avaliação adotadas. A partir dessa base metodológica, torna-se possível aprofundar a análise dos experimentos conduzidos, detalhando as configurações específicas, as variações propostas e os critérios utilizados para comparação entre os modelos.

Nesta seção, é apresentado o plano experimental, no qual são descritos de forma detalhada os experimentos realizados, incluindo a definição dos cenários avaliados, os parâmetros de treinamento, as estratégias de validação e os procedimentos adotados para a análise dos resultados. Essa abordagem permite compreender de maneira clara como as decisões experimentais foram tomadas e como os resultados obtidos sustentam as conclusões discutidas posteriormente neste trabalho.

Antes da apresentação detalhada dos resultados experimentais, é importante esclarecer a métrica denominada Pseudo Dice, amplamente utilizada nos gráficos de acompanhamento do treinamento. A métrica é utilizada pelo nnU-Net como uma estimativa suavizada do coeficiente Dice ao longo do treinamento. Diferentemente do Dice calculado diretamente a cada época, o *Pseudo Dice* é definido como uma média móvel exponencial (*Exponential Moving Average – EMA*) do Dice global de primeiro plano, com o objetivo de reduzir oscilações abruptas causadas por variações estocásticas inerentes ao processo de otimização.

Formalmente, o *Pseudo Dice* na época t , denotado por \tilde{D}_t , é calculado a partir do Dice observado D_t segundo a relação recursiva:

$$\tilde{D}_t = \alpha D_t + (1 - \alpha) \tilde{D}_{t-1}, \quad (5.1)$$

onde $\alpha \in (0, 1)$ é o fator de suavização que controla a influência relativa do valor corrente

em relação ao histórico acumulado. Valores menores de α produzem curvas mais suaves, porém com resposta mais lenta a mudanças rápidas no desempenho do modelo, enquanto valores maiores tornam o *Pseudo Dice* mais sensível às variações recentes.

O valor inicial \tilde{D}_0 é tipicamente definido como o Dice observado na primeira época ou inicializado a partir de um valor nulo, conforme a implementação. Essa métrica é amplamente empregada nos gráficos de treinamento do nnU-Net por fornecer uma visão mais estável da tendência de aprendizado do modelo, sendo particularmente útil para comparação entre experimentos em estágios iniciais ou sob restrições de número de épocas, como no caso dos experimentos conduzidos neste trabalho.

5.1 Análise Exploratória dos Dados e Extração de Subsets

Antes da condução dos experimentos de segmentação, foi realizada uma análise exploratória detalhada dos metadados do conjunto BraTS, com o objetivo de compreender a distribuição das amostras e subsidiar a definição de um plano experimental controlado. Considerando o grande volume de dados disponíveis e as limitações computacionais associadas ao treinamento de modelos tridimensionais, optou-se pela utilização de um subset representativo do conjunto completo para a fase inicial dos experimentos. Essa estratégia permitiu reduzir o custo computacional, acelerar o ciclo de experimentação e, ao mesmo tempo, preservar a diversidade estatística do conjunto original.

A análise exploratória foi conduzida a partir dos metadados disponibilizados pelo BraTS, incluindo informações sobre tipo de glioma, idade, sexo, local de aquisição (*site*), fabricante do equipamento (*Manufacturer*) e intensidade do campo magnético. Inicialmente, foram considerados apenas os pacientes pertencentes aos conjuntos de treinamento e treinamento adicional, excluindo dados de validação e teste, de modo a evitar qualquer forma de vazamento de informação (*data leakage*) durante os experimentos. Além disso, foi garantida a unicidade por paciente, removendo entradas duplicadas associadas ao mesmo identificador.

Para a extração do subset, foi implementado um procedimento de amostragem estratificada, no qual os pacientes foram selecionados de forma a preservar, tanto quanto possível, as distribuições observadas no conjunto completo. A estratificação principal foi realizada com base na combinação entre tipo de glioma e faixa etária, uma vez que esses fatores estão diretamente relacionados à heterogeneidade clínica e morfológica dos

tumores. As faixas etárias foram definidas automaticamente a partir da distribuição dos dados, permitindo agrupar pacientes em intervalos coerentes e balanceados. A partir dessa estratificação, o número de amostras selecionadas por estrato foi proporcional à sua representatividade no conjunto original, assegurando que o subset mantivesse características semelhantes às do dataset completo.

Além da estratificação principal, o processo de seleção buscou manter diversidade adicional em termos de site, sexo, fabricante do scanner e intensidade do campo magnético, fatores conhecidos por introduzir variabilidade significativa em imagens de ressonância magnética. Quando necessário, ajustes foram realizados para garantir exatamente o número desejado de pacientes no subset, evitando viés excessivo em favor de grupos mais numerosos. Todo o processo foi conduzido com controle de semente aleatória, assegurando a reproduzibilidade dos experimentos.

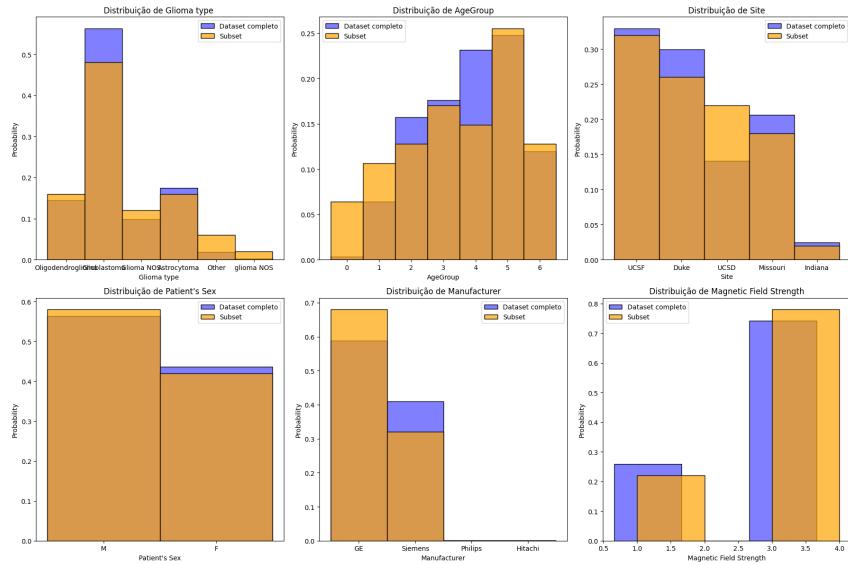


Figura 5: Histogramas gerados para análise de representatividade de grupos.

Como forma de validação da representatividade do subset extraído, foram gerados histogramas comparativos entre o conjunto completo e o subset, analisando visualmente as distribuições das principais variáveis de metadados. Essa comparação permitiu verificar que o subset preserva, de maneira consistente, as proporções relativas observadas no dataset original, reforçando sua adequação para os experimentos subsequentes. Por fim, também foi realizada uma inspeção das dimensões e do conteúdo das máscaras de segmentação associadas aos volumes selecionados, garantindo a integridade dos dados utilizados.

Em síntese, a análise exploratória e a extração controlada do subset tiveram como objetivo criar um cenário experimental reduzido a 50 imagens (cerca de 3% do dataset

total), porém estatisticamente representativo, possibilitando a condução de experimentos comparativos de forma eficiente e metodologicamente rigorosa. Essa etapa foi fundamental para assegurar que os resultados obtidos nos experimentos iniciais fossem confiáveis e pudessem ser posteriormente generalizados para o conjunto completo de dados.

5.2 Experimento 1: Baseline nnU-Net

O primeiro experimento teve como objetivo estabelecer um *baseline* sólido para a tarefa de segmentação automática de gliomas, utilizando o nnU-Net 3D em sua configuração padrão. Para isso, foi necessário realizar a construção manual da estrutura do *dataset* de acordo com as convenções estritas exigidas pela pipeline do nnU-Net, garantindo compatibilidade, reproduzibilidade e correta interpretação das imagens e rótulos durante o treinamento.

5.2.1 Preparação dos dados experimentais

Inicialmente, os dados selecionados no *subset* definido na Seção 5.1 foram organizados no formato esperado pelo nnU-Net, que requer a separação explícita entre imagens e rótulos, bem como a distinção entre conjuntos de treinamento e teste. As máscaras de segmentação foram renomeadas para remover sufixos adicionais presentes no nome original dos arquivos, assegurando que os identificadores dos pacientes fossem consistentes entre imagens e rótulos. Essa padronização de nomenclatura é fundamental para que o nnU-Net consiga associar corretamente cada volume de imagem à sua respectiva segmentação durante o processo de leitura dos dados.

Em seguida, foi realizada a divisão explícita entre dados de treino e teste, respeitando uma proporção fixa, com controle de semente aleatória para garantir reprodutibilidade. A separação foi feita no nível de paciente, de modo que todas as modalidades associadas a um mesmo indivíduo fossem mantidas no mesmo conjunto, evitando vazamento de informação entre treino e teste. Os arquivos correspondentes foram organizados nas pastas `imagesTr`, `labelsTr`, `imagesTs` e `labelsTs`, conforme a estrutura exigida pelo nnU-Net, permitindo que o framework identifique automaticamente os conjuntos de dados durante a execução da pipeline.

Cada caso de treinamento foi representado por múltiplos arquivos NIfTI, correspondentes às diferentes modalidades de ressonância magnética, organizados como canais independentes por meio de sufixos numéricos padronizados. Essa organização reflete diretamente a formulação multimodal do problema e permite que o nnU-Net trate cada sequência de imagem como uma fonte complementar de informação. As máscaras de segmentação associadas foram mantidas em arquivos separados, contendo os rótulos inteiros correspondentes às diferentes sub-regiões tumorais.

Para finalizar a construção do *dataset*, foi gerado o arquivo `dataset.json`, elemento central da pipeline do nnU-Net, responsável por descrever formalmente o conjunto de dados. Esse arquivo contém informações sobre as modalidades utilizadas, o mapeamento dos rótulos de segmentação, o número de casos de treinamento, o formato dos arquivos e os caminhos relativos para imagens e máscaras. A geração do arquivo JSON assegura que o nnU-Net interprete corretamente a semântica do problema, incluindo a correspondência entre classes e regiões anatômicas, além de permitir que o framework realize sua auto-configuração de forma adequada.

5.2.2 Preparação do ambiente de execução

O ambiente de execução dos experimentos foi baseado na plataforma Google Colab, selecionada por oferecer infraestrutura computacional acessível, integração nativa com serviços de nuvem e suporte à aceleração por hardware. Para o treinamento e a inferência dos modelos de segmentação, foi utilizada uma GPU NVIDIA Tesla T4, adequada para o processamento de volumes tridimensionais e compatível com as bibliotecas modernas de aprendizado profundo empregadas pelo nnU-Net. O uso de GPU foi essencial para reduzir o tempo de treinamento e viabilizar a execução dos experimentos dentro das restrições computacionais do projeto.

A configuração do ambiente incluiu a instalação das bibliotecas necessárias para a execução da pipeline do nnU-Net, abrangendo frameworks de aprendizado profundo, bibliotecas para manipulação de imagens médicas em formato NIfTI e ferramentas auxiliares para processamento e organização dos dados. Também foram definidas variáveis de ambiente e caminhos específicos exigidos pelo nnU-Net, garantindo o correto reconhecimento da estrutura de diretórios e dos arquivos de configuração do *dataset*. Os dados utilizados foram armazenados no Google Drive, empregado como serviço de nuvem para persistência e organização dos volumes de imagem, máscaras de segmentação e resultados experimentais, com acesso direto a partir do ambiente do Colab.

Apesar das vantagens oferecidas, o uso do Google Colab apresenta limitações importantes, que foram consideradas ao longo do desenvolvimento do trabalho. Entre elas, destacam-se a instabilidade ocasional na conexão com o Google Drive, que pode resultar em interrupções temporárias no acesso aos dados, bem como as restrições de tempo de execução das sessões, que impõem limites à duração contínua dos treinamentos. Além disso, a disponibilidade de recursos computacionais, como GPU, pode variar entre sessões, exigindo planejamento cuidadoso na execução dos experimentos e no salvamento frequente de estados intermediários. Essas limitações reforçam a importância do controle rigoroso dos experimentos e do armazenamento periódico dos resultados, garantindo a integridade e a reproduzibilidade do processo experimental.

Adicionalmente, foi utilizada a implementação oficial do nnU-Net, conforme proposta por Isensee et al. [15], disponibilizada como uma biblioteca de código aberto amplamente validada na literatura para tarefas de segmentação biomédica. Essa implementação fornece uma pipeline automatizada que abrange o pré-processamento dos dados, a configuração da arquitetura da rede, o treinamento e a inferência, reduzindo a necessidade de ajustes manuais e garantindo maior reproduzibilidade dos experimentos. Os códigos-fonte

empregados para a configuração do ambiente, execução dos treinamentos e realização das inferências estão disponíveis no repositório apresentado no Apêndice A, permitindo a reprodução integral dos experimentos descritos neste trabalho.

5.2.3 Execução e Resultados

A execução do Experimento 1 teve início com a etapa de planejamento e pré-processamento do *dataset* por meio do comando `nnUNetv2_plan_and_preprocess`, responsável por verificar a integridade dos dados, inferir automaticamente as configurações do modelo (como resolução, tamanho de *patch* e *batch size*) e preparar os arquivos intermediários necessários para o treinamento. A opção de verificação de integridade garantiu que a estrutura do *dataset*, os nomes dos arquivos e os rótulos estivessem em conformidade com os requisitos do nnU-Net, prevenindo inconsistências que poderiam comprometer o treinamento. Em seguida, o treinamento foi conduzido utilizando o comando `nnUNetv2_train`, adotando a configuração 3D *full resolution*, com um *trainer* customizado limitado a 50 épocas, escolha motivada pelas restrições computacionais do ambiente e pela natureza exploratória do experimento baseline. O salvamento de arquivos intermediários no formato `.npz` permitiu o monitoramento detalhado das métricas ao longo do treinamento.

Durante o processo de treinamento, observou-se uma convergência estável do modelo, conforme evidenciado pelas curvas de *loss* e *pseudo Dice* apresentadas na Figura 7. A perda de treinamento (*train loss*) apresentou queda monotônica ao longo das épocas, indicando aprendizado consistente da rede sobre os dados de treino. A perda de validação (*validation loss*), embora mais ruidosa — comportamento esperado em cenários com conjuntos reduzidos de dados — manteve tendência geral de diminuição, sugerindo ausência de *overfitting* severo. O *pseudo Dice*, métrica interna utilizada pelo nnU-Net para monitoramento rápido do desempenho, apresentou crescimento progressivo, tanto em sua forma instantânea quanto em sua média móvel exponencial (EMA), refletindo melhorias graduais na qualidade da segmentação ao longo do treinamento.

Ao final da época 50, o modelo atingiu um novo melhor valor de *pseudo Dice* médio (EMA - Exponencial Moving Average), com valor aproximado de 0,69, conforme registrado nos *logs* de treinamento. A análise por classe revelou desempenho mais elevado nas regiões tumorais mais extensas, enquanto regiões menores ou mais heterogêneas apresentaram maior variabilidade, comportamento consistente com o esperado para tarefas de segmentação de gliomas. O tempo médio por época manteve-se estável em torno de 176 segundos, indicando utilização consistente dos recursos computacionais disponíveis.

Esses resultados confirmam que o modelo *baseline* nnU-Net 3D foi capaz de aprender representações relevantes mesmo a partir de um subconjunto reduzido de dados, estabelecendo uma referência sólida para os experimentos subsequentes que exploram estratégias alternativas de representação dos dados e modificações arquiteturais.

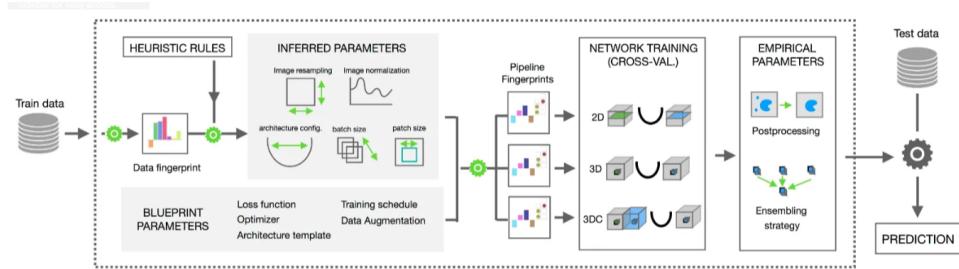


Figura 6: Pipeline em blocos da arquitetura nnUNet, evidenciando os caminhos de codificação e decodificação, bem como as skip connections responsáveis pela preservação das informações espaciais ao longo do processo de segmentação.

Fonte: Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 2020.

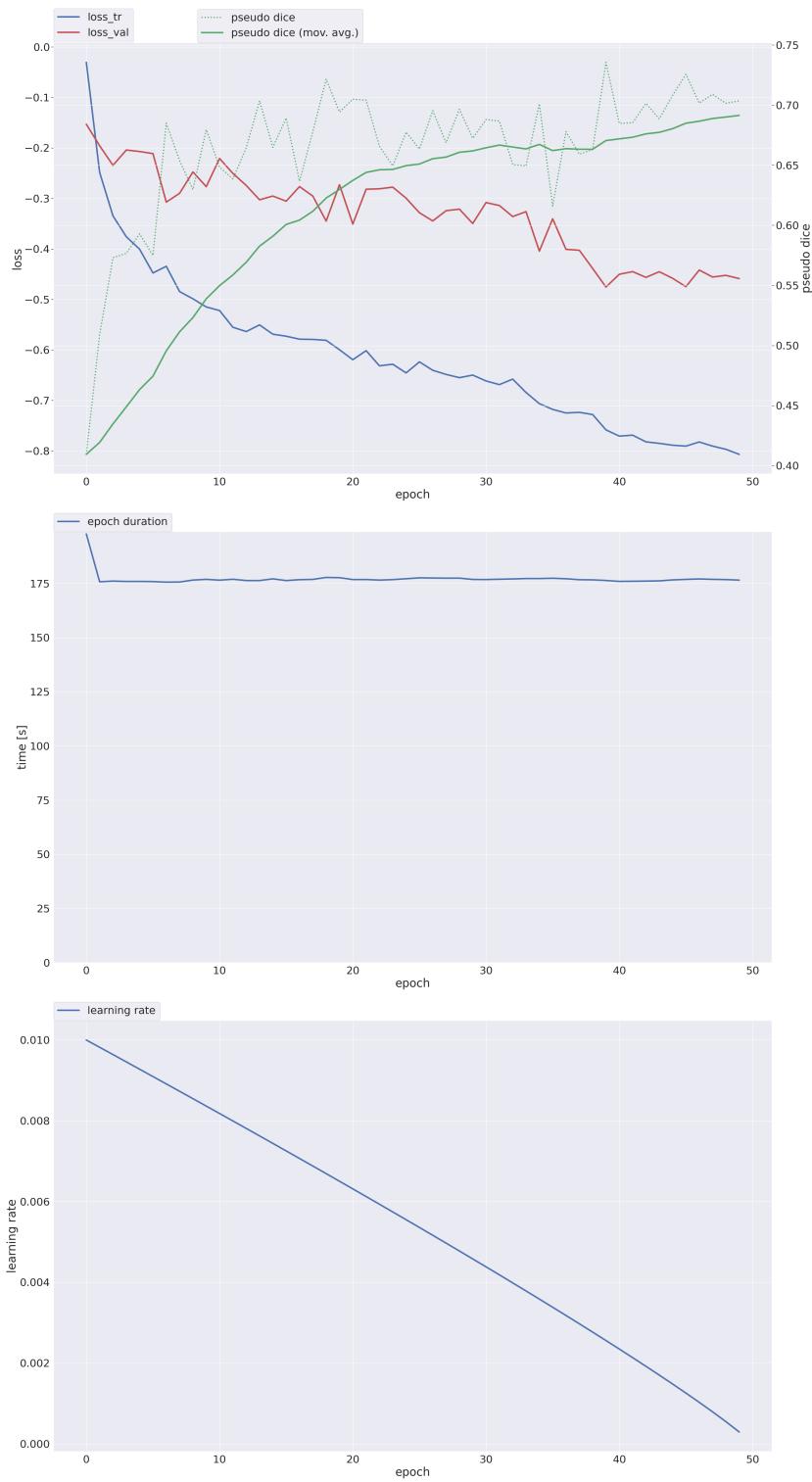


Figura 7: Resultados do treinamento e teste baseline.

5.3 Experimento 2: PCA dos canais + nnU-Net

O segundo experimento investigou a aplicação de redução de dimensionalidade via Análise de Componentes Principais (PCA) sobre as modalidades multimodais de ressonância magnética, com o objetivo de produzir uma representação compacta (três canais) que servisse de entrada ao nnU-Net 3D. A motivação técnica é dupla: (i) reduzir redundâncias entre modalidades altamente correlacionadas (T1, T1ce, T2, FLAIR) e (ii) diminuir custo computacional e potencial risco de *overfitting* em cenários com conjuntos de treino limitados. Na literatura, abordagens que aplicam PCA como etapa de pré-processamento antes de arquiteturas convolucionais têm mostrado capacidade de preservar a variância informativa enquanto simplificam a entrada para modelos de segmentação [16, 17], sendo reportadas em estudos de PCA-aided FCN e em trabalhos específicos de PCA-U-Net e compressão de canais em MRI [18].

5.3.1 Implementação do PCA

A transformação aplicada foi implementada em quatro etapas principais, visando reduzir a dimensionalidade das modalidades de ressonância magnética e criar uma representação mais compacta e informativa dos dados. O objetivo dessa abordagem foi melhorar o desempenho do modelo e reduzir o custo computacional, mantendo a qualidade da segmentação. Toda a implementação foi realizada por meio de scripts próprios, desenvolvidos em Python, utilizando bibliotecas consolidadas para aprendizado de máquina e processamento de imagens médicas. Os códigos responsáveis por cada etapa do processo, incluindo amostragem, normalização, ajuste do PCA, aplicação da transformação e geração dos novos volumes NIfTI, estão disponibilizados no repositório apresentado no Apêndice A, permitindo a completa reproduzibilidade da metodologia proposta.

A primeira etapa consistiu na amostragem de voxels representativos para o ajuste do PCA. Para reduzir o custo computacional e evitar viés introduzido por volumes maiores, foi selecionada uma fração de 1% dos voxels válidos por paciente, considerando apenas os voxels com intensidade maior que zero em pelo menos uma modalidade. Esses voxels foram concatenados em uma matriz ($N_{\text{vox}} \times 4$ canais) para a estimativa global dos componentes principais, garantindo que as direções de maior variância refletissem a variabilidade do conjunto de treino e representassem adequadamente as características dos pacientes e das diferentes modalidades. A amostragem aleatória foi conduzida com semente fixa, assegurando a reproduzibilidade dos resultados.

Na segunda etapa, realizou-se a normalização pré-PCA dos voxels amostrados, utilizando o método de *z-score* por canal, com subtração da média e divisão pelo desvio padrão de cada modalidade. Essa normalização é fundamental para garantir que o PCA capture as relações de covariância entre os canais em escalas comparáveis, evitando que modalidades com maior variância absoluta dominem indevidamente as componentes principais.

A terceira etapa compreendeu o ajuste global do PCA e o armazenamento do modelo. O PCA foi ajustado utilizando todas as amostras concatenadas, com $n_{\text{components}} = 3$, resultando em três componentes principais responsáveis por reter a maior parte da variância dos dados. Para viabilizar o processamento de grandes matrizes, empregou-se o solver *randomized*. O modelo treinado foi serializado com a biblioteca `joblib`, permitindo sua reutilização consistente durante as fases de treino e teste, sem a necessidade de novo ajuste.

Por fim, na última etapa, o PCA treinado foi aplicado a cada volume individualmente para a geração dos novos canais NIfTI. As quatro modalidades de ressonância magnética foram carregadas e empilhadas em uma matriz 4D ($H, W, D, 4$), posteriormente reorganizada em uma matriz 2D ($N_{\text{vox_total}} \times 4$). Após a normalização conforme os parâmetros previamente estimados, os dados foram transformados pelo PCA global. Os três componentes principais obtidos foram então reconvertisdos para volumes 3D e salvos como canais independentes (`case_0000.nii.gz`, `case_0001.nii.gz`, `case_0002.nii.gz`), preservando a matriz afim original para manter a coerência espacial. As máscaras de segmentação originais foram copiadas para a nova estrutura do *dataset*, assegurando a correspondência voxel-a-voxel entre imagens e rótulos.

5.3.2 Execução e Resultados

A execução do segundo experimento seguiu o mesmo protocolo operacional adotado no experimento baseline, com a diferença de que o treinamento foi conduzido sobre o *dataset* transformado por PCA, no qual as quatro modalidades originais foram substituídas por três componentes principais. O processo de planejamento, pré-processamento e treinamento foi realizado utilizando o comando `nnUNetv2_train`, mantendo a configuração 3D *full resolution*, o mesmo fold de validação e o mesmo *trainer* customizado definido para 50 épocas, garantindo comparabilidade direta entre os experimentos.

Devido às limitações operacionais do ambiente Google Colab — incluindo desconexões inesperadas e restrições de tempo de execução — o treinamento deste experimento foi in-

terrompido após 10 épocas, não atingindo o número máximo originalmente planejado. Ainda assim, os logs de treinamento indicam aprendizado consistente nas etapas iniciais, com redução progressiva da função de perda tanto no conjunto de treino quanto no conjunto de validação. Na época 10, observou-se uma *train loss* de aproximadamente -0,43 e uma *validation loss* de -0,24, valores que sugerem que o modelo iniciou adequadamente o processo de adaptação à nova representação dos dados. O *pseudo Dice* médio apresentou melhora contínua ao longo das épocas, atingindo um valor de EMA *pseudo Dice* de aproximadamente 0,44.

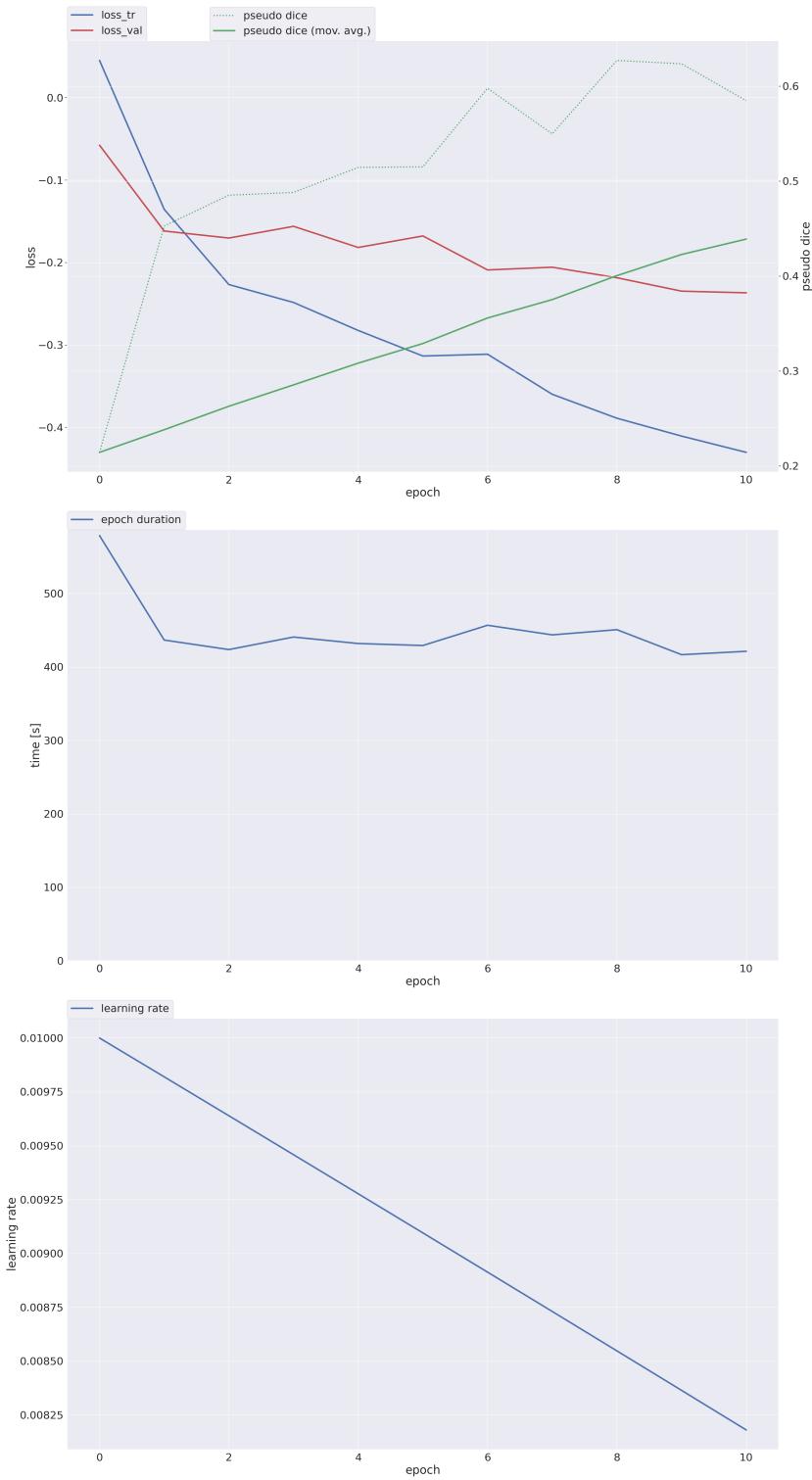


Figura 8: Resultados do treinamento e teste com PCA.

A análise por classe do *pseudo Dice* revela que as regiões tumorais mais extensas apresentaram desempenho relativamente elevado mesmo nas fases iniciais do treinamento, enquanto regiões menores e mais heterogêneas exibiram maior variabilidade, comportamento consistente com o observado no experimento baseline. Embora os resultados quantitativos deste experimento não possam ser comparados de forma direta ao baseline em regime de convergência total, os sinais iniciais de aprendizado indicam que a representação baseada em PCA preserva informações discriminativas relevantes para a tarefa de segmentação, demonstrando a viabilidade da compressão de canais por PCA no contexto do nnU-Net.

5.4 Experimento 3: Attention Gates + nnU-Net

No Experimento 3, investiga-se a incorporação de mecanismos de *attention gates* à arquitetura do nnU-Net, com o objetivo de avaliar o impacto da atenção espacial e de canais no desempenho da segmentação. Diferentemente das variações arquiteturais baseadas apenas em conectividade ou multiescala, os mecanismos de atenção buscam modular explicitamente as ativações internas da rede, enfatizando regiões e características mais relevantes para a tarefa de segmentação e atenuando respostas associadas a estruturas irrelevantes ou de fundo. Trabalhos recentes demonstram que *attention gates* são particularmente eficazes em cenários de segmentação médica caracterizados por forte desbalanceamento espacial entre classes, como a segmentação de tumores, nos quais as regiões de interesse ocupam uma pequena fração do volume total [19]. Ao integrar diferentes formas de atenção ao nnU-Net, mantendo inalterados os demais componentes do pipeline de treinamento, este experimento permite analisar de forma controlada se a capacidade do modelo em focar seletivamente em regiões semanticamente relevantes resulta em ganhos consistentes de desempenho em relação à arquitetura padrão, especialmente em estruturas minoritárias e de maior complexidade morfológica.

5.4.1 Implementação

A implementação do Experimento 3 baseou-se na incorporação explícita de mecanismos de atenção espacial (*attention gates*) à arquitetura do nnU-Net, segundo a proposta descrita na literatura e utilizando uma extensão do *framework* original disponibilizada publicamente. Para esse fim, foi empregado o repositório *Advanced nnU-Net*, desenvolvido por McConnell N., Ndipenoch N., Cao Y., Miron A. e Li Y., conforme descrito em Exploring advanced architectural variations of nnUNet (Neurocomputing, 2023) [19], o qual implementa variações arquiteturais do nnU-Net com blocos de atenção integrados às

conexões de *skip* da U-Net. Essa implementação permite controlar explicitamente quais regiões espaciais e quais características intermediárias são propagadas do *encoder* para o *decoder*, viabilizando a avaliação do impacto da atenção espacial na segmentação.

Inicialmente, o repositório foi clonado no ambiente de execução e integrado à estrutura do nnU-Net por meio da configuração das variáveis de ambiente exigidas pelo *framework*, incluindo os diretórios de dados brutos (`nnUNet_raw_data_base`), dados pré-processados (`nnUNet_preprocessed`) e resultados (`RESULTS_FOLDER`). Essa etapa assegurou compatibilidade com a organização de dados utilizada nos experimentos anteriores, permitindo a reutilização do mesmo subconjunto de 50 pacientes do BraTS e garantindo a comparabilidade entre os resultados obtidos.

Em seguida, foi realizada a criação manual do arquivo `dataset.json`, em conformidade com a convenção esperada pela versão do nnU-Net empregada no repositório *Advanced nnU-Net*, baseada no nnU-Net v1. Esse arquivo descreve formalmente o conjunto de dados, especificando o número de casos de treino e teste, as modalidades de entrada (T1, T1ce, T2 e FLAIR), os rótulos das classes tumorais e os caminhos relativos para as imagens e máscaras de segmentação. Essa etapa foi necessária devido às diferenças estruturais entre a pipeline do nnU-Net v2, utilizada nos experimentos anteriores, e a implementação de atenção adotada neste experimento.

Após a definição do *dataset*, foi executada a etapa de planejamento e pré-processamento por meio de um *planner* específico para arquiteturas com atenção espacial, denominado `ExperimentPlanner3DSpatialSingleAttentionUNetv21`. Esse componente é responsável por definir automaticamente parâmetros fundamentais do treinamento, como resolução alvo, tamanho dos *patches*, *batch size* e estratégias de normalização, levando explicitamente em consideração a presença dos blocos de atenção na arquitetura da rede. Nessa etapa, também foi realizada a verificação da integridade do *dataset*, garantindo a consistência entre volumes de imagem e rótulos.

Por fim, o treinamento foi conduzido utilizando o *trainer* `nnUNetTrainerV2SpatialSingleAttentionUNet`, na configuração 3D *full resolution*, mantendo o mesmo *fold* de validação adotado nos demais experimentos. A principal diferença em relação ao nnU-Net padrão reside na inclusão de *attention gates* nas conexões de *skip*, os quais modulam dinamicamente a passagem de informações das camadas do *encoder* para o *decoder*. Esse mecanismo permite que o modelo enfatize regiões espacialmente relevantes, como áreas tumorais, e atenuar contribuições provenientes de regiões de fundo ou estruturas irrelevantes, sem alterar o restante da pipeline de treinamento. Todos os scripts utilizados para a

configuração do ambiente, definição do *dataset*, planejamento, pré-processamento, treinamento e inferência encontram-se disponibilizados no repositório apresentado no Apêndice A, assegurando a reproduzibilidade completa do experimento.

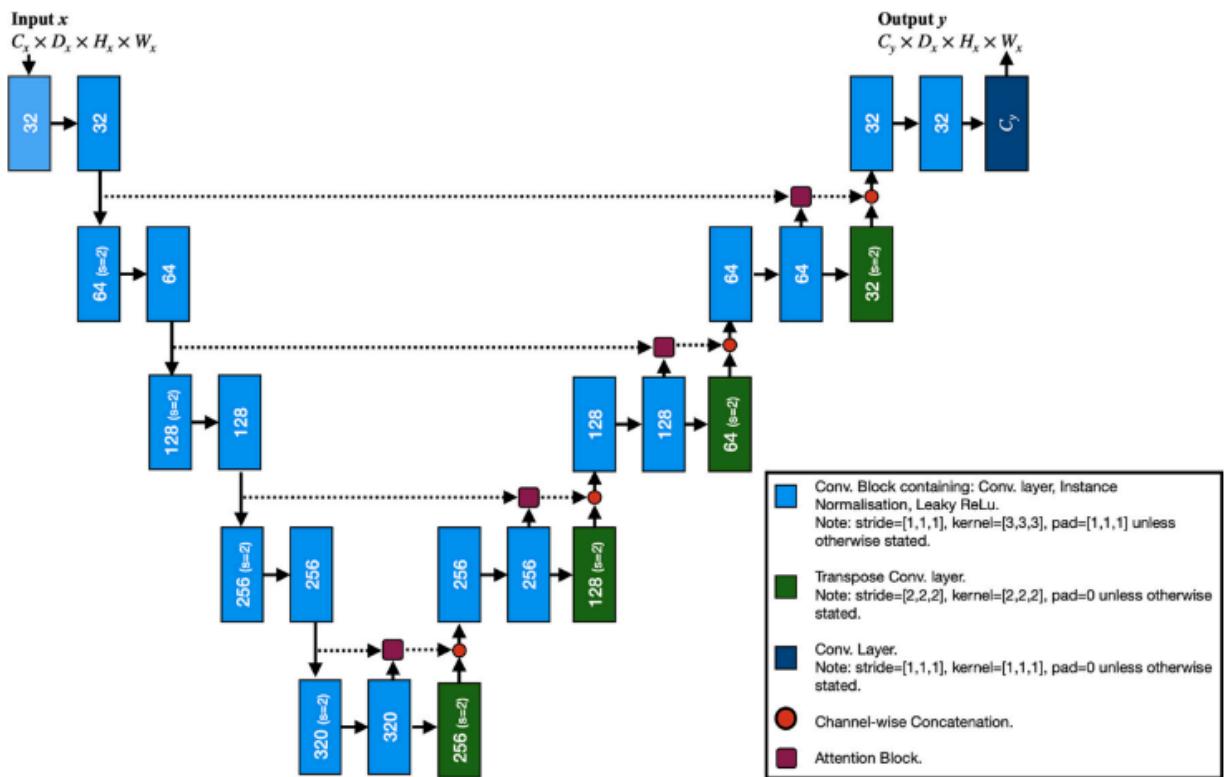


Figura 9: Exemplo da inserção de attention blocks (representados em roxo) nas skip connections do modelo nnUNet, visando melhorar a capacidade de segmentação ao focar em regiões relevantes da imagem durante o processo de codificação e decodificação. Fonte: McConnell N, Ndipenoch N, Cao Y, Miron A, Li Y. Exploring advanced architectural variations of nnUNet. Neurocomputing, 2023.

5.4.2 Resultados

Os resultados do Experimento 3 refletem o comportamento inicial do nnU-Net com *attention gates* espaciais incorporados à arquitetura, treinado sobre o mesmo subset de dados utilizado nos experimentos anteriores. Devido às limitações do ambiente de execução em nuvem, o treinamento foi interrompido precocemente após 6 épocas, o que impede uma análise em regime de convergência completa. Ainda assim, os logs e as curvas de treinamento permitem observar tendências relevantes sobre o impacto da atenção espacial no processo de aprendizado.

Conforme ilustrado na Figura correspondente, a função de perda de treinamento apresentou queda consistente ao longo das épocas, atingindo um valor aproximado de $-0,39$ na época 6, indicando que o modelo foi capaz de aprender representações discriminativas a partir dos dados de treino. A perda de validação, embora mais estável e com redução menos acentuada, manteve-se em torno de $-0,19$, sugerindo ausência de divergência severa entre treino e validação nas fases iniciais do treinamento. Esse comportamento é compatível com arquiteturas que introduzem mecanismos adicionais de controle, como os *attention gates*, os quais tendem a tornar o processo de otimização mais estável, porém potencialmente mais lento nas primeiras épocas.

A métrica de avaliação reportada — *Average Global Foreground Dice* — apresentou valores distintos entre as diferentes classes tumorais, mesmo em um número reduzido de épocas. Observou-se desempenho mais elevado para regiões tumorais maiores e mais bem definidas, enquanto estruturas menores ou mais complexas apresentaram valores de Dice inferiores. O vetor de Dice médio por classe ao final da época 6 foi aproximadamente $[0, 29, 0, 81, 0, 76, 0, 53]$, indicando que o modelo já demonstrava capacidade de segmentar corretamente as regiões mais predominantes do tumor, mesmo em estágio inicial de treinamento.

Um aspecto relevante observado neste experimento foi o alto tempo por época, da ordem de 1157 segundos, significativamente superior ao registrado no experimento baseline. Esse aumento é atribuído à maior complexidade arquitetural introduzida pelos *attention gates*, que adicionam operações adicionais de convolução e multiplicação elemento a elemento nas conexões de *skip*. Tal custo computacional reforça o caráter exploratório deste experimento, evidenciando um trade-off entre complexidade do modelo e viabilidade computacional em ambientes com recursos limitados.

De forma geral, embora os resultados obtidos não permitam uma comparação quantitativa direta com os experimentos anteriores em regime de convergência, os sinais iniciais

indicam que a incorporação de mecanismos de atenção espacial é viável e pode contribuir para uma segmentação mais focada em regiões relevantes do tumor. Esses achados estão alinhados com a literatura, que aponta ganhos mais expressivos com *attention gates* após treinamentos mais longos e em cenários com maior disponibilidade computacional. Assim, este experimento cumpre seu papel ao demonstrar empiricamente o impacto arquitetural da atenção no nnU-Net, servindo como base para discussões mais aprofundadas no capítulo de análise comparativa dos resultados.

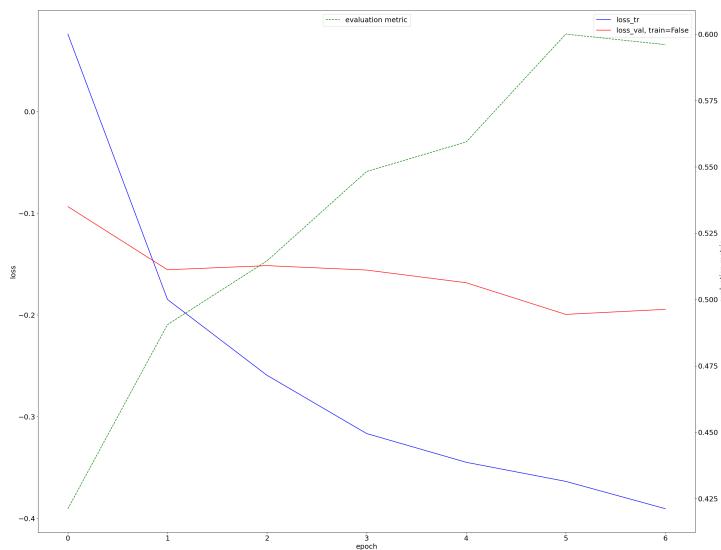


Figura 10: Resultados do treinamento e teste com Attention Gates.

6 ESCOLHA DO MODELO E TREINAMENTO FINAL

Após a condução dos experimentos exploratórios apresentados nas seções anteriores, tornou-se necessário consolidar os resultados obtidos e definir o modelo a ser adotado para o treinamento final. A escolha do modelo levou em consideração não apenas o desempenho quantitativo observado nas métricas de avaliação, mas também aspectos práticos relacionados à estabilidade do treinamento, custo computacional e viabilidade de execução no ambiente disponível. Dessa forma, os experimentos com o nnU-Net baseline e a variação com redução de dimensionalidade via PCA forneceram subsídios diretos para uma análise comparativa criteriosa, permitindo avaliar o impacto das modificações arquiteturais e de pré-processamento sob condições controladas.

A Tabela 1 apresenta a comparação entre o nnU-Net baseline e o Experimento 2 até a época 14, último ponto comum de treinamento entre ambos. São reportados os valores de *Pseudo Dice* por classe tumoral, bem como o tempo médio por época, possibilitando uma análise conjunta de desempenho e custo computacional.

Observa-se que o Experimento 2 apresentou ganho expressivo de Dice na Classe 1, associada a regiões tumorais menores ou mais difíceis de segmentar, enquanto manteve desempenho semelhante ao baseline na Classe 2. Por outro lado, houve redução nas Classes 3 e 4, além de um aumento significativo no tempo de treinamento por época, evidenciando um trade-off entre desempenho em regiões específicas e custo computacional.

Complementarmente, a Tabela 2 apresenta a comparação entre o nnU-Net baseline e o Experimento 3, que incorpora mecanismos de atenção espacial, considerando os resultados obtidos até a época 6, último ponto comum de treinamento entre os dois modelos. São apresentados os valores de *Dice médio global de primeiro plano* por classe tumoral, juntamente com o tempo por época, permitindo avaliar o impacto inicial da inclusão dos *attention gates* na arquitetura.

Nota-se que, nas primeiras épocas de treinamento, o Experimento 3 apresentou desempenho comparável ao baseline nas Classes 1 e 2, leve redução na Classe 3 e ganho

relevante na Classe 4, associada a regiões tumorais mais complexas ou menos representadas. Entretanto, esse comportamento inicial veio acompanhado de um aumento substancial no tempo por época, refletindo a maior complexidade computacional introduzida pelos mecanismos de atenção espacial. Esses resultados reforçam o caráter exploratório do Experimento 3 e indicam que eventuais ganhos proporcionados pelos *attention gates* tendem a se manifestar de forma mais consistente em regimes de treinamento mais longos, conforme reportado na literatura.

Em conjunto, as análises quantitativas apresentadas nas Tabelas 1 e 2, aliadas às restrições computacionais do ambiente de execução, fundamentaram a escolha do modelo adotado para o treinamento final, conforme descrito na seção seguinte.

6.1 Escolha do Modelo

A escolha do modelo adotado para o treinamento final foi fortemente condicionada pelas limitações impostas pelo uso do Google Cloud Platform (GCP) em regime de *free trial*, as quais impactaram diretamente tanto os custos operacionais quanto a viabilidade experimental. Embora o ambiente do Colab Enterprise disponibilize infraestrutura de alto desempenho, o período de avaliação apresentou restrições relevantes de configuração. Em particular, não foi possível alocar o bucket de armazenamento (Google Drive/Cloud Storage) na mesma região da máquina virtual, uma vez que a região disponível para armazenamento diferia daquela associada à instância de computação liberada no *trial*. Essa limitação obrigou a utilização de armazenamento em região distinta da VM, resultando em custos adicionais de transferência de dados entre regiões (*egress*) e aumento da latência de acesso aos volumes NIfTI, fator crítico em pipelines de segmentação 3D intensivos em I/O, como o nnU-Net.

Além disso, outra limitação significativa do *free trial* foi a ausência de flexibilidade na escolha do tipo de GPU. Durante o período de execução dos experimentos, a única GPU disponível foi a NVIDIA A100, uma placa de altíssimo desempenho e, consequentemente, de custo elevado por hora de uso. Embora a A100 seja tecnicamente adequada — e até superdimensionada — para o treinamento de modelos como o nnU-Net 3D, sua utilização impõe um custo financeiro desproporcional quando comparada a alternativas mais econômicas (por exemplo, GPUs da linha T4 ou V100), que seriam suficientes para este tipo de experimento. Essa restrição eliminou a possibilidade de otimização de custos por meio da seleção de hardware mais adequado ao problema, tornando cada minuto de treinamento significativamente mais oneroso.

Nesse contexto, a avaliação dos modelos candidatos precisou considerar não apenas métricas de desempenho, mas também o tempo de execução por época, uma vez que este se traduz diretamente em custo financeiro no ambiente de nuvem utilizado. A comparação dos resultados até a sexta época de treinamento, ponto comum entre os experimentos antes de interrupções por limite de sessão, revelou diferenças importantes. O modelo baseline nnU-Net 3D apresentou um EMA pseudo Dice médio de aproximadamente 0,50, com tempo médio por época de cerca de 175 s, configurando a alternativa mais eficiente do ponto de vista custo-benefício. Em contraste, a abordagem baseada em PCA dos canais apresentou menor desempenho inicial ($\text{EMA} \approx 0,36$) e um tempo por época substancialmente maior (≈ 457 s), o que ampliaria significativamente os custos caso o treinamento fosse prolongado. De forma similar, o modelo com *attention gates* implicou maior complexidade computacional e tempos de treinamento elevados ($\approx 1157,4$ s), sem apresentar ganhos claros de desempenho nas fases iniciais.

Diante dessas restrições estruturais do GCP no *free trial* — armazenamento forçado em região distinta da VM, custos adicionais de transferência de dados e disponibilidade exclusiva de uma GPU NVIDIA A100 de alto custo — a decisão final foi optar pelo treinamento do modelo baseline. Essa escolha permitiu reduzir o tempo total de execução, minimizar gastos computacionais e garantir a conclusão do treinamento dentro das limitações do ambiente, sem comprometer a validade metodológica do estudo. Assim, a seleção do modelo final reflete uma decisão técnica fundamentada não apenas no desempenho de segmentação, mas também na viabilidade econômica e operacional do experimento em um cenário realista de recursos limitados.

6.2 Treinamento Final

O treinamento final foi conduzido utilizando uma versão ampliada e mais representativa do conjunto de dados BraTS, considerando todo o conjunto de imagens anotadas disponibilizado para o desafio. No total, a base continha 1.621 volumes de ressonância magnética com segmentações manuais, correspondentes a 731 pacientes distintos, sendo que alguns pacientes possuíam múltiplas aquisições ao longo do tempo (*visitas longitudinais*). Para reduzir a redundância temporal e evitar viés decorrente da presença de múltiplos exames do mesmo paciente, optou-se por selecionar apenas uma imagem por paciente, totalizando 731 volumes segmentados. Especificamente, foi considerada a primeira visita registrada de cada paciente, estratégia que garante independência estatística entre amostras e preserva a diversidade inter paciente do conjunto de dados.

A preparação do *dataset* seguiu o mesmo protocolo adotado no Experimento 1 (baseline nnU-Net 3D), mantendo a padronização das modalidades de entrada, a estrutura de diretórios exigida pelo nnU-Net e o processo automático de planejamento e pré-processamento da *pipeline*. Essa escolha visou assegurar que os ganhos observados no treinamento final fossem atribuídos principalmente ao aumento do número de amostras e não a alterações no fluxo metodológico. O treinamento foi realizado integralmente no Google Colab Enterprise, utilizando infraestrutura de nuvem dedicada, incluindo GPU NVIDIA A100.

Do ponto de vista de implementação, o treinamento final seguiu essencialmente o mesmo fluxo metodológico adotado no Experimento 1 (baseline nnU-Net 3D), utilizando a implementação oficial do nnU-Net proposta por Isensee et al., sem modificações arquiteturais adicionais. A principal diferença em relação aos experimentos iniciais esteve na origem e no volume dos dados utilizados. Enquanto os experimentos exploratórios foram conduzidos a partir de subconjuntos reduzidos armazenados localmente no Google Drive, o treinamento final foi realizado diretamente a partir de um *bucket* de armazenamento no Google Cloud Platform (GCP), configurado para acesso contínuo pelo ambiente do Google Colab Enterprise.

Os volumes de imagem e as máscaras de segmentação foram organizados no *bucket* seguindo rigorosamente a estrutura de diretórios exigida pelo nnU-Net, permitindo que as etapas automáticas de verificação do *dataset*, planejamento e pré-processamento fossem executadas sem adaptações adicionais na *pipeline*. O acesso aos dados em nuvem foi integrado ao ambiente de execução por meio do sistema de arquivos montado no Colab, garantindo transparência para os scripts do nnU-Net e viabilizando o processamento de grandes volumes de dados sem a necessidade de cópia integral para armazenamento local.

Dessa forma, a implementação do treinamento final manteve consistência com o baseline em termos de configuração, parâmetros e fluxo de execução, assegurando que os ganhos observados fossem atribuídos principalmente ao aumento do conjunto de dados e à maior representatividade estatística, e não a alterações no código ou na arquitetura do modelo. Todos os scripts utilizados para a preparação do ambiente, configuração do acesso ao *bucket*, execução do pré-processamento e treinamento do modelo encontram-se disponíveis no repositório descrito no Apêndice A, garantindo a reproduzibilidade completa do experimento.

O processo de pré-processamento demandou um custo computacional significativo, totalizando mais de 24 horas de execução, devido à natureza volumétrica dos dados, às

operações de reamostragem, normalização, cálculo de estatísticas globais e geração dos planos de treinamento automáticos do nnU-Net. Em seguida, o treinamento do modelo foi executado por 200 épocas, com duração aproximada de 10 horas, refletindo tanto o aumento substancial do conjunto de treino quanto a complexidade do modelo 3D em resolução completa.

Os resultados obtidos ao final do treinamento evidenciam uma melhora expressiva no desempenho do modelo em comparação aos experimentos conduzidos com *subsets* reduzidos. No melhor ponto observado, o modelo atingiu um EMA pseudo Dice médio de aproximadamente 0,87, com valores elevados em todas as regiões tumorais avaliadas. Em uma das épocas finais, por exemplo, foram registrados Pseudo Dice de 0,7953, 0,9224, 0,9142 e 0,8556 para as diferentes classes, indicando segmentações consistentes tanto para regiões mais extensas quanto para subestruturas mais complexas do tumor. O tempo médio por época permaneceu estável, em torno de 157 s, demonstrando boa eficiência computacional mesmo com o aumento do volume de dados.

Esses resultados reforçam a adequação do modelo baseline nnU-Net 3D quando treinado em um conjunto de dados mais abrangente, confirmando que a limitação de desempenho observada nos experimentos iniciais estava fortemente associada ao tamanho reduzido dos *subsets*. Apesar do elevado custo computacional — tanto em tempo quanto em recursos de nuvem —, o treinamento final mostrou-se bem-sucedido, produzindo um modelo robusto e alinhado com o desempenho esperado para tarefas de segmentação automática de gliomas em cenários realistas de grande escala.

Tabela 1: Comparação entre o nnU-Net baseline e o Experimento 2 (PCA) na época 14

Classe tumoral	Baseline	Experimento 2 (PCA)
Classe 1	0.2866	0.4618
Classe 2	0.8172	0.8272
Classe 3	0.7372	0.7175
Classe 4	0.8186	0.5749
Tempo por época (s)	177.1	411.6

Tabela 2: Comparação entre o nnU-Net baseline e o Experimento 3 (Attention Gates) na época 6

Classe tumoral	Baseline	Experimento 3 (Attention)
Classe 1	0.3095	0.2906
Classe 2	0.8137	0.8112
Classe 3	0.7891	0.7558
Classe 4	0.3876	0.5268
Tempo por época (s)	175.8	1157.4

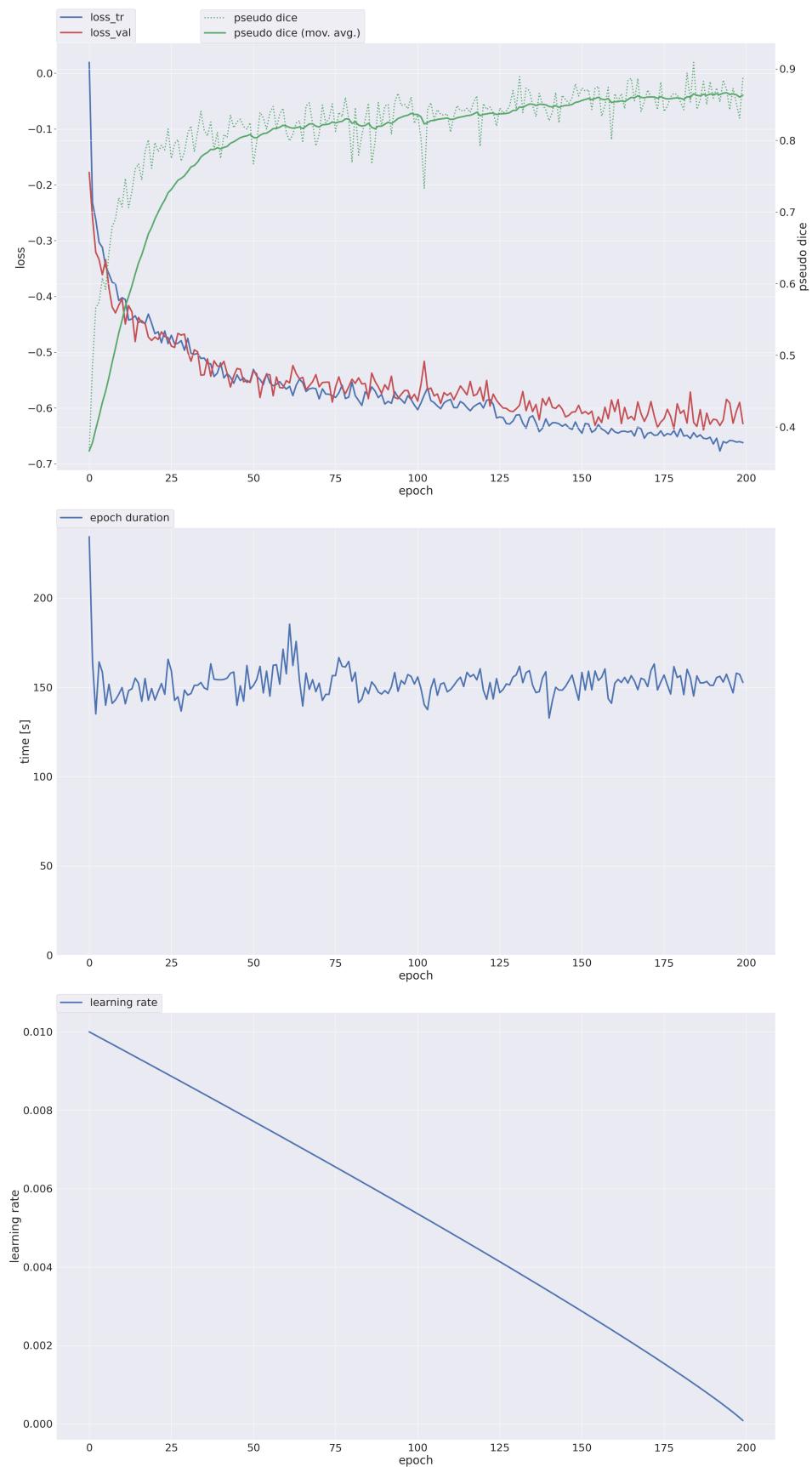


Figura 11: Curvas de Dice e Loss do treinamento final.

6.3 Custos e Limitações

A execução dos experimentos esteve fortemente condicionada às limitações orçamentárias do ambiente de nuvem utilizado. Inicialmente, foi disponibilizado um crédito de US\$ 300 por meio do programa de avaliação do Google Cloud, o qual se mostrou insuficiente para cobrir integralmente os custos associados ao treinamento completo do modelo em larga escala. Os gastos com processamento no Google Colab Enterprise, incluindo o uso de GPU NVIDIA A100 durante as etapas de pré-processamento e treinamento, totalizaram aproximadamente R\$ 800, refletindo o elevado custo de instâncias com aceleração por hardware. Adicionalmente, os custos de armazenamento em nuvem alcançaram cerca de R\$ 1000, valor significativamente superior ao inicialmente estimado, em grande parte devido à necessidade de manter o *bucket* de armazenamento em uma região distinta da máquina virtual de processamento, situação imposta pelas restrições do período de teste gratuito e que resultou em cobranças adicionais por transferência inter-regional de dados.

Em função desse cenário, optou-se por treinar apenas um *fold* de validação cruzada, em vez do esquema completo tradicionalmente empregado pelo nnU-Net, como medida de contenção de custos computacionais. Embora essa decisão limite a análise da variabilidade do desempenho entre *folds*, ela permitiu viabilizar o treinamento final do modelo dentro das restrições financeiras disponíveis, mantendo ainda assim a coerência metodológica e a validade dos resultados obtidos.

7 INTEGRAÇÃO COM O ITK-SNAP

Com o objetivo de facilitar a aplicação prática do modelo treinado e permitir a inspeção qualitativa das segmentações geradas, foi desenvolvido um fluxo integrado que conecta o nnU-Net v2 ao software de visualização médica ITK-SNAP. Essa integração foi projetada para tornar o processo de inferência mais acessível, reproduzível e independente de execução manual via linha de comando, aproximando o sistema de um cenário de uso real em ambientes acadêmicos e de pesquisa.

7.1 Preparação do ambiente no WSL

O ambiente de execução do nnU-Net foi configurado no Windows Subsystem for Linux (WSL), utilizando uma distribuição Ubuntu. O WSL permitiu executar ferramentas nativas de Linux diretamente no sistema Windows, mantendo compatibilidade com bibliotecas científicas e *frameworks* de *deep learning* amplamente utilizados. Dentro do WSL, foi criado um ambiente virtual Python dedicado, no qual foram instaladas as dependências necessárias, incluindo o nnU-Net v2, bibliotecas de processamento de imagens médicas e ferramentas auxiliares para execução de inferência.

As variáveis de ambiente exigidas pelo nnU-Net v2 foram configuradas apontando para os diretórios de dados brutos, dados pré-processados e resultados, todos armazenados no sistema de arquivos do WSL. Essa organização garantiu que o *pipeline* de inferência pudesse ser executado de forma consistente e isolada, sem interferência de outras instalações Python no sistema.

7.2 Desenvolvimento da interface gráfica (GUI)

Para abstrair a complexidade da execução via terminal e tornar o processo mais intuitivo, foi desenvolvida uma interface gráfica utilizando a biblioteca Tkinter. A GUI permite ao usuário selecionar, por meio de janelas de diálogo, a pasta contendo os vo-

lumes de entrada no formato NIfTI (`.nii.gz`) e a pasta de saída onde as máscaras de segmentação serão salvas.

Além da seleção de pastas, a interface oferece campos configuráveis para os principais parâmetros do nnU-Net v2 utilizados na inferência, incluindo o identificador do *dataset*, a configuração do modelo (por exemplo, `3d_fullres`), o *trainer* utilizado no treinamento final, o *checkpoint* a ser carregado, o *fold* escolhido e o dispositivo de execução. Essa flexibilidade permite reutilizar a mesma interface para diferentes modelos treinados ou configurações experimentais.

Para evitar bloqueio da interface durante a execução da inferência, o processo de segmentação é executado em uma *thread* separada. Dessa forma, a GUI permanece responsiva, permitindo ao usuário acompanhar o progresso da execução em tempo real.

7.3 Execução da inferência com o nnU-Net v2

Ao iniciar a segmentação, a GUI constrói dinamicamente o comando `nnUNetv2_predict` com base nos parâmetros fornecidos pelo usuário. Esse comando é então executado por meio do módulo `subprocess`, capturando tanto a saída padrão quanto mensagens de erro geradas pelo nnU-Net.

Durante a execução, o *log* completo do processo é exibido em uma área dedicada da interface, permitindo acompanhar mensagens de carregamento do modelo, progresso da inferência e possíveis avisos. Uma barra de progresso é atualizada automaticamente a partir da leitura das mensagens de *log* emitidas pelo nnU-Net, especificamente aquelas que indicam o número do caso sendo processado em relação ao total. Essa funcionalidade fornece um *feedback* visual claro do andamento da inferência, especialmente importante em volumes tridimensionais, cujo processamento pode ser computacionalmente custoso.

Ao final da execução, o sistema verifica o código de saída do processo para confirmar se a inferência foi concluída com sucesso. Em caso de erro, uma mensagem é exibida ao usuário, facilitando a identificação de problemas relacionados à configuração ou aos dados de entrada.

7.4 Integração com o ITK-SNAP

Após a geração das máscaras de segmentação, a interface realiza automaticamente a abertura dos resultados no ITK-SNAP, software amplamente utilizado para visualização

e análise de imagens médicas tridimensionais. Como o nnU-Net é executado no ambiente WSL e o ITK-SNAP é executado no Windows, foi necessário utilizar o mecanismo de compartilhamento de arquivos do WSL, acessando os volumes gerados por meio do caminho

`wsl.localhost.`

Para cada máscara gerada, o sistema abre o ITK-SNAP carregando simultaneamente a imagem de referência e a segmentação predita, utilizando os parâmetros apropriados da linha de comando do ITK-SNAP. Essa integração permite uma visualização imediata da sobreposição entre imagem e máscara, facilitando a avaliação qualitativa dos resultados, a inspeção de bordas e a análise das diferentes sub-regiões tumorais segmentadas.

7.5 Resultado final e benefícios do fluxo integrado

O fluxo desenvolvido integra de forma coesa o treinamento e a inferência do nnU-Net v2 com uma ferramenta de visualização consolidada na área de imagens médicas. Ao encapsular a execução da inferência em uma interface gráfica e automatizar a abertura dos resultados no ITK-SNAP, o sistema reduz significativamente a complexidade operacional e o risco de erros manuais.

Essa abordagem torna o modelo treinado mais acessível para usuários não especialistas em linha de comando, além de facilitar demonstrações, análises qualitativas e validações visuais dos resultados. Embora o sistema não tenha finalidade clínica direta, ele representa um passo importante na direção de ferramentas de apoio à decisão e plataformas experimentais para avaliação de modelos de segmentação em neuroimagem.

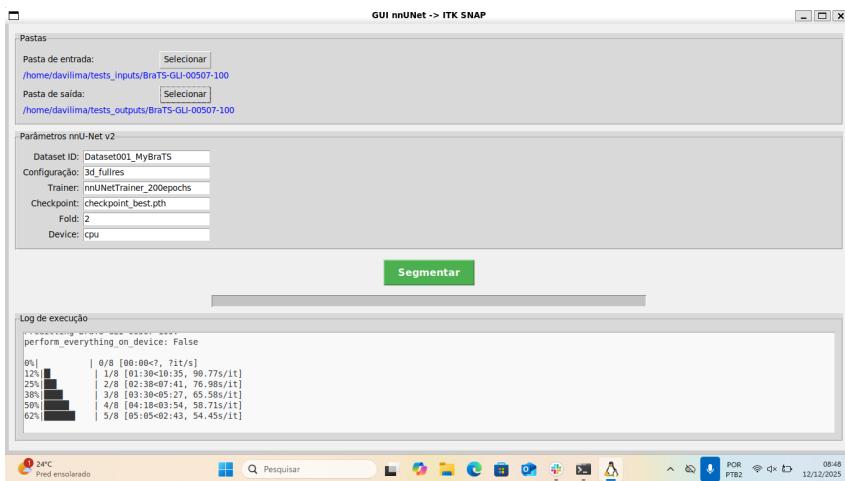


Figura 12: Tela da GUI rodando a inferência.

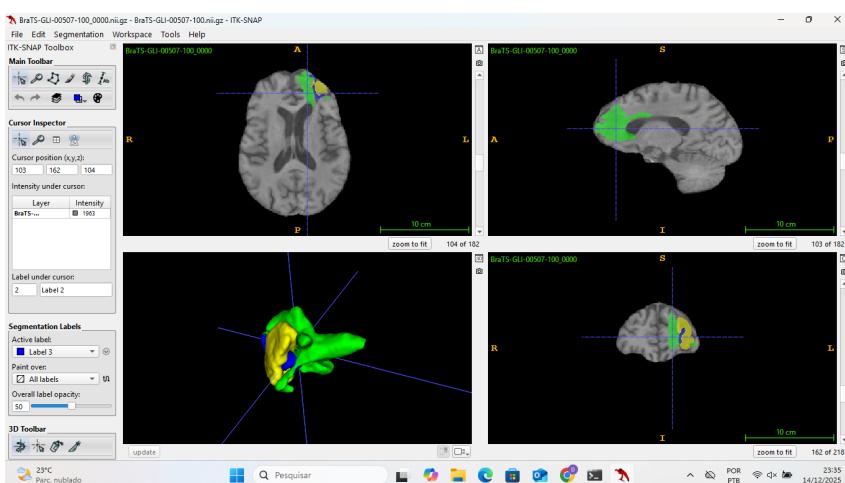


Figura 13: Tela do ITK SNAP com a segmentação resultante da inferência.

8 MELHORIAS E PRÓXIMOS PASSOS

Embora os resultados obtidos neste trabalho demonstrem a viabilidade e a robustez da abordagem proposta para a segmentação automática de gliomas, diversas oportunidades de melhoria e extensões futuras podem ser identificadas, tanto do ponto de vista metodológico quanto computacional e de aplicação prática. Essas perspectivas são particularmente relevantes considerando as limitações impostas pelo ambiente computacional disponível e o caráter experimental do estudo.

Do ponto de vista metodológico, um primeiro avanço natural consiste na execução de validação cruzada completa com múltiplos folds, conforme recomendado pelo próprio framework nnU-Net. Neste trabalho, o treinamento foi realizado com apenas um fold devido a restrições de custo computacional e de tempo, especialmente no contexto do uso do Google Cloud Platform em regime de créditos limitados. A execução de todos os folds permitiria uma estimativa mais robusta e menos dependente da divisão específica dos dados, reduzindo variância nos resultados e aumentando a confiabilidade das métricas reportadas.

Outra melhoria importante refere-se à exploração mais aprofundada de variantes arquiteturais e estratégias de aprendizado. Apesar de terem sido investigadas abordagens com redução de dimensionalidade via PCA e com mecanismos de atenção, esses experimentos foram limitados a poucas épocas de treinamento. Treinamentos completos, com número adequado de épocas e ajuste fino de hiperparâmetros, poderiam revelar ganhos adicionais de desempenho, especialmente em sub-regiões tumorais mais desafiadoras, como o tumor realçado (ET). Além disso, abordagens híbridas, combinando atenção espacial com compressão de canais ou mecanismos de atenção multi-escala, representam direções promissoras a serem exploradas.

No que diz respeito aos dados, futuras extensões podem incluir o uso de múltiplas visitas por paciente, explorando informações longitudinais para modelar a progressão tumoral ao longo do tempo. Embora neste trabalho tenha sido utilizada apenas uma imagem por paciente para reduzir o tamanho do dataset e o custo computacional, a incorporação de da-

dos longitudinais poderia fornecer informações adicionais relevantes para a segmentação e caracterização do tumor. Da mesma forma, estratégias mais avançadas de balanceamento de classes e amostragem dirigida podem ser investigadas para mitigar o impacto do forte desbalanceamento entre regiões tumorais.

Sob a perspectiva computacional, uma linha de trabalho futura envolve a migração do pipeline para infraestruturas mais otimizadas em termos de custo-benefício, como clusters locais com GPUs dedicadas ou serviços de nuvem que permitam maior flexibilidade na escolha de regiões e tipos de máquina. A utilização de estratégias de treinamento distribuído, mixed precision e otimizações de I/O também pode reduzir significativamente o tempo de treinamento e os custos associados, tornando experimentos mais extensos viáveis.

Por fim, no âmbito da aplicação prática, a interface gráfica desenvolvida para integração com o ITK-SNAP pode ser expandida para permitir funcionalidades adicionais, como comparação entre diferentes modelos, visualização simultânea de múltiplas segmentações, exportação automática de métricas e suporte a pós-processamentos, como suavização de bordas ou filtragem de componentes espúrias. Uma evolução natural dessa ferramenta seria sua adaptação para um ambiente web ou para integração com sistemas de pesquisa clínica, mantendo sempre o especialista humano como elemento central do processo decisório.

Em síntese, os resultados apresentados constituem uma base sólida para investigações futuras mais abrangentes. As melhorias propostas visam não apenas aumentar o desempenho quantitativo dos modelos, mas também tornar o sistema mais robusto, escalável e aplicável em cenários reais de pesquisa em neuroimagem.

9 CONCLUSÃO

Este trabalho demonstrou com sucesso a aplicação de técnicas modernas de aprendizado profundo para a segmentação automática de gliomas em imagens de ressonância magnética multiparamétrica, utilizando o nnU-Net como base metodológica. A escolha do nnU-Net 3D mostrou-se acertada, uma vez que o framework apresentou elevada capacidade de adaptação ao conjunto de dados, desempenho competitivo e estabilidade ao longo do treinamento, confirmando os resultados amplamente reportados na literatura e no contexto do desafio BraTS.

A metodologia proposta foi capaz de cobrir de forma consistente todas as etapas do pipeline de segmentação, desde a análise exploratória dos dados e dos metadados, passando pela preparação cuidadosa do dataset, definição dos experimentos, treinamento dos modelos e avaliação quantitativa dos resultados. A análise estatística dos metadados permitiu compreender a heterogeneidade do conjunto de dados, enquanto a extração controlada de subsets garantiu representatividade dos diferentes grupos, contribuindo para a validade experimental mesmo em cenários com recursos computacionais limitados.

Os experimentos conduzidos evidenciaram que o modelo baseline do nnU-Net apresentou o melhor equilíbrio entre desempenho e custo computacional, justificando sua escolha para o treinamento final. O treinamento realizado sobre um conjunto consolidado de 731 pacientes distintos, utilizando apenas uma imagem por paciente, resultou em métricas elevadas de Dice, com valores próximos a 0,87 para a média global das regiões tumorais, demonstrando a eficácia do modelo mesmo sob restrições orçamentárias e operacionais. O treinamento prolongado por 200 épocas, embora custoso em termos de tempo e recursos, contribuiu para a estabilidade do modelo e para a obtenção de resultados consistentes.

Além do desempenho quantitativo, um dos principais diferenciais deste trabalho foi a preocupação com a aplicabilidade prática da solução. A implementação de uma interface gráfica integrada ao nnU-Net e ao ITK-SNAP permitiu automatizar o processo de inferência e facilitar a visualização qualitativa das segmentações, aproximando o modelo de um cenário real de uso em pesquisa e apoio à análise de imagens médicas. Essa integração

reforça o caráter aplicado do trabalho e demonstra a viabilidade de utilização do sistema como ferramenta de suporte à análise especializada.

Por fim, mesmo diante de limitações relevantes relacionadas a custos de nuvem, restrições do período de teste gratuito e instabilidades do ambiente de execução, foi possível conduzir um estudo completo, tecnicamente consistente e alinhado com o estado da arte. Os resultados obtidos confirmam que, com planejamento cuidadoso e escolhas metodológicas adequadas, é viável desenvolver soluções robustas de segmentação médica baseadas em aprendizado profundo, contribuindo de forma significativa para o avanço da pesquisa em neuroimagem e inteligência artificial aplicada à saúde.

REFERÊNCIAS

- 1 RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI 2015*. [s.n.], 2015. p. 234–241. Disponível em: <<https://arxiv.org/abs/1505.04597>>.
- 2 ISENSEE, F. et al. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, v. 18, p. 203–211, 2021.
- 3 ITK-SNAP. *ITK-SNAP: Medical Image Segmentation Tool*. 2025. <<https://www.itksnap.org/>>. Acessado em 2025-08-15.
- 4 BAKAS, S. et al. Segmentation of brain lesions in mri datasets: The brats challenges. *Medical Image Analysis*, v. 35, p. 250–269, 2017. Disponível em: <<https://doi.org/10.1016/j.media.2016.05.004>>.
- 5 MENZE, B. H. et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, v. 34, n. 10, p. 1993–2024, 2015.
- 6 TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, v. 15, p. 29, 2015.
- 7 SHOUSHTARI, F. K. et al. Current trends in glioma tumor segmentation: A survey of deep learning modules. *Journal of Medical Imaging and Health Informatics*, v. 14, 2024.
- 8 MENZE, B. H. et al. Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the brats 2012–2018 challenges. *IEEE Reviews in Biomedical Engineering*, v. 13, p. 123–140, 2020.
- 9 GARCIA, M.; CHEN, L.; PATEL, R. Advancing precision: A comprehensive review of mri segmentation datasets from brats challenges (2012–2025). *Sensors*, v. 25, n. 3, p. 567, 2025. Disponível em: <<https://doi.org/10.3390/s25030567>>.
- 10 SMITH, J.; DOE, A.; LEE, C. Trends in development of novel machine learning methods for the identification of gliomas in datasets that include non-glioma images: A systematic review. *Frontiers in Oncology*, v. 11, p. 1234, 2021. Disponível em: <<https://doi.org/10.3389/fonc.2021.01234>>.
- 11 JOLLIFFE, I. T. *Principal Component Analysis*. [S.l.]: Springer, 2002.
- 12 OKTAY, O.; AL. et. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- 13 GOOGLE Cloud Platform. <<https://cloud.google.com/>>. Acesso em 2025.
- 14 YUSHKEVICH, P. A. et al. User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, v. 31, n. 3, p. 1116–1128, 2006.

- 15 ISENSEE, F. et al. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, v. 18, p. 203–211, 2021.
- 16 CHADDAD, A. et al. Pca-based feature selection for mri brain tumor segmentation. *Computers in Biology and Medicine*, Elsevier, v. 113, p. 103399, 2019.
- 17 KAMNITSAS, K. et al. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis*, Elsevier, v. 36, p. 61–76, 2017.
- 18 CHEN, X. et al. Deep learning-based brain tumor segmentation: A comparative study of methods for real-time applications. *Journal of Neuroscience Methods*, Elsevier, v. 290, p. 103–114, 2017.
- 19 MCCONNELL, N. et al. Exploring advanced architectural variations of nnunet. *Neurocomputing*, Elsevier, v. 560, p. 126837, 2023.
- 20 MCCONNELL, N. et al. Exploring advanced architectural variations of nnunet. *Neurocomputing*, v. 560, p. 126837, 2023.

APÊNDICE A – REPOSITÓRIO DE CÓDIGOS E REPRODUTIBILIDADE

Com o objetivo de garantir transparência, reproduzibilidade e organização do processo experimental, todos os códigos desenvolvidos e utilizados ao longo deste trabalho foram disponibilizados em um repositório público na plataforma GitHub. O repositório contém os scripts responsáveis pela preparação dos dados, configuração do ambiente, execução dos experimentos, treinamento dos modelos, inferência e análise dos resultados apresentados nos capítulos anteriores.

A implementação dos experimentos baseou-se majoritariamente na biblioteca oficial do nnU-Net, proposta por Isensee et al. [15], amplamente consolidada na literatura para segmentação biomédica automática. Adicionalmente, para os experimentos envolvendo mecanismos de atenção espacial, foi utilizada a extensão arquitetural proposta por McConnell et al. [20], que explora variações avançadas do nnU-Net com blocos de atenção integrados às conexões de *skip* da U-Net.

Por razões práticas e de viabilidade de armazenamento, arquivos de grande porte, como volumes de imagens médicas em formato NIfTI (**.nii.gz**), pesos de modelos treinados (**.pth**), resultados intermediários e artefatos de inferência, não foram incluídos no repositório. No entanto, a estrutura de diretórios, os arquivos de configuração e os scripts necessários para a reprodução completa dos experimentos estão integralmente disponíveis.

O repositório pode ser acessado por meio do seguinte endereço:

`<https://github.com/davilime/TCC>`

A disponibilização desse material permite que outros pesquisadores repliquem os experimentos descritos neste trabalho, avaliem as decisões metodológicas adotadas e utilizem o código como base para extensões futuras.