



UFCSPA

**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CIÊNCIAS
DA SAÚDE DE PORTO ALEGRE**

**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS
DA INFORMAÇÃO E GESTÃO EM SAÚDE (PPGTIGSAÚDE)**

BRUNO YUKIO TAKARA

**AVALIAÇÃO DE REDES NEURAIIS CONVOLUCIONAIS PARA A DELIMITAÇÃO DE
METÁSTASES CEREBRAIS EM IMAGENS DE RESSONÂNCIA MAGNÉTICA**

PORTO ALEGRE - RS

2024

BRUNO YUKIO TAKARA

AVALIAÇÃO DE REDES NEURAIS CONVOLUCIONAIS PARA
A DELIMITAÇÃO DE METÁSTASES CEREBRAIS EM IMAGENS
DE RESSONÂNCIA MAGNÉTICA

Dissertação no Programa de Mestrado Acadêmico em
Tecnologias da Informação e Gestão em Saúde da
Universidade Federal de Ciências da Saúde de Porto
Alegre.

Área de concentração: Tecnologias da Informação
Inteligente

Linha de pesquisa: Sistemas Inteligentes e Aplicações na
Saúde

Orientador: Prof. Dr. Mirko Salomón Alva Sánchez

Coorientadora: Profa. Dra. Viviane Rodrigues Botelho

PORTO ALEGRE - RS
2024

Catálogo na Publicação

Takara, Bruno Yukio

Avaliação de redes neurais convolucionais para a delimitação de metástases cerebrais em imagens de ressonância magnética / Bruno Yukio Takara. -- 2024. 126 p. : il., graf., tab. ; 30 cm.

Dissertação (mestrado) -- Universidade Federal de Ciências da Saúde de Porto Alegre, Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde, 2024.

Orientador(a): Prof. Dr. Mirko Salomón Alva Sánchez ;
coorientador(a): Profa. Dra. Viviane Rodrigues Botelho.

1. Aprendizado Profundo. 2. Neoplasias Encefálicas. 3. Imageamento por Ressonância Magnética. 4. Radiocirurgia.
I. Título.

BRUNO YUKIO TAKARA

**AVALIAÇÃO DE REDES NEURAIS CONVOLUCIONAIS PARA
A DELIMITAÇÃO DE METÁSTASES CEREBRAIS EM IMAGENS
DE RESSONÂNCIA MAGNÉTICA**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre como requisito parcial para a obtenção do título de Mestre em Sistemas Inteligentes e Aplicações na Saúde.

Orientador: Prof. Dr. Mirko Salomón Alva Sánchez

Co-orientadora: Profa. Dra. Viviane Rodrigues Botelho

Aprovada em: Porto Alegre - RS, 31/07/2024

BANCA EXAMINADORA

Profa. Dra. Carla Diniz Lopes Becker
Universidade Federal de Ciências da Saúde
de Porto Alegre (UFCSPA)

Profa. Dra. Sandra Guzmán
Universidad Nacional Federico Villarreal
(UNFV)

Profa. Dra. Patrícia Nicolucci
Universidade de São Paulo (USP)

Dedico àqueles que sempre estiveram comigo,
minha família.

Agradecimentos

Agradeço ao meu orientador Mirko Salomón Alva Sánchez pela oportunidade de fazer parte do programa de pós-graduação, pelas discussões sobre a pesquisa, pelos ensinamentos na área da física da radioterapia, pela paciência e pelas orientações profissionais, acadêmicas e pessoais durante minha jornada científica. Agradeço à minha co-orientadora Viviane Rodrigues Botelho, pelas discussões em inteligência artificial, pelas recomendações de leitura, pelas frases de motivação e também puxões de orelha nos momentos em que nada parecia fluir e que direcionou a pesquisa por novos caminhos quando esses momentos chegaram.

Também gostaria de agradecer a CAPES, que realizou o financiamento da pesquisa e por também tornar possível todas as outras pesquisas no Brasil. Agradeço aos Professores do PPGTIGSaúde que promoveram e promovem o programa em um ambiente tão acolhedor e tão humano, sempre com vigor, amor e paixão por ensinar. Fico muito honrado de fazer parte de um programa tão bonito e com tantos profissionais empenhados em sua qualidade. Agradeço também à Coordenação do programa, em especial à Coordenadora a Professora Mellina da Silva Terres, que sempre foi muito solícita e promoveu diversos momentos de integração, transparência e acolhimento.

Aos meus colegas de trabalho em inteligência artificial, o Felipe Ferreira de Freitas que me introduziu ao tema e me ensinou a resolver os problemas com o código, meu orientador Mirko Salomón Alva Sánchez por proporcionar um espaço de discussão em inteligência artificial e aos estudantes Fábio e Brenda pelas discussões no grupo de pesquisa. Não poderia faltar um agradecimento ao ChatGPT que me ajudou, e muito, a debugar os códigos e que me deu inúmeras respostas certas e erradas que me guiaram aos resultados obtidos nesse trabalho.

Também agradeço à Secretaria do programa, por tirar todas as dúvidas e por todos os problemas resolvidos. Agradeço aos meus colegas de turma que foram unidos nessa empreitada, sempre nos ajudando pelo nosso grupo do whats.

Agraço à minha namorada Victória, pela parceria e companheirismo, por ter estado ao meu lado nos momentos difíceis, por ter me acolhido nos dias em que ocorrem as enchentes, pela paciência nos dias em que as demandas eram infinitas e que não podíamos nos ver, obrigado por fazer tudo ser tão leve e por me fazer tão bem.

Por fim, porém não menos importante, agradeço à minha família, minha mãe Eni, meu pai Claiton e minha irmã Juliana, por sempre apoiarem minhas decisões e por estarem sempre presentes em minha vida, sem vocês nada disso teria sido possível.

“A vida é fundamentalmente um processo de comunicação constante e mútua, e comunicar-se é emitir e receber ondas e radiações “*minika ye minienie*”. Esse processo de receber e liberar ou transmitir, “*tâmbula ye tambikisa*”, é a chave para o jogo de sobrevivência do ser humano. Uma pessoa é constantemente banhada pela carga das radiações, “*zitu kia minienie*”. A carga “*zitu/-demo*” das radiações pode ter um impacto negativo ou positivo sobre qualquer pequeno ser, por exemplo, uma pessoa, que representa o “*kolo*” (nó) mais vibrátil das relações.”
— *Kimbwandênde Kia Bunseki Fu-Kiau*

Resumo

As metástases cerebrais múltiplas acometem cerca de 20% dos pacientes com algum tipo de tumor e seu tratamento pode ser realizado, entre outras formas, através da radiocirurgia estereotáxica. O delineamento realizado por um *software* das metástases cerebrais para o planejamento do tratamento pode ter grande valor na prática radioterapêutica, visto que essa tarefa costuma demandar bastante tempo e ser de difícil identificação quando as metástases se apresentam em formatos pequenos. Assim, é proposta a aplicação de redes neurais artificiais em imagens de ressonância magnética (RM) para o delineamento automático de regiões tumorais com metástases e órgãos de risco, para o preparo de um planejamento radioterápico mais acurado e com menor risco de sobredosagem ao paciente. Para isso, três modelos de *Deep Learning* que realizam a tarefa de segmentação de metástases cerebrais encontrados na literatura foram utilizados em três bancos de dados de metástases cerebrais disponíveis publicamente, os resultados obtidos por esses modelos foram comparados para determinar adequação na prática clínica. Além disso, as publicações envolvendo esses modelos foram avaliadas através dos *checklists* CLAIM e CLAMP. Adicionalmente foi proposto e avaliado um modelo de segmentação de metástases cerebrais utilizando uma rede U-Net 2D com diferentes combinações de hiperparâmetros e funções de perda. As métricas de performance incluirão o coeficiente Dice por *slice* (swDSC), coeficiente Dice volumétrico (3D-DSC) e distância de Hausdorff (HD). Os modelos Met-Seg, AURORA e Raidionics apresentaram valores médios de 3D-DSC iguais a $0,44 \pm 0,23$, $0,55 \pm 0,29$ e $0,53 \pm 0,29$, e valores médios de HD iguais a 101 ± 42 mm, 116 ± 81 mm e 119 ± 82 mm, respectivamente. Para o desenvolvimento do modelo, os hiperparâmetros $\eta = 9$ e $\psi = 2$ com peso de balanceamento de classes igual a 0,2 da função de perda BCE somada à Tversky e *Boundary* foram o que resultaram na melhor performance, com um valor de swDSC de 0,48 para o conjunto de teste. Por fim, para a aplicação na prática clínica, ainda é necessário o desenvolvimento de modelos com melhores performances na segmentação de metástases cerebrais, principalmente para volumes tumorais pequenos e em maior quantidade. Com o passar do tempo, espera-se que mais bancos de dados com imagens de metástases cerebrais se tornem disponíveis para treinamento, o que deve melhorar a performance dos modelos desenvolvidos. Além disso, uso da função de perda BCE somada às funções de Tversky e *Boundary* pode diminuir a quantidade de falsos positivos e falsos negativos do resultado final e pode ser um caminho para esse desenvolvimento.

Palavras-chave: Aprendizado Profundo. Neoplasias Encefálicas. Imageamento por Ressonância Magnética. Radiocirurgia.

Abstract

Multiple brain metastases affect about 20% of patients with some type of tumor, and their treatment can be carried out, among other methods, through stereotactic radiosurgery. The automatic delineation of brain metastases for treatment planning, can be highly valuable in radiotherapy practice, as this task often demands considerable time and can be challenging when the metastases are small in size. So, the application of artificial neural networks on magnetic resonance imaging (MRI) for the automatic delineation of tumoral regions with metastases and organs at risk is proposed, aiming for a more accurate radiotherapy planning with lower risk of ionizing radiation overexposure to the patient. To accomplish this task, three Deep Learning models that perform brain metastasis segmentation, found in the literature, were used on three publicly available brain metastasis databases. The results obtained by these models were compared to determine their suitability for clinical practice. Additionally, publications involving these models were evaluated using the CLAIM and CLAMP checklists. Furthermore, a brain metastasis segmentation model was proposed and evaluated using a 2D U-Net network with different combinations of hyperparameters and loss functions. The performance metrics included the slice-wise Dice coefficient (swDSC), volumetric Dice coefficient (3D-DSC), and Hausdorff distance (HD). The Met-Seg, AURORA, and Raidionics models showed mean 3D-DSC values of 0.44 ± 0.23 , 0.55 ± 0.29 , and 0.53 ± 0.29 , and mean HD values of 101 ± 42 mm, 116 ± 81 mm, and 119 ± 82 mm, respectively. For the development of the model, the hyperparameters $\eta = 9$ and $\psi = 2$ with a class balance weight of 0.1 for the BCE loss function combined with Tversky and Boundary, resulted in the best performance, with a swDSC value of 0.48 for the test set. All in all, for clinical application, the development of models with better performance in brain metastasis segmentation is still needed, particularly for small and numerous tumor volumes. Over time, it is expected that more databases with brain metastasis images will become available for training, which should improve the performance of the developed models. Additionally, the use of the BCE loss function combined with Tversky and Boundary functions may reduce the number of false positives and false negatives in the final result, potentially paving the way for this development.

Keywords: Deep Learning. Brain Neoplasms. Magnetic Resonance Imaging. Radiosurgery.

Lista de figuras

Figura 1 – Imagens de RM ponderadas em T1 com contraste em hipersinal anelar para casos de (A) toxoplasmose, (B) neurocisticercose, (C) glioblastoma, (D) esclerose múltipla tumefativa, (E) linfoma e (F) metástase de tumor pulmonar.	3
Figura 2 – Exemplo do uso da ferramenta de desenho à mão livre do TPS Eclipse da Varian.	8
Figura 3 – Diferentes modos de aquisição de imagem de metástases cerebrais. TC - tomografia computadorizada. T1, FLAIR e T2* - sequências de pulso para imagens de ressonância magnética.	10
Figura 4 – Imagens de RM ponderadas em (A) T1 com contraste de gadolínio, (B) DWI sem contraste e (C) mapa ADC sem contraste em caso de metástase cerebral única de tumor primário de mama em paciente mulher de 44 anos.	11
Figura 5 – Relação entre os objetos que compõem um modelo de DL.	12
Figura 6 – Representação esquemática dos blocos de uma rede neural convolucional	12
Figura 7 – Exemplo do cálculo da distância de Hausdorff entre a máscara de segmentação de referência (<i>Ground Truth</i>) e a segmentação feita pelo modelo, ou predição. Os números na imagem ao centro indicam a posição em <i>pixels</i>	21
Figura 8 – Fluxo de inclusão dos estudos seguindo o protocolo PRISMA para revisões sistemáticas.	33
Figura 9 – Publicações ao longo dos anos dos artigos encontrados nas bases PubMed, IEEE Xplore e Epistemonikos que foram selecionados para a leitura completa	33
Figura 10 – Quantidade das publicações incluídas na revisão por país	34
Figura 11 – Fluxo de trabalho e análises realizadas nessa pesquisa.	52
Figura 12 – Exemplos das imagens em PNG do banco <i>BrainMetShare</i>	53
Figura 13 – Exemplos das segmentações encontradas no <i>dataset</i> do TCIA, em que (a) é a região com edema peritumoral (em verde) encontrada na sequência FLAIR, (b) possui as marcações encontradas em T1 pós contraste para a região com necrose (em vermelho) e a região tumoral com absorção de contraste (em azul) e em (c) encontram-se as máscaras de segmentação.	54
Figura 14 – Exemplo de um <i>slice</i> e sua respectiva máscara de segmentação de um paciente do banco de dados de Ocaña-Tienda <i>et al.</i> (2023)	55

Figura 15 – Plot violino da quantidade de volumes tumorais encontrados por exame, nos diferentes bancos de dados.	56
Figura 16 – Plot violino do volume tumoral total por exame, nos diferentes bancos de dados.	56
Figura 17 – Planos axial, sagital e coronal originais, com o uso do <i>software</i> FSL-BET e com o uso da biblioteca HD-BET para a extração cerebral na etapa de pré-processamento do banco de dados de Ocaña-Tienda <i>et al.</i> (2023).	59
Figura 18 – Exemplo do pré-processamento em que ocorre a remoção da região de edema peritumoral da máscara de segmentação das imagens do banco do TCIA.	60
Figura 19 – Fluxograma dos passos metodológicos.	63
Figura 20 – Diagrama da rede U-Net	63
Figura 21 – Exemplo das transformações de Data Augmentation aplicadas em duas imagens de treino da rede neural.	64
Figura 22 – Validação cruzada em <i>k-folds</i> para $k = 3$	65
Figura 23 – Soma acumulada do número de volumes tumorais por exame, identificados pelo algoritmo de contagem de <i>clusters</i> para o banco de dados de Ocaña-Tienda <i>et al.</i> (2023).	68
Figura 24 – Coeficiente Dice médio por <i>slice</i> swDSC calculado em todas as combinações de sequências do modelo de (OTTESEN et al., 2023) para o banco de dados BMS em formato PNG considerando-se os <i>slices</i> compostos exclusivamente de verdadeiros negativos (BG) e desconsiderando-se esses <i>slices</i> (noBG) para o cálculo do valor de swDSC. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.	73
Figura 25 – Coeficientes de correlação de Pearson dos valores de swDSC do modelo de (OTTESEN et al., 2023) para o banco de dados BMS em seu formato PNG com a remoção de alguma modalidade em relação ao todo para o caso em que se considera os verdadeiros negativos dos <i>slices</i> sem metástases como tendo o valor de DSC = 1. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.	74
Figura 26 – Coeficientes de correlação de Pearson dos valores de DSC do modelo de (OTTESEN et al., 2023) para o banco de dados do BMS em seu formato PNG com a remoção de alguma modalidade em relação ao todo para o caso em que não se considera os verdadeiros negativos dos <i>slices</i> sem metástases. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.	75

Figura 27 – Coeficientes de correlação de Pearson dos valores de swDSC do modelo Met-Seg (OTTESEN et al., 2023) para o banco de dados do BMS em formato NIFTI nativo com a remoção de alguma modalidade em relação ao todo para o caso em que não se considera os verdadeiros negativos dos <i>slices</i> sem metástases. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.	76
Figura 28 – Valores dos coeficientes de correlação de Pearson para os valores dos coeficientes Dice médio por <i>slice</i> do modelo Met-Seg para o banco de dados do TCIA entre os pares com e sem a remoção de alguma(s) modalidade(s) de imagem, para o caso em que não se considera os verdadeiros negativos dos <i>slices</i> sem metástases. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.	78
Figura 29 – Coeficiente Dice calculado em todas as combinações de sequências do modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA considerando-se os <i>slices</i> compostos exclusivamente de verdadeiros negativos (BG) e desconsiderando-se esses <i>slices</i> (noBG) para o cálculo do valor médio de DSC por <i>slice</i> . Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.	79
Figura 30 – Exemplo para dois <i>slices</i> em um caso do paciente de índice 92 do banco de dados do TCIA, em que o modelo Met-Seg realiza erroneamente (em vermelho) as segmentações de volumes tumorais. Em azul, nas duas primeiras imagens do topo, está destacada a região com presença metastática delimitada por radiologistas (<i>ground truth</i>).	80
Figura 31 – Valores dos coeficientes de correlação de Pearson para os valores dos coeficientes Dice volumétricos (3D-DSC) do modelo AURORA para o banco de dados BMS em seu formato NIFTI nativo entre os pares com e sem a remoção de alguma(s) modalidade(s) de imagem. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.	82
Figura 32 – Coeficiente Dice médio por <i>slice</i> versus a quantidade de volumes tumorais por paciente para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: <i>BrainMetShare</i> (em vermelho), TCIA (em azul) e Ocanã-Tienda (em amarelo).	86

Figura 33 – Coeficiente Dice médio por <i>slice</i> versus a o volume total tumoral por paciente para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: <i>BrainMetShare</i> (em vermelho), TCIA (em azul) e Ocaña-Tienda (em amarelo).	87
Figura 34 – Coeficiente Dice médio por <i>slice</i> versus a mediana de volumes tumorais por paciente para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: <i>BrainMetShare</i> (em vermelho), TCIA (em azul) e Ocaña-Tienda (em amarelo).	88
Figura 35 – Coeficiente Dice versus o tamanho da imagem do exame para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: <i>BrainMetShare</i> (em vermelho), TCIA (em azul) e Ocaña-Tienda (em amarelo).	89
Figura 36 – Coeficientes Dice (DSC) médio por <i>slice</i> para as diferentes origens de tumor primário, calculados para os três modelos: Met-Set (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) no banco de dados Ocaña-Tienda (OCAÑA-TIENDA et al., 2023). SCLC - <i>Small Cell Lung Cancer</i> , NSCLC - <i>Non Small Cell Lung Cancer</i>	90
Figura 37 – Distâncias de Hausdorff (HD) calculadas para os três modelos: Met-Set (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) em cada um dos três bancos de dados: <i>BrainMetShare</i> (BMS) (GRØVIK et al., 2020), TCIA (RAMAKRISHNAN et al., 2024) e Ocaña-Tienda (OCAÑA-TIENDA et al., 2023). Note que o banco BMS possui duas versões, a primeira em PNG e a segunda em NifTI, e o modelo ideal apresenta valores de distância de Hausdorff próximas de 0 milímetros. Além disso, os <i>outliers</i> muito acima do restante do conjunto de previsões representam erros em que o modelo não conseguiu identificar a presença da metástase.	91

Figura 38 – Coeficientes Dice (DSC) calculados para os três modelos: Met-Set (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) em cada um dos três bancos de dados: BrainMetShare (GRØVIK et al., 2020), TCIA (RAMAKRISHNAN et al., 2024) e Ocaña-Tienda (OCAÑA-TIENDA et al., 2023). Note que o banco BMS possui duas versões, a primeira em PNG e a segunda em NIfTI, e o modelo ideal apresenta valores de Coeficientes Dice próximos a 1.	92
Figura 39 – Comparação das métricas de coeficientes Dice por <i>slice</i> médio (slicewise DSC), coeficiente Dice para todo o volume (3D DSC), e o valor da interseção sobre a união (IoU) calculados para os três modelos: Met-Set (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) em cada um dos três bancos de dados: <i>BrainMetShare</i> em PNG (old_BMS) e em NIfTI (BMS) (GRØVIK et al., 2020), TCIA (RAMAKRISHNAN et al., 2024) e Ocaña-Tienda (Ocana) (OCAÑA-TIENDA et al., 2023). Note que o banco BMS possui duas versões, a primeira em PNG e a segunda em NIfTI, e o modelo ideal apresenta valores de DSC e IoU próximos a 1.	93
Figura 40 – Arquitetura de rede utilizada para identificação	95
Figura 41 – Matrizes de confusão para diferentes sequências de pulso onde: <i>tp</i> - <i>true positive</i> , <i>fp</i> - <i>false positive</i> , <i>tn</i> - <i>true negative</i> , <i>fn</i> - <i>false negative</i>	95
Figura 42 – Imagem com uma sequência em cada canal RGB (direita) e Imagem original FLAIR (esquerda)	97
Figura 43 – Imagens mostrando a eficiência da U-Net em realizar segmentações. Imagem com os três canais de cores para cada modalidade de aquisição (acima à esquerda), imagem da aquisição FLAIR (acima à direita), máscara de segmentação feita pelos neurorradiologistas (abaixo à esquerda) e máscara de segmentação realizada pela U-Net (abaixo à direita) para duas imagens (a) com 2 metástases e em (b) com 1 metástase pequena.	98
Figura 44 – Mapas de calor dos coeficientes Dice para a variação dos coeficientes $\frac{\eta}{2}$ e $\frac{\psi}{2}$ em números inteiros de 1 a 10 da função perda Tversky. O mapa da esquerda foi feito para a <i>learning rate</i> de 0,01 e o mapa da direita para a <i>learning rate</i> de 0,001.	100
Figura 45 – Mapas de calor dos coeficientes Dice para a variação dos coeficientes η e ψ entre 0,25 e 4 da função perda Tversky ponderada para a <i>learning rate</i> de 0,001. Os quadrados destacados em vermelho e amarelo representam, respectivamente, as funções perda Dice e IoU	101

Figura 46 – Comparação dos valores do coeficiente Dice para o conjunto de teste com o uso de três funções perdas e com variação dos pesos para o balanceamento das classes, cujos valores vão de de 0,1 até 10^{-16} e estão apresentados na legenda pelo negativo do logaritmo.	102
Figura 47 – Mapa de calor dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.	102
Figura 48 – Mapa de calor dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.	103
Figura 49 – Gráfico de barras empilhadas dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.	103
Figura 50 – Gráfico de barras empilhadas dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.	104
Figura 51 – Gráfico de barras empilhadas dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE, Tversky e Boundary ponderada com diferentes pesos para o balanceamento de classe.	104
Figura 52 – Imagem de teste e sua máscara de segmentação. Esse exemplo foi escolhido por representar uma metástase múltipla de difícil identificação devido ao tamanho.	105
Figura 53 – Resultados da segmentação do modelo com uso de <i>Data Augmentation</i> para diferentes composições de η e ψ utilizando o peso de balanceamento de classes igual a 0,1 na função de perda BCE combinada à Tversky ponderada e à <i>Boundary</i> . O melhor e o pior modelo estão destacados no canto superior direito com uma seta dentro de um círculo verde e vermelho, respectivamente.	106
Figura 54 – Resultados da segmentação do modelo com uso de <i>Data Augmentation</i> para diferentes composições de η e ψ utilizando o peso de balanceamento de classes igual a 0,001 na função de perda BCE combinada à Tversky ponderada e à <i>Boundary</i> . O melhor e o pior modelo estão destacados no canto superior direito com uma seta dentro de um círculo verde e vermelho, respectivamente.	107
Figura 55 – Curvas de aprendizado para o modelo final em treino realizado utilizando imagens apenas em T1 com contraste do banco de dados <i>Brain-MetShare</i> em formato PNG.	108

Lista de tabelas

Tabela 1 – Comparação do número N de casos/pacientes no treino o <i>split</i> em treino, validação e teste (T/V/T), o coeficiente Dice (DSC), a precisão (PPV), a sensibilidade (TPR) e a quantidade de falsos positivos por paciente (FPP) dos estudos encontrados em literatura para a segmentação de metástases cerebrais. Os melhores resultados estão destacados em negrito.	35
Tabela 2 – Comparação do número N de casos/pacientes no treino o <i>split</i> em treino, validação e teste (T/V/T), o coeficiente Dice (DSC), a precisão (PPV), a sensibilidade (TPR) e a quantidade de falsos positivos por paciente (FPP) dos estudos encontrados em literatura para a segmentação de metástases cerebrais. Os melhores resultados estão destacados em negrito.	36
Tabela 3 – Comparação dos parâmetros de aquisição dos três bancos de dados, em que os parâmetros das 5 primeiras linhas apresentam os valores de mediana acompanhada dos valores mínimo e máximo entre parênteses. TE - tempo de eco, TR - tempo de repetição, N/A - informações não disponíveis. Os números entre parênteses nas intensidades de campo indicam a quantidade de imagens para cada intensidade.	57
Tabela 4 – Contagem da quantidade (Qtd.) de <i>clusters</i> para o banco de dados de Ocaña-Tienda <i>et al.</i> (2023) bem como a quantidade de acertos do número de <i>clusters</i> por paciente (TP); Os valores 591 e 373 representam os valores <i>Ground Truth</i> (GT), e vx indica o tamanho mínimo do <i>voxel</i> para ser incluído na contagem	68
Tabela 5 – Número de volumes considerados para cada um dos arquivos de segmentação, os números entre parênteses são as quantidades de volumes formados por menos do que 6 <i>voxels</i> , que foram considerados como ruído e não foram incluídos na contagem. <i>Ground Truth</i> é a anotação feita por radiologistas, Met-Seg (OTTESEN <i>et al.</i> , 2023), AURORA (BUCHNER <i>et al.</i> , 2022) e Raidionics (BOUGET <i>et al.</i> , 2023) são os modelos de redes neurais. *Não foi possível carregar um paciente no modelo Raidionics, com um volume tumoral, do banco de Ocaña-Tienda <i>et al.</i> (2023).	70
Tabela 6 – Valores médios para a contagem de volumes tumorais, o volume tumoral total (V_{total}) e a mediana do volume tumoral individual ($V_{mediano}$) para os bancos de dados BMS, TCIA e Ocaña-Tienda <i>et al.</i> (2023). . .	70

Tabela 7 – Métricas de performance do coeficiente Dice médio por slice (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), sensibilidade geral (Sens.), sensibilidade por paciente (Sens./pat.) e falsos positivos por paciente (FP/pat.) para as predições com todas as modalidades presentes do banco de dados <i>BrainMetShare</i> .	71
Tabela 8 – Métricas de performance do coeficiente Dice médio por <i>slice</i> (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com algumas das modalidades faltantes, denotadas por mFLAIR, mBRAVO, mT1pos, mT1pre e mT1 para, respectivamente, os casos com a remoção das sequências FLAIR, BRAVO, T1 pós contraste, T1 pré contraste e ambas as sequências T1 do modelo de (OTTESEN et al., 2023) para o banco de dados <i>BrainMetShare</i> .	72
Tabela 9 – Métricas de performance do coeficiente Dice médio por <i>slice</i> (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com apenas uma das sequências FLAIR, BRAVO, T1 pós contraste (T1pos), T1 pré contraste (T1pre) e ambas as sequências T1 do modelo de (OTTESEN et al., 2023) para o banco de dados <i>BrainMetShare</i> .	72
Tabela 10 – Métricas de performance do coeficiente Dice médio por <i>slice</i> (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com algumas das modalidades faltantes, denotadas por mFLAIR, mT1pos e mT1pre para, respectivamente, os casos com a remoção das sequências FLAIR, T1 pós contraste e T1 pré contraste do modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA.	76
Tabela 11 – Métricas de performance do coeficiente Dice (DSC) médio por slice para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com a presença de todas as sequências utilizando o modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA e o <i>BrainMetShare</i> .	77
Tabela 12 – Métricas de performance do coeficiente Dice médio por <i>slice</i> (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com apenas uma das sequências FLAIR, T1 pós contraste (T1pos) e T1 pré contraste (T1pre) do modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA.	77

Tabela 13 – Comparação dos resultados do coeficiente Dice volumétrico (3D-DSC) obtidos para as predições realizadas pelo modelo AURORA (BUCHNER et al., 2023) utilizando todas as modalidades de imagens presentes nos bancos de dados dos autores, BMS, TCIA e Ocaña-Tienda <i>et al.</i> (2023).	80
Tabela 14 – Média dos coeficientes Dice volumétricos (3D-DSC) das predições do modelo AURORA (BUCHNER et al., 2023) para o banco de dados BMS em seu formato nativo NIfTI com a remoção da sequência T1 pré-contraste (mT1pre), utilizando apenas a sequência FLAIR, apenas as sequências T1 pré e pós contraste (T1) e apenas a sequência T1 pré (T1pre) e pós (T1pos) contraste. Em todos os casos a sequência BRAVO não foi utilizada, pois o modelo não foi treinado para esse tipo de imagem.	81
Tabela 15 – Comparação dos resultados do coeficiente Dice volumétrico (3D-DSC) obtidos para as predições realizadas pelo modelo Raidionics (BOUGET et al., 2023) utilizando apenas a modalidade T1 pós contraste dos bancos de dados dos autores, BMS, TCIA e Ocaña-Tienda <i>et al.</i> (2023).	83
Tabela 16 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelos modelos Met-Seg, AURORA e Raidionics nos bancos de dados <i>BrainMetShare</i> em formato original PNG, <i>BrainMetShare</i> em formato original NIfTI, TCIA e Ocaña-Tienda <i>et al.</i> (2023).	85
Tabela 17 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelos modelos Met-Seg, AURORA e Raidionics nas imagens em ponderação T1 com contraste dos bancos de dados <i>BrainMetShare</i> em formato original NIfTI, TCIA e Ocaña-Tienda <i>et al.</i> (2023).	85
Tabela 18 – Resultados dos treinos para determinação das melhores sequências de pulso no diagnóstico de metástases cerebrais pela rede neural, com medições de M - modelo, P - precisão, A - acurácia, S - sensibilidade, E - especificidade e F1 - <i>F1-Score</i> . Os melhores valores estão destacados em negrito.	96
Tabela 19 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelo nosso modelo e comparado aos modelos Met-Seg, AURORA e Raidionics nos bancos de dados <i>BrainMetShare</i> em formato original NIfTI, TCIA e Ocaña-Tienda <i>et al.</i> (2023). Os melhores resultados por banco de dados estão destacados em negrito.	108

Tabela 20 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelo modelo desenvolvido e os modelos da literatura Met-Seg, AURORA e Raidionics nas imagens em ponderação T1 com contraste dos bancos de dados *BrainMetShare* em formato original NIfTI, TCIA e Ocaña-Tienda *et al.* (2023). Os melhores resultados por banco de dados estão destacados em negrito. 109

Lista de quadros

Quadro 1 – Checklists CLAIM e CLAMP para as publicações dos três modelos. . 84

Lista de abreviaturas e siglas

2D	Bidimensional
3D	Tridimensional
3D-DSC	Coeficiente Dice volumétrico.
ADC	Coeficiente de difusão aparente, do inglês <i>Apparent Diffusion Coefficient</i>
AuROCC	<i>Area under Receiver Operating Characteristic Curve</i>
BMS	<i>BrainMetShare</i>
BN	<i>Batch Normalization</i>
BraTS	<i>Brain Tumor Segmentation Dataset</i>
BRAVO	<i>Inversion-Recovery-Prepared Fast-Spoiled Gradient Recalled Brain Volume</i> ponderada em T_1
BS	<i>Batch Size</i>
CAD	<i>Computer-aided Detection</i>
CADx	<i>Computer-aided Diagnosis</i>
CBCT	<i>Cone Beam Computed Tomography</i>
CC	Com Contraste
CK	<i>Cyber Knife</i>
CLAIM	<i>Checklist for Artificial Intelligence in Medical Imaging</i>
CLAMP	<i>Checklist for AI in Medical Physics</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CEP	Comitê de Ética em Pesquisa
CNN	Rede Neural Convolucional, do inglês <i>Convolutional Neural Networks</i>
CTV	<i>Clinical Target Volume</i>
CUBE	<i>Fast Spin Echo</i> com ângulo de rotação variável ponderada em T_1 da <i>General Electric</i>

DICOM	<i>Digital Imaging and Communications in Medicine</i>
DL	Aprendizado profundo, do inglês <i>Deep Learning</i>
DSC	<i>Dice Similarity Coefficient</i>
DVH	Histograma Dose-Volume, do inglês <i>Dose Volume Histogram</i>
DWI	<i>Diffusion-Weighted Imaging</i>
FCN	<i>Fully Connected Network</i>
FLAIR	Recuperação da Inversão Atenuada por Fluidos ponderada em T_2
FP	Falso Positivo
FPP	Falso Positivo por Paciente
FN	Falso Negativo
GK	<i>Gamma Knife</i>
GPU	<i>Graphics Processing Unit</i>
GRE	Gradiente Eco
GT	<i>Ground Truth</i>
GTV	<i>Gross Target Volume</i>
HD	<i>Hausdorff Distance</i>
IA	Inteligência Artificial
ICRU	<i>International Commission on Radiation Units and Measurements</i>
IGRT	Radioterapia guiada por imagem, do inglês <i>Image Guided Radiation Therapy</i>
IoU	<i>Intersection over Union</i>
IM	Margem Interna, do inglês <i>Internal Margin</i>
IMRT	Radioterapia com intensidade modulada, do inglês <i>Intensity Modulated Radiation Therapy</i>
ITV	<i>Internal Target Volume</i>
LR	Taxa de aprendizagem, do inglês <i>Learning Rate</i>
MAE	Erro absoluto médio, do inglês <i>Mean Absolute Error</i>

MIP	Projeção de Intensidade Máxima, do inglês <i>Maximum Intensity Projection</i>
ML	Aprendizado de Máquina, do inglês <i>Machine Learning</i>
MPR	Reconstrução Multi-planar, do inglês <i>Multi Planar Reconstruction</i>
NIFTI	<i>Neuroimaging Informatics Technology Initiative</i>
OAR	Órgão de Risco, do inglês <i>Organ at Risk</i>
PET	Tomografia por emissão de pósitrons, do inglês <i>Positron Emission Tomography</i>
PPGTIGS	Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde
PPV	Precisão, do inglês <i>Positive Predicted Value</i>
PTV	<i>Planning Target Volume</i>
RGB	Canais <i>Red, Green, and Blue</i>
RM	Ressonância Magnética
RMSProp	<i>Root Mean Squared Propagation</i>
ROI	Região de interesse, do inglês <i>Region of Interest</i>
RT	Radioterapia
SM	Margem de Setup, do inglês <i>Setup Margin</i>
SP	Sequência de Pulso
SPACE	Sequência <i>Fast Spin Echo</i> ponderada em T1
SPGR	<i>Spoiled Gradient Echo</i>
SRS	Radiocirurgia, do inglês <i>Stereotactic Radiosurgery</i>
SSD	<i>Single-shot detector</i>
swDSC	Coefficiente Dice médio por <i>slice</i> , do inglês <i>slice-wise Dice</i>
TC	Tomografia Computadorizada
TCIA	<i>The Cancer Imaging Archive</i>
TN	Verdadeiro Negativo, do inglês <i>True Negative</i>

TP	Verdadeiro Positivo, do inglês <i>True Positive</i>
TPR	Sensibilidade, do inglês <i>True Positive Rate</i>
TPS	Sistema de planejamento radioterápico, do inglês <i>Treatment Planning System</i>
T/V/T	Divisão de treino, validação e teste;
UFCSPA	Universidade Federal de Ciências da Saúde de Porto Alegre
VGG	<i>Visual Geometry Group</i>
VMAT	Arcoterapia Volumétrica Modulada, do inglês <i>Volumetric Modulated Arc Therapy</i>
WBRT	Radioterapia Cerebral Total, do inglês <i>Whole Brain Radiation Therapy</i>

Sumário

1 – Introdução	1
2 – Objetivos	5
2.1 Objetivo Geral	5
2.2 Objetivos Específicos	5
3 – Referencial Teórico	6
3.1 Metástases Cerebrais e técnicas de tratamento	6
3.2 Delineamento de metástases cerebrais em imagens de RM	8
3.3 Deep Learning	11
3.3.1 Redes Neurais Convolucionais	12
3.3.2 <i>Transfer Learning</i>	15
3.3.3 Redes de segmentação	16
3.3.3.1 U-Net	16
3.3.3.2 SegNet	16
3.3.3.3 DeepMedic	17
3.3.3.4 GoogLeNet	17
3.3.3.5 HRNetV2	17
3.4 Métricas de Avaliação de performance	18
3.4.1 Precisão	18
3.4.2 Sensibilidade ou <i>Recall</i>	18
3.4.3 Acurácia	19
3.4.4 Área sob a curva ROC (AuROCC)	19
3.4.5 <i>F1-Score</i>	19
3.4.6 Coeficiente Dice (DSC)	19
3.4.7 Intersecção sobre a união (IoU)	20
3.4.8 Distância de Hausdorff (HD)	20
3.5 Funções de Perda	21
3.5.1 <i>Cross-Entropy Loss</i>	22
3.5.2 <i>Balanced Cross-Entropy Loss</i>	23
3.5.3 <i>Dice Loss</i>	23
3.5.4 <i>Generalized Dice Loss</i>	24
3.5.5 <i>Binary Focal Loss</i>	24
3.5.6 <i>Boundary Loss</i>	25
3.5.7 <i>Tversky Loss</i>	25

3.6	Recomendações para publicação de trabalhos relacionados à Física Médica com uso de Inteligência Artificial	26
3.6.1	CLAMP	26
3.6.2	CLAIM	28
4	–Revisão da Literatura	32
5	–Materiais e Métodos	52
5.1	Bancos de dados	52
5.1.1	<i>BrainMetShare</i> (BMS)	52
5.1.2	<i>The Cancer Imaging Archive</i> (TCIA)	53
5.1.3	Ocaña-Tienda <i>et al.</i> (2023)	54
5.2	Comparação entre os bancos de dados	55
5.3	Pré-Processamento	57
5.3.1	BrainMetShare (BMS)	57
5.3.2	Ocaña-Tienda	58
5.3.3	<i>The Cancer Imaging Archive</i> (TCIA)	58
5.4	Modelos de Segmentação	59
5.4.1	Met-Seg	60
5.4.2	AURORA	60
5.4.3	Raidionics	61
5.4.4	Proposição de um modelo de segmentação	62
5.5	Análise e Validação dos Modelos	65
6	–Resultados e Discussão	67
6.1	Algoritmo de contagem de metástases	67
6.2	Análise dos modelos da literatura	71
6.2.1	Modelo MetSeg, desenvolvido por Ottesen <i>et al.</i> (2023)	71
6.2.2	Modelo AURORA, desenvolvido por Buchner <i>et al.</i> (2023)	79
6.2.3	Modelo Raidionics, desenvolvido por Bouget <i>et al.</i> (2023)	82
6.2.4	Análise dos <i>checklists</i> CLAMP e CLAIM para as publicações envolvendo os modelos da literatura	83
6.2.5	Comparação dos resultados obtidos nos três modelos da literatura	85
6.3	Proposição de um modelo para segmentação de metástases cerebrais	94
6.3.1	Modelo para avaliação das melhores sequências de pulso para identificação de Metástases Cerebrais	94
6.3.2	Avaliação das funções de perda do modelo de segmentação	98
6.3.2.1	<i>Binary Cross Entropy</i>	98
6.3.2.2	<i>Binary Balanced Focal Cross Entropy</i>	99
6.3.2.3	Tversky	99

6.3.2.4	BCE + Dice + <i>Boundary</i>	99
6.3.2.5	BCE + Dice + HD	100
6.3.2.6	BCE + Tversky ponderada	100
6.3.2.7	BCE + Tversky ponderada + <i>Boundary</i>	104
6.3.3	Comparação dos resultados do modelo desenvolvido com os modelos de literatura	108
7	– Conclusão	110
	Referências	112
	Apêndices	125
	APÊNDICE A – Contagem de <i>clusters</i> em Python	126

1 Introdução

As metástases cerebrais múltiplas são os tumores intracranianos malignos mais comuns que acometem adultos, com uma incidência de 20% entre todos os pacientes com câncer (CHO et al., 2020; LIN; DEANGELIS, 2015; ACHROL et al., 2019). Esse tipo de neoplasia é decorrente de outros tumores em outras localidades do corpo, uma vez que as células tumorais podem ser conduzidas via corrente sanguínea e sistema linfático para o sistema nervoso central, permitindo que tais células adentrem a barreira hematoencefálica, dando início à metástase (ACHROL et al., 2019; AMSBAUGH; KIM, 2023). Um estudo realizado em 2005, (SHAW et al., 2005) informa que o tempo de sobrevivência de pacientes diagnosticados com esse tipo de câncer era, em média, de 4 a 6 meses após tratamento radioterápico. Em menos de duas décadas, os avanços nas formas de diagnóstico e tratamento elevaram esse tempo para até 2 anos, dependendo do tipo de tumor, como mostra o estudo feito por (ACHROL et al., 2019) que também indica que fatores prognósticos como sexo, idade e histórico patológico influenciam diretamente no tempo de sobrevida.

O tratamento das metástases cerebrais pode ser realizado através da ressecção cirúrgica, quando o número de metástases é pequeno e estão localizadas em regiões de fácil acesso, a radioterapia de cérebro inteiro (WBRT, do inglês *Whole Brain Radiation Therapy*), quando o status de bem-estar do paciente avaliado pelo médico responsável, for baixo ou se a quantidade de metástases for muito alta. Alternativamente, a radiocirurgia estereotáxica (SRS, do inglês *Stereotactic Radiosurgery*) é recomendada quando o número de metástases estiver entre 1 e 4, podendo chegar a 10 em alguns casos, quando o status de bem-estar do paciente for normal (AMSBAUGH; KIM, 2023). Outras modalidades de terapia como a quimioterapia ou a imunoterapia possuem eficácias limitadas, devido à dificuldade de agentes anticancerígenos em penetrar a barreira hematoencefálica, e podem ser utilizadas em conjunto com a radioterapia a depender do contexto do paciente e da origem do tumor primário (TOZUKA et al., 2024; LI et al., 2024).

A recorrência metastática após a ressecção cirúrgica é relativamente alta, podendo chegar a 43% em um período de 12 meses, o que pode ser melhor controlado aplicando-se as técnicas de radioterapia (RT) após a ressecção. No entanto, a aplicação dessas técnicas podem afetar negativamente algumas funções neurocognitivas (AMSBAUGH; KIM, 2023).

A SRS é uma técnica de radioterapia que se provou com menor potencial de danos após o tratamento (TSAO et al., 2012), podendo ser benéfica para o controle tumoral e para manter a função cognitiva dos neurônios, além de ser uma das técnicas mais utilizadas para tratamentos de neoplasias cranianas múltiplas, juntamente com craniotomia, quimioterapia ou WBRT. O processo da SRS começa na aquisição de imagens médicas,

que formarão um molde 3D no sistema de planejamento (TPS, do inglês *Treatment Planning System*), onde serão elaborados as melhores condições da entrega de uma alta dose de radiação aos volumes alvo, geralmente realizada em uma única entrega, e tendo em consideração as estruturas circunvizinhas às regiões irradiadas (BIBAULT et al., 2021).

De acordo com a *International Commission on Radiation Units and Measurements* (ICRU) (ICRU, 2015), em um planejamento radioterápico, busca-se realizar o delineamento de uma região tumoral nas imagens em Ressonância Magnética (RM), Tomografia Computadorizada (TC) ou na fusão de ambas. Em algumas situações é interessante utilizar os dois tipos de modalidades de imagens que são complementares uma à outra, pois a RM possui uma capacidade maior de diferenciar tecido nervoso de estrutura cerebral, porém está sujeita a um maior número de artefatos e distorções espaciais, enquanto que a TC possui uma excelente resolução espacial e mede diretamente as densidades eletrônicas para os cálculos de dose do planejamento radioterápico, porém possui baixo contraste nas regiões de tecidos moles (HU et al., 2019).

Dentro do planejamento radioterápico, a região tumoral a ser delineada para tratamento é composta por quatro volumes de características distintas: o *Gross Tumour Volume* (GTV), o *Clinical Target Volume* (CTV), o *Internal Target Volume* (ITV) e o *Planning Target Volume* (PTV), sendo essa ordem indo do menor volume para o maior, ou seja, cada volume maior engloba o volume menor. O GTV é o volume do tumor identificado, o CTV é o volume que engloba o tumor e considera patologias subclínicas em torno do GTV, o ITV leva em consideração a variação da posição espacial do CTV pela movimentação do paciente, por exemplo pela respiração ou diferença no posicionamento, adicionando uma margem interna (IM, do inglês *Internal Margin*) ao CTV, e o PTV é o volume do planejamento considerando a margem de *Setup* (SM, do inglês *Setup Margin*) adicionada ao ITV, e que correspondente às variações dadas pelo equipamento ou geometrias utilizadas (MACKIE; GREGOIRE, 2011), sendo esse o volume de tratamento.

As definições para as margens IM e SM são padronizadas dependendo do tipo de órgão, do tipo de modalidade de imagem e do tipo da neoplasia a ser tratada, essas margens são usualmente definidas com valores entre 1 milímetro a 1,5 centímetro, sendo a SM levemente maior do que a IM (LEE, 2013). Além desses volumes de tratamentos, outras regiões de interesse para o tratamento (ROIs, do inglês *Regions of Interest*) são propostas, como os órgãos de risco (OARs, do inglês *Organs at Risk*), que são delineados de modo a quantificar a absorção de dose de radiação por OARs saudáveis e responsáveis por funções vitais, tendo em vista que suas atividades fisiológicas podem sofrer alteração após o tratamento.

O delineamento dessas regiões, na maioria dos casos, é realizado manualmente e demanda um trabalho extensivo e sujeito a variações entre radiologistas. Esse delineamento é geralmente feito em imagens de TC utilizando o *software* do TPS e consiste no

contorno manual utilizado uma ferramenta de lápis ou pincel, sendo realizado *slice* por *slice* da imagem. Há a opção de interpolação entre *slices* não consecutivos para o preenchimento automático dos *slices* intermediários, para evitar que o profissional precise realizar o delineamento em todos os *slices* da imagem. Esse é um trabalho que demanda tempo e pode ser facilitado com o uso de ferramentas de IA para o delineamento automático. Além disso, a diferenciação de metástases cerebrais de outras patologias é uma tarefa complexa, já que os achados radiológicos entre elas são parecidos. Na Figura 1, estão alguns exemplos de 5 outras patologias que também demonstraram hipersinal em forma de anel ao redor da lesão, em uma sequência ponderada em T1 após aplicação de contraste de Gadolínio. Com isso em vista, técnicas de Inteligência Artificial (IA) podem auxiliar no delineamento como uma ferramenta de apoio à tomada de decisão do especialista.

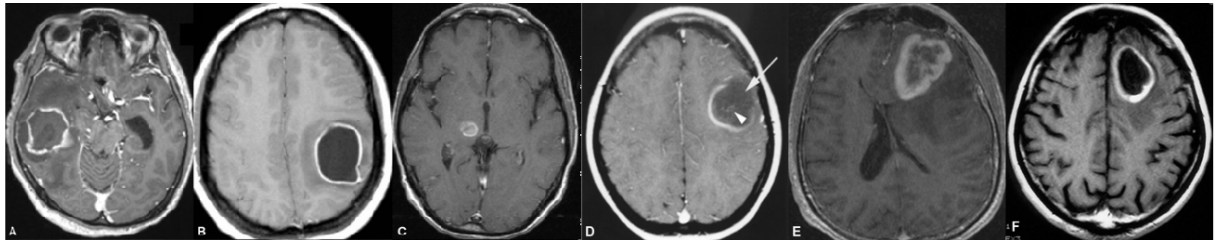


Figura 1 – Imagens de RM ponderadas em T1 com contraste em hipersinal anelar para casos de (A) toxoplasmose, (B) neurocisticercose, (C) glioblastoma, (D) esclerose múltipla tumefativa, (E) linfoma e (F) metástase de tumor pulmonar.

Fonte: Adaptado de (POPE, 2018)

O *Deep Learning* (DL) ou Aprendizado Profundo, é uma subárea do *Machine Learning* (ML) ou Aprendizado de Máquina, que utiliza camadas sucessivas de operações matemáticas cujos valores são atualizados conforme é realizado o treinamento do modelo. As Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) são um dos diversos tipos de redes de aprendizado profundo, que utilizam operações de convoluções em sub-regiões de uma imagem, extraindo características de imagens (CHOLLET, 2017). Esse tipo de tecnologia pode ser uma alternativa para realizar o delineamento das metástases cerebrais, pois permite que o delineamento seja feito mais rapidamente, além de oferecer uma segunda checagem para identificar o volume a ser delineado.

Algumas pesquisas envolvendo a aplicação de *Deep Learning* para a segmentação de estruturas cerebrais podem ser utilizadas no delineamento do planejamento radioterápico. Apesar dos trabalhos demonstrarem resultados promissores, é preciso uma avaliação mais aprofundada sobre o potencial de utilização destes modelos na prática. Com ênfase na aplicação de segmentação de metástases cerebrais, por exemplo, alguns estudos apontam para a dificuldade em se detectar metástases pequenas (GRØVIK et al., 2020; BIBAULT et al., 2021; CHARRON et al., 2018) quando o modelo também é treinado para metástases maiores, o que pode dificultar a definição de um modelo mais geral. No entanto, o tamanho mínimo a partir do qual uma metástase é considerada pequena não

é um consenso entre os autores, tampouco a unidade de medida utilizada para essa avaliação, como por exemplo o tamanho de $1,5 \text{ cm}^3$ por Liu *et al.* (2017), 2 mLs por Bouget *et al.* (2023), 5 mm por Park *et al.* (2021), 7 mm para Grøvik *et al.* (2020), diâmetro de 8 mm por Charron *et al.* (2018), 10 mm por Yang *et al.* (2023) ou 1,16 mLs para Buchner *et al.* (2023). Além dessas divergências, não são todos os autores que informam se o tamanho se trata de um raio, um diâmetro ou outra distância. Alguns treinos realizados apenas com metástases menores melhoram a eficiência de detecção, como é o caso do estudo de Bousabarah *et al.* (2020) que treinou o modelo apenas com metástases de tamanhos menores do que 0,4 mLs e apresentou uma sensibilidade de detecção de 97% para metástases maiores do que 0,06 mL, enquanto que o modelo de Grøvik *et al.* (2020) teve uma sensibilidade de 50% para metástases menores do que 7 mm. No entanto, esses modelos precisam passar por refinamentos e seus resultados não são satisfatórios quando comparados aos resultados obtidos para metástases maiores. Além disso, esses modelos podem apresentar resultados promissores para um banco de dados em específico, e ter uma performance pior para outros bancos de dados.

Sendo assim, o questionamento que guia este trabalho é como a utilização de *Deep Learning* pode auxiliar no delineamento automático de metástases cerebrais? Para responder a essa pergunta, foi feita uma análise crítica de modelos da literatura em diferentes bancos de dados e, adicionalmente, foi desenvolvido um modelo de redes neurais para a mesma tarefa. Dessa forma, espera-se que com os resultados obtidos neste estudo, os tratamentos de metástases cerebrais utilizando radioterapia tenham uma ferramenta embasada em *Deep Learning* para auxiliar as equipes médicas durante o planejamento radioterápico, com ênfase na etapa de identificação e segmentação das regiões metastáticas, principalmente aquelas menores e de difícil identificação devido ao tamanho, uma vez que em muitos casos o processo de delineamento é demorado, podendo ser acelerado com a utilização de *Machine Learning*.

2 Objetivos

2.1 Objetivo Geral

Avaliar modelos de aprendizado profundo para automatizar o delineamento de metástases cerebrais múltiplas com imagens de ressonância magnética.

2.2 Objetivos Específicos

- I Avaliar diferentes bancos de dados públicos contendo imagens de RM para pacientes diagnosticados com metástases cerebrais.
- II Levantamento e seleção de modelos de aprendizado profundo disponíveis na literatura para o delineamento de metástases cerebrais em imagens de RM.
- III Avaliação da qualidade dos modelos selecionados através de indicadores convencionais do ramo de *Deep Learning* utilizando diferentes bancos de dados.
- IV Avaliação do potencial de aplicação dos modelos selecionados considerando aspectos relevantes para a prática clínica.
- V Avaliação dos modelos selecionados frente os critérios estabelecidos no *Checklist for Artificial Intelligence in Medical Imaging* (CLAIM) e no *Checklist for AI in Medical Physics* (CLAMP).
- VI Proposição de um modelo de aprendizado profundo para o delineamento de metástases cerebrais.
- VII Avaliar concordância da segmentação de metástases cerebrais obtidas com redes neurais em relação aos delineamentos realizados por profissionais da área da saúde.

3 Referencial Teórico

Neste capítulo, serão apresentados alguns tópicos centrais do trabalho, começando por uma breve explicação sobre o que são as metástases cerebrais e quais são as técnicas de tratamento, como são realizados os delineamentos nas imagens em RM, uma breve introdução ao *Deep Learning* com ênfase nas características trabalhadas nesse estudo, nas métricas de avaliação utilizadas e uma descrição das funções perdas estudadas no desenvolvimento do modelo. Essas funções perda fazem parte de um modelo de *Deep Learning* e o motivo para não apresentar as funções perda dentro da seção de *Deep Learning*, foi que alguns conceitos centrais para seu entendimento serão introduzidos na seção das métricas de avaliação de performance e, por essa razão, foi colocada após a seção das métricas de avaliação de performance. Ao final do capítulo, serão apresentados brevemente os *checklists* CLAMP e CLAIM que compõem as recomendações para publicação de trabalhos relacionados à Física Médica com uso de Inteligência Artificial.

3.1 Metástases Cerebrais e técnicas de tratamento

As metástases cerebrais são o tipo de tumor intracraniano mais comum (ACHROL et al., 2019; AMSBAUGH; KIM, 2023; CHO et al., 2020; LIN; DEANGELIS, 2015), com ocorrência na faixa de 10% a 26% de todos os óbitos por câncer (AMSBAUGH; KIM, 2023). As células metastáticas mais recorrentes são provenientes de câncer de pulmão, mama e melanoma, mas também podem ser transportadas de tumores na próstata, cabeça e pescoço (AMSBAUGH; KIM, 2023).

Esse transporte de uma célula metastática ocorre nos sistemas sanguíneo ou linfático, através da quebra da barreira hematoencefálica, levando à proliferação dessas células que podem causar invasão de regiões cerebrais, deslocamento de estruturas cerebrais, inflamações e edemas. Os locais de acúmulo no cérebro dependem do subtipo histológico da célula, e também se há alto fluxo sanguíneo na região (AMSBAUGH; KIM, 2023).

Em casos de suspeita de ocorrência de metástase, o histórico do paciente é importante, e sintomas como dores de cabeça, visão embaçada e náusea podem ser indícios da presença metastática. Além disso, testes físicos devem ser realizados de modo a se avaliar a força, sensações, coordenação, reflexos, função cerebral, propriocepção, função dos nervos cranianos, fala, pensamento, visão e memória. Outras informações que podem ser relevantes são a idade, status de bem-estar e da carga do câncer sistêmico (AMSBAUGH; KIM, 2023). Outros tipos de exames que podem ser realizados incluem a contagem de hemácias, o painel metabólico e análise da função do fígado (AMSBAUGH; KIM, 2023), cujas informações podem ser utilizadas como radiômica para complementar a análise feita

pelas imagens (BIBAULT et al., 2021).

Após estadiamento do câncer metastático do paciente e verificada a necessidade de submissão à radioterapia, são obtidas imagens médicas em TC e é definida a modalidade da radioterapia para o planejamento radioterápico. As técnicas da radioterapia utilizadas para o tratamento de metástases cerebrais podem variar de acordo com o número, volume e localização das metástases, o status de bem-estar do paciente, a idade, presença de doença extracraniana e doença leptomeníngea, localização do tumor primário, local da ressecção e a preferência do paciente (ARORA et al., 2023). Uma das diferentes técnicas de radioterapia utilizadas para o tratamento de metástases cerebrais é a radioterapia cerebral total (WBRT, do inglês *Whole Brain Radiation Therapy*), que é uma modalidade utilizada no tratamento de metástases cerebrais que irradia todo o cérebro com doses totais de 20 a 30 Gy, distribuídas em sessões que podem ir até 10 (DESAI; RICH, 2020; ARORA et al., 2023). Outra forma é através da radioterapia com intensidade modulada (IMRT, do inglês, *Intensity Modulated Radiation Therapy*), que é uma modalidade de radioterapia que envolve o uso de feixes de fótons provenientes de aceleradores lineares em diferentes intensidades e formatos, de modo a entregar a dose na região desejada que geralmente fica localizada na intersecção das trajetórias dos feixes. Já a Arcoterapia Volumétrica Modulada (VMAT, do inglês *Volumetric Modulated Arc Therapy*) ou *RapidArc*, é um tipo de tratamento que entrega a dose em feixes volumétricos em arco, através da rotação do acelerador linear ao redor da região a ser tratada sendo mais rápida e apresenta *falloff* de dose parecidos com a radiocirurgia (BALLANGRUD et al., 2018). Por fim, a radiocirurgia estereotáxica (SRS, do inglês *Stereotactic Radiosurgery*) é uma técnica de radioterapia que envolve a aplicação única e bem localizada de uma alta dose de radiação ionizante na região tumoral a ser tratada, dois dos aparelhos mais comuns são o *Gamma Knife* (GK) e o *Cyber Knife* (CK), sendo o primeiro baseado em radiações gama oriundas de fontes de Cobalto-60 e o segundo em raios-X de um acelerador linear acoplado a um braço mecânico.

Para o caso da radiocirurgia, a precisão de localização da região a ser tratada é importante para o tratamento mais acurado, sendo possível fazer uso da combinação das imagens em RM e tomografia por emissão de pósitrons (PET, do inglês *Positron Emission Tomography*) para se obter uma melhor resolução espacial e acurácia na determinação das regiões para o tratamento, já que a imagem da RM é utilizada para se obter informações anatômicas enquanto que a imagem do PET é utilizada para se obter informações metabólicas, permitindo a diferenciação de uma recorrência tumoral de uma necrose (MORI et al., 2006) ou de alguma alteração após um primeiro tratamento (CASTELLANO et al., 2021). Além disso, a partir da imagem de RM não é possível saber se o acúmulo de contraste foi dado pelo aumento da permeabilidade da barreira hematoencefálica ou se há um tumor na região. Por esse motivo, técnicas mais recentes em RM podem ser mais eficazes na detecção tumoral, como a espectroscopia por RM, RM com difusão e RM

com perfusão cerebral, bem como a utilização em conjunto com as imagens de PET para possibilitar a diminuição da região a ser delineada no planejamento radioterápico (CASTELLANO et al., 2021). Essas técnicas, no entanto, carregam informações prévias do estado do paciente, que podem estar alteradas no dia do tratamento. Visando contornar essa limitação da precisão da localização em tempo real, o uso de aparelhos de tomografia computadorizada de feixe cônico (CBCT, do inglês *Cone Beam Computed Tomography*) podem auxiliar a guiar o posicionamento do paciente no momento da aplicação do feixe de tratamento.

3.2 Delineamento de metástases cerebrais em imagens de RM

A identificação de metástases cerebrais é geralmente realizada através de imagens de RM com aplicação de contraste de gadolínio, que permite determinar o número de metástases e as regiões anatômicas em que se encontram, além de determinar o grau do edema relacionado ao tumor (AMSBAUGH; KIM, 2023). No entanto, para a RT, é viável utilizar uma imagem de TC devido à maior precisão espacial do tumor, uma vez que a imagem da RM está sujeita à artefatos e distorções de reconstrução (HU et al., 2019), sendo também importante para a dosimetria por ser possível determinar os coeficientes de atenuação dos diferentes tecidos através dos os valores para os números de Hounsfield obtidos com a imagem de TC (JAYAMANI et al., 2019). Atualmente, nos sistemas de planejamento, o delineamento é realizado utilizando-se ferramentas de desenho, como o lápis ou o pincel, como ilustrado na Figura 2 e, como o planejamento radioterápico se faz através do uso de altas doses que podem gerar toxicidades, é necessário que esse delineamento seja o mais preciso possível. Por isso, o uso da combinação entre TC e RM têm se mostrado bastante eficaz (DEMIRAL et al., 2019).

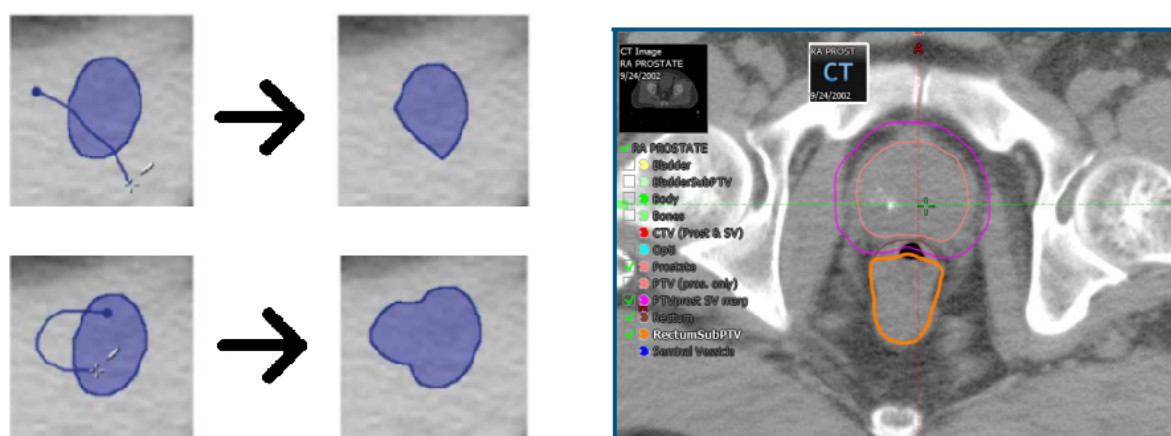


Figura 2 – Exemplo do uso da ferramenta de desenho à mão livre do TPS Eclipse da Varian.

Fonte: Adaptado de (VARIAN, 2014) e (VARIAN, 2022)

As imagens são usualmente armazenadas em formato *Digital Imaging and Communications in Medicine* (DICOM), que é o formato padrão para o intercâmbio de dados biomédicos (BIDGOOD JR et al., 1997), e conta com o armazenamento de informações não apenas referentes à imagem, mas também dados do paciente e do equipamento utilizado para realizar o exame. Existem diversas bibliotecas em linguagem *Python* específicas para realizar a leitura e manuseio de dados no formato DICOM, como a *Pydicom* (MASON, 2011) ou a recente *Imagedata* (ANDERSEN, 2022). Na próxima seção, encontra-se um panorama geral sobre as imagens em RM para a identificação de metástases cerebrais.

As imagens médicas obtidas em Ressonância Magnética podem ser adquiridas em três ponderações, obtidas pela combinação de diferentes tempos de eco e de repetição do pulso de radiofrequência do aparelho, que são as ponderações T1, T2 e densidade de prótons (MAZZOLA, 2015). Além disso, podem ser utilizadas diferentes Sequências de Pulso (SP), que podem variar de acordo com a fabricante do aparelho de RM. Cada tipo de modalidade possui parâmetros de aquisição distintos e trazem informações estruturais complementares sobre a região imageada, como é o caso da sequência em ponderação T2 denominada FLAIR, do inglês *Fluid Attenuated Inversion Recovery*, que atenua o sinal do líquido cefalorraquidiano, aumenta o contraste entre a substância branca e cinzenta (BANGERTER et al., 2006) e permite determinar regiões com lesões com mais precisão, apesar de também ser mais suscetível à artefatos de imagem (OKUDA et al., 1999).

A ponderação T1 pode ser realizada em SP Spin Eco ou Gradiente Eco (GRE), sendo que essa segunda permite a aplicação de contraste formado por elementos paramagnéticos, como o Gadolínio, e possibilita a verificação de diferenças entre as imagens obtidas antes e após a aplicação do contraste, que normalmente se acumulam em regiões com edema, rompimento da barreira hematoencefálica ou necrose (KINOSHITA et al., 2005; MEHRABIAN et al., 2019). Entre as sequências de pulso em ponderação T1, destacam-se a *Cube*, com e sem contraste, e a BRAVO, do inglês *Brain Volume*, para as imagens utilizadas nessa pesquisa. Por outro lado, o hipsinal na SP em T2 FLAIR permite identificar as regiões com edema vasogênico próximas aos tumores (MEHRABIAN et al., 2019), que está correlacionado tanto com o tamanho dos tumores quanto com o local do tumor primário. Esse edema ocorre preferencialmente na região da substância branca, por ter uma distribuição radial de fibras nervosas, e como depende do tamanho da metástase, o hipsinal não deve ser usado como regra para identificação, principalmente para o caso de metástases pequenas (SCHNEIDER et al., 2017). Um exemplo das diferenças entre as sequências pode ser visto na Figura 3, onde também é possível fazer a comparação com imagens em TC. Nota-se a importância do contraste para a identificação de hipsinal em anel ao redor das massas tumorais, e o hipsinal na sequência FLAIR indicativo de edema vasogênico.

As imagens de RM com contraste são recomendadas para identificação de metás-

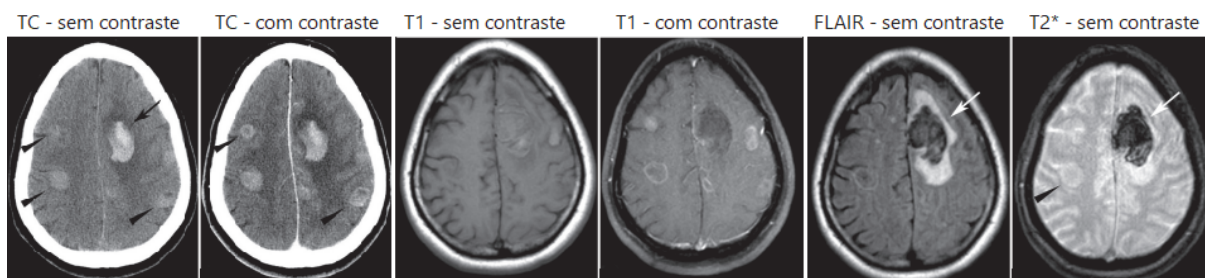


Figura 3 – Diferentes modos de aquisição de imagem de metástases cerebrais. TC - tomografia computadorizada. T1, FLAIR e T2* - sequências de pulso para imagens de ressonância magnética.

Fonte: Adaptado de (FINK; FINK, 2013)

tases cerebrais, e provaram-se superior às imagens obtidas em TC com contraste ou RM sem contraste. No entanto, é preciso cuidado ao se administrar doses de contraste maiores do que as recomendadas, pois apesar de melhorar o diagnóstico de pequenas metástases, o número de falsos positivos também aumenta, além do risco de desenvolvimento de fibrose nefrogênica sistêmica (ZAKARIA et al., 2014).

Entre as sequências de pulso mais utilizadas, destacam-se a *Spoiled Gradient Echo* (SPGR) e a FLAIR para diferenciar anormalidades menores do que 3 milímetros (ZAKARIA et al., 2014). Já a ressonância de difusão (DWI, do inglês *Diffusion-Weighted Imaging* com o mapa de coeficiente de difusão aparente (ADC, do inglês *Apparent Diffusion Coefficient*), permite diferenciar tumores de abscessos, esses que restringem a difusão de água. Um outro fator que deve ser levado em consideração é que depois que o tipo de tratamento é determinado pela equipe médica, novas imagens podem ser adquiridas para uma melhor resolução espacial.

A Figura 4 apresenta hipersinal em T1 com contraste em forma de anel em torno do tumor, o que pode indicar metástase, glioma de alto grau ou abscesso. O hipossinal (mais escuro) em DWI pode ser um indício de edema vasogênico, que em teoria é mais difusivo ao redor de metástases do que de gliomas de alto grau, que possuem mais células em torno do tumor (ZAKARIA et al., 2014). Quando a restrição da difusão se apresenta em hipersinal em DWI e hipossinal no mapa ADC, pode ser um indício de coexistência de metástase e abscesso ou necrose com edema intracelular (YOUSEF et al., 2014).

O sinal encontrado para metástases não possui um padrão bem definido (ZAKARIA et al., 2014; YOUSEF et al., 2014) e varia de acordo com o local do tumor primário. Um exemplo é a metástase de melanoma, cuja melanina possui propriedades paramagnéticas e aparece em hipersinal em T1 sem contraste e hipossinal em T2 FLAIR. De modo geral, tem-se que o sinal é isointenso ou hipointenso em T1, e hiperintenso em T2, com alterações em casos de hemorragia e para casos de metástases provindas de melanomas. Ademais, o contraste é tipicamente hiperintenso, podendo ter um formato uniformemente distribuído na região tumoral, também podendo ser pontual ou ao redor do tumor, em

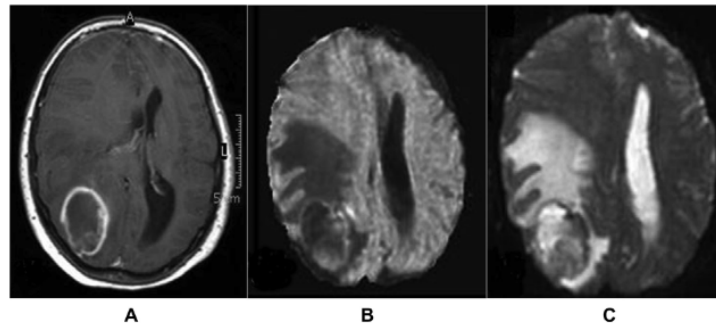


Figura 4 – Imagens de RM ponderadas em (A) T1 com contraste de gadolínio, (B) DWI sem contraste e (C) mapa ADC sem contraste em caso de metástase cerebral única de tumor primário de mama em paciente mulher de 44 anos.

Fonte: (YOUSEF et al., 2014)

forma de anel (FINK; FINK, 2013).

Um estudo piloto realizado por Yoon *et al.* (2018) para a comparação de diferentes sequências de pulso na detecção de metástases cerebrais indicou que o uso da sequência T1 CUBE, que tem como base a sequência *fast spin echo* 3D com ângulos de inversão de refocagem modulados, um esquema de imagens paralelas no espaço híbrido com autocalibração e um esquema de ordenação de visualização otimizado para a grade $k_y - k_z$ (STANLEY et al., 2009), sendo essa a modalidade que apresenta maior sensibilidade (aumento de 5%) e reduzidos falsos positivos e falsos negativos quando comparada com a sequência BRAVO, em um estudo com dois neurorradiologistas com mais de 10 anos de experiência (YOON et al., 2018).

3.3 Deep Learning

Deep Learning (DL) é um ramo de *Machine Learning* que se utiliza de redes neurais com camadas profundas para a determinação de regras aplicadas a um conjunto de dados através do fornecimento dos dados de entrada, ou *inputs*, e as respostas que a rede deve alcançar.

Em um modelo de *Deep Learning*, alguns objetos são fundamentais para a sua construção, como as camadas, os dados de entrada ou *inputs*, a função perda, também chamada de função de custo, *loss function* ou simplesmente *loss*, e o otimizador. A relação entre esses objetos pode ser visto na Figura 5, e uma explicação mais detalhada sobre as funções perda serão apresentadas na seção após as métricas de avaliação.

A utilidade do DL aplicado na RT é diversa, tendo funções como o aprimoramento da reconstrução da imagem, remoção de ruídos, segmentação automática de estruturas, adequação às recomendações de delineamento, dosimetria, previsão de resposta ao tratamento, previsão de toxicidades, entre outros (BIBAULT et al., 2021). No âmbito da segmentação automática, utiliza-se a tecnologia de Visão Computacional através das re-

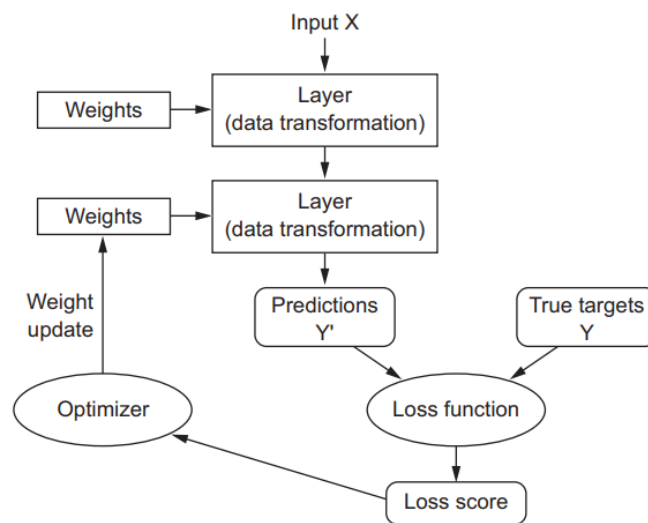


Figura 5 – Relação entre os objetos que compõem um modelo de DL.

Fonte: (CHOLLET, 2017)

des de aprendizado profundo, que podem ser de três tipos: as redes convolucionais, as redes *Deep Belief* e as máquinas de Boltzmann profundas (VOULODIMOS et al., 2018). Nesse estudo, serão aplicadas redes neurais convolucionais, e um detalhamento maior sobre essas redes está descrito na subseção a seguir.

3.3.1 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNNs), do inglês *Convolutional Neural Networks*, foram inspiradas pela estrutura do sistema visual biológico, que processa o *input* em um padrão de grade de modo a adquirir características espaciais sobre as bordas, texturas, contraste da imagem, entre outras características (YAMASHITA et al., 2018). Uma CNN é constituída por três tipos de camadas principais: a camada convolucional, a camada de *pooling* e a camada *fully connected*, conforme ilustrado na Figura 6.

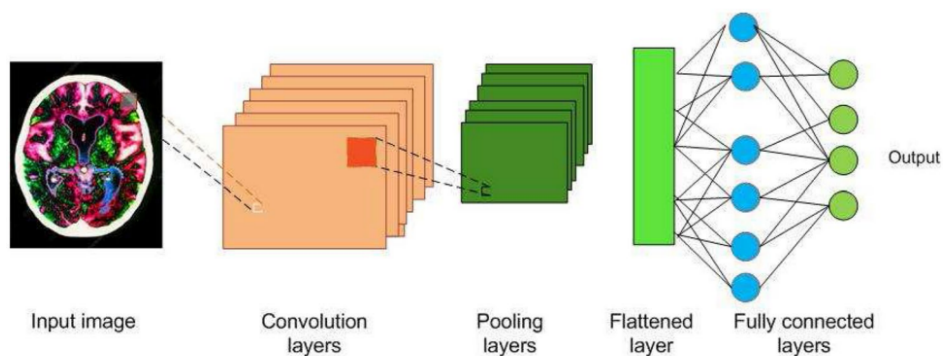


Figura 6 – Representação esquemática dos blocos de uma rede neural convolucional

Fonte: (SARVAMANGALA; KULKARNI, 2022)

A camada de convolução irá extrair algumas características da imagem e formar o que se chama de mapa de características, enquanto que a camada de *pooling* irá reduzir a dimensão espacial da imagem, em um processo que pode ocasionar perda de informação mas que também previne o chamado *overfitting*, que é quando a rede treinada se ajusta perfeitamente aos dados de treino mas sem grandes aplicações para dados que estejam fora dele. Esses mapas de características com redução de dimensionalidade passam por diversas camadas de convolução e *pooling*, a depender da arquitetura selecionada, cuja complexidade de características aumenta a cada camada, até chegar à última camada *fully connected* que agrega todos os valores obtidos das camadas anteriores em um tensor unidimensional, e é responsável em trazer o resultado final da rede (VOULODIMOS et al., 2018; YAMASHITA et al., 2018).

A fins de elucidação, e tendo como base os artigos de (VOULODIMOS et al., 2018; YAMASHITA et al., 2018) e o livro de (CHOLLET, 2017), alguns termos específicos utilizados no contexto das redes neurais convolucionais e que serão mencionados ao longo deste trabalho estão descritos a seguir:

- **Input:** O dado que entra na rede neural, nesse estudo é a sequência de valores da intensidade de *pixels* que compõe uma imagem.
- **Output:** O dado que sai da rede neural, nesse estudo é a máscara de segmentação, constituída de *pixels* com valores 0 e 1. O *pixel* de valor 0 representa a região sem presença de metástase, e o *pixel* com valor 1 representa a região identificada como região tumoral. O valor desses *pixels* pode ser multiplicado por 255 para a representação visual que se diferencia entre preto e branco.
- **Parâmetro:** A variável que está sendo ajustada durante o processo de treinamento da rede neural.
- **Hiperparâmetro:** A variável que o usuário ajusta e que é definida antes do treino ser iniciado e que não se alteram durante o treino, a variação desses valores acarreta resultados distintos para treinos distintos. Os hiperparâmetros influenciam diretamente na qualidade do modelo e consiste no ajuste de valores como o tamanho do *kernel*, o número de *kernels*, o tamanho do *padding*, o tamanho do *stride*, a probabilidade de *Dropout*, os otimizadores utilizados, a *learning rate* (LR), o *batch size* (BS), o número de épocas do treinamento, a utilização de *callbacks* como *model checkpoint*, *early stopping* e *decay*, que, respectivamente, salva o melhor modelo, interrompe o treino caso não haja variações nos resultados e altera dinamicamente os valores de alguns hiperparâmetros para evitar que os pesos do modelo convirjam em mínimos locais, entre outros hiperparâmetros que podem ser definidos antes do treino.

- **Kernel:** É um conjunto de parâmetros variáveis da operação de convolução.
- **Pesos:** Um dos parâmetros que se alteram no treino, dentro do Kernel há diversos pesos e para cada neurônio há um número distinto de pesos associados a cada elemento do *input*.
- **Feature Maps:** Ou mapa de características, é o conjunto de valores adquiridos em uma imagem após a aplicação da convolução. São responsáveis por extrair características locais da imagem, como bordas, linhas, contrastes, entre outras características.
- **Weight Sharing:** Uma das características das CNNs é que os pesos dos kernels são compartilhados para todas as posições da imagem, permitindo extrair características de forma mais generalizada.
- **Época:** É um ciclo do treino, quando todo o conjunto de dados passa ao menos uma vez pelas camadas da rede para a atualização dos valores dos pesos.
- **Batch size:** Uma época é dividida em treinos nos subconjuntos de imagens, e o *batch size* (BS) determina quantos *inputs* serão treinados ao mesmo tempo. O valor do BS pode ir de 1 até o número de amostras totais e, quanto menor seu tamanho, mais passos a rede irá precisar para finalizar uma época.
- **Patch:** É um recorte da imagem original, geralmente utilizado para diminuir o tamanho da imagem treinada, diminuindo a quantidade de poder computacional necessária para o treino, mantendo a resolução da imagem original.
- **Learning rate:** A *learning rate* (LR) é um hiperparâmetro que representa a taxa de aprendizagem da rede, isto é, o valor da atualização da variação dos pesos presentes nos *kernels*.
- **Overfitting:** Representa a condição em que o modelo treinado apresente resultados otimizados para a validação com o conjunto de dados utilizado no treino, porém não apresenta bons resultados quando aplicado em dados externos ao treino.
- **Padding:** É um hiperparâmetro que representa o preenchimento que as redes neurais utilizam para que não haja perda da dimensionalidade do *input*, uma vez que a grade de convolução possui dimensões não unitárias, é preciso preencher as bordas da imagem original para que a nova imagem possua a mesma dimensão.
- **Stride:** É um hiperparâmetro que determina a distância, em *pixels* ou *voxels*, que o *kernel* percorre em cada passo de uma operação de convolução, geralmente seu valor utilizado é unitário.

- **Dropout:** É uma técnica de regularização que desativa alguns dos neurônios da rede de modo a não criar dependências de uma característica com algum neurônio em específico, e permite a atualização mais regular de todos os pesos da rede.
- **Pooling:** São importantes para adicionar uma invariância translacional em relação à pequenas mudanças espaciais entre diferentes amostras, bem como algumas distorções. É uma camada que diminui a dimensionalidade espacial dos mapas de características, e resulta também na diminuição do número de parâmetros de treino.
- **Função perda:** Também chamada de *loss function*, é a função que será utilizada durante o treino para avaliar a eficiência do modelo, quanto menor a função perda, mais o modelo possui previsões que se aproximam do resultado desejado. Para o caso de segmentação, é utilizada a função perda do coeficiente Dice.
- **Função de ativação da última camada:** Essa função é aplicada no *output* da última camada de modo a padronizar o resultado obtido, geralmente é empregada a função *Rectified Linear Units* (ReLU) para a tarefa de classificação, mas também pode ser a função sigmoide ou tangente hiperbólica.
- **Otimizadores:** São algoritmos que permitem a minimização da função perda durante o treino, um dos otimizadores mais utilizados é o Adam (KINGMA; BA, 2014), que tem como base o gradiente de descida estocástica e calcula derivadas de primeira ordem da função perda, outros otimizadores como o Adagrad (DUCHI et al., 2011) e o RMSProp, que é um otimizador não publicado que foi apresentado em uma aula¹ do professor Geoff Hinton da Universidade de Toronto (RUDER, 2016).
- **Data Augmentation:** É um processo de aumento de dados utilizado para treinar o modelo com mais variações de uma imagens, aumentando o tamanho de amostras e permitindo que o modelo fique mais robusto em relação às variações. Geralmente são aplicadas transformações na imagem que envolvem rotação, inversão, ajuste de brilho, recortes, entre outras.

3.3.2 Transfer Learning

O treino de uma rede neural pode demandar muita capacidade computacional e levar de dias a semanas para finalizar. Por isso, uma das estratégias de treino de redes neurais envolve o uso de parâmetros obtidos com um treino prévio, feito no mesmo ou em outro banco de dados, que são reutilizados para uma tarefa semelhante. Geralmente, esses modelos pré-treinados utilizaram bancos de dados maiores e tiveram um maior uso de recursos computacionais, como é o caso dos modelos treinados no ImageNet. Para a tarefa de segmentação de imagens médicas, um estudo feito para investigar as representações

¹ https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

aprendidas com o *Transfer Learning*, concluiu que o ganho nas métricas de coeficiente Dice e distância de Hausdorff foram positivos (KARIMI et al., 2021). Nesse estudo, os autores concluíram que o *transfer learning* pode ser utilizado para um novo conjunto de dados apenas no caso de ambos os bancos de dados tiverem a tarefa em comum, caso contrário a transferência do aprendizado não será estatisticamente significativa (KARIMI et al., 2021). Além disso, para aplicar o *transfer learning*, é necessário que o modelo possua uma arquitetura idêntica ao que foi treinado previamente, para que os parâmetros sejam transferidos neurônio a neurônio, o que impede sua aplicação em modelos com arquiteturas distintas.

3.3.3 Redes de segmentação

A arquitetura da rede neural pode ser alterada e, entre as CNNs utilizadas para realizar a segmentação de estruturas anatômicas, algumas possuem melhor performance para a detecção de metástases cerebrais. Alguns exemplos de diferentes arquiteturas utilizadas em estudos similares a este são a U-Net (RONNEBERGER et al., 2015), a SegNet (BADRINARAYANAN et al., 2017), a DeepMedic (KAMNITSAS et al., 2017) a GoogleNet (GRØVIK et al., 2020) e a HRNetV2 (SUN et al., 2019).

A seguir, encontram-se listadas as principais CNNs estado da arte com uma breve descrição sobre cada uma delas.

3.3.3.1 U-Net

Em 2015, foi apresentado um modelo de rede neural por (RONNEBERGER et al., 2015), que consiste em uma CNN com blocos residuais, e é bastante utilizada na segmentação de imagens médicas para diagnóstico (HU et al., 2019; RUDIE et al., 2021). Essa arquitetura de rede não necessita de grandes conjuntos de dados para realizar um treino com resultados aceitáveis, e utiliza a técnica de *data augmentation* em poucos dados anotados e com camadas de *upsampling* para compensar as camadas de *pooling* realizadas na primeira etapa do processo de treinamento. Essa arquitetura foi a vencedora do desafio do *International Symposium on Biomedical Imaging (ISBI)* em 2012 para a segmentação de estruturas neurais em microscopia eletrônica e, desde então, tem sido amplamente utilizada para a função de segmentação na área médica. A U-Net pode ser encontrada em arquiteturas tanto para a convolução 2D quanto para a convolução 3D, e uma hibridização de ambas as arquiteturas também pode ser utilizada, como foi feita por (JALALIFAR et al., 2020) na tarefa de segmentação de metástases cerebrais.

3.3.3.2 SegNet

A arquitetura SegNet (BADRINARAYANAN et al., 2017) tem como base uma rede na forma *encoder-decoder*, formada por redes *Visual Geometry Group (VGG)* e VGG

inversa, que diminui a dimensão do *input* para extração de características e depois aumenta a resolução em um processo de *upsampling*, essa arquitetura consiste em camadas convolucionais seguidas de camadas de *Batch Normalisation* com ativação ReLU. Como os índices dos mapas de características são salvos nas camadas de *MaxPooling* da parte *decoder* da rede, há um menor uso de poder computacional quando comparado com outras arquiteturas de rede, podendo ser um atrativo do uso dessa arquitetura. A aplicação proposta por (BADRINARAYANAN et al., 2017) foi feita para imagens de ambientes externos e internos para a segmentação de objetos, e não há registros de utilização para a segmentação de imagens médicas.

3.3.3.3 DeepMedic

Criada em 2017, a arquitetura de rede DeepMedic (KAMNITSAS et al., 2017) realiza convoluções 3D em dois caminhos de resoluções distintas da imagem do *input*, uma com alta resolução outra com baixa resolução, com *kernels* de tamanhos distintos para cada um dos caminhos e os agregando ao final, de modo a encontrar características globais e locais das lesões cerebrais. Essa rede é otimizada para identificação de lesões cerebrais do conjunto de dados do desafio do *Brain Tumor Segmentation* (BraTS) e demonstrou eficiência nos estudos publicados por (HU et al., 2019; LIU et al., 2017; CHARRON et al., 2018; PENNIG et al., 2021; JÜNGER et al., 2021) para a segmentação de metástases cerebrais. No entanto, pode trazer resultados superestimados quando utilizado em testes desse mesmo *dataset* (CHARRON et al., 2018).

3.3.3.4 GoogLeNet

A GoogLeNet é uma rede neural de 2014 que possui uma característica de concatenar operações de convolução com *kernels* de diferentes tamanhos antes de passar para uma próxima camada, também chamados de módulos *Inception*, e foi utilizada no estudo por (GRØVIK et al., 2020), alcançando resultados promissores na tarefa de segmentação de metástases cerebrais.

3.3.3.5 HRNetV2

HighResolution Net V2 é uma versão melhorada da HighResolution Net (SUN et al., 2019) que trabalha com imagens 2D de alta resolução, dividindo em sub-redes paralelas com *inputs* resoluções menores e agregando as informações de todas as sub-redes em paralelo. Essa arquitetura mostrou resultados promissores na publicação realizada por (OTTESEN et al., 2023) para a segmentação de metástases cerebrais, que é um dos modelos avaliados neste trabalho.

3.4 Métricas de Avaliação de performance

As métricas de avaliação de performance indicam como o modelo desempenha na função para a qual foi proposta, e podem variar de acordo com o tipo de rede neural utilizada e sua finalidade. Antes de tudo, é necessário realizar uma análise de risco para responder à seguinte pergunta: Qual é o risco associado aos erros gerados pelo resultado modelo na tarefa de segmentação de metástases cerebrais?

Para responder essa pergunta, tomamos um posicionamento conservador, em que o fato de não encontrar uma metástase cerebral é mais danoso do que classificar tecido sadio como sendo metástase cerebral, pois, *a priori*, o exame de imagem é realizado para auxiliar o tratamento das metástases.

Para esse estudo, as métricas utilizadas estão listadas nas subseções a seguir, e uma breve explicação sobre as abreviaturas estão listadas a seguir:

- **TP:** Número de *pixels* ou *voxels* classificados como verdadeiros positivos, isto é, onde o modelo apontou ser metástase e acertou.
- **TN:** Número de *pixels* ou *voxels* classificados como verdadeiros negativos, isto é, onde o modelo apontou ser tecido saudável e acertou.
- **FP:** Número de *pixels* ou *voxels* classificados como falsos positivos, isto é, onde o modelo apontou ser metástase e errou, pois se trata de metástase.
- **FN:** Número de *pixels* ou *voxels* classificados como falsos negativos, isto é, onde o modelo apontou ser tecido saudável e errou, pois se trata de tecido saudável.

3.4.1 Precisão

A precisão (PPV, do inglês *Predicted Positive Value*) é dada a partir da Equação (1):

$$PPV = \frac{TP}{TP + FP} \quad (1)$$

Nota-se que essa métrica depende de qual etiqueta está sendo utilizada como positiva ou negativa. No nosso caso, o positivo indica a presença de metástase na região analisada.

3.4.2 Sensibilidade ou *Recall*

A sensibilidade ou *Recall* (TPR, do inglês *True Positive Rate*) é dada a partir da Equação (2):

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

Essa métrica irá dizer o quando o modelo acerta em relação a todos os casos em que ocorrem metástases.

3.4.3 Acurácia

A acurácia é uma métrica que mede o total de acertos em relação ao total de previsões, ela é medida através da Equação (3):

$$Acuracia = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

3.4.4 Área sob a curva ROC (AuROCC)

A área sob a curva ROC (AuROCC, do inglês *Area under Receiver Operating Characteristics Curve*) é uma métrica que indica o quanto o modelo acerta em suas previsões tanto para as características positivas quanto para os negativas. Quanto mais próxima a área sob a curva estiver do valor unitário, maior é a confiabilidade do modelo em acertar seus previsões.

3.4.5 *F1-Score*

A *F1-Score* é uma métrica que relaciona a precisão e a sensibilidade através de uma média harmônica entre elas, e pode ser utilizada em conjunto com a acurácia para avaliar a qualidade do modelo, é descrita pela Equação (4):

$$F1 = 2 \times \frac{(PPV \times TPR)}{PPV + TPR} \quad (4)$$

Em que PPV é a precisão e TPR a sensibilidade.

3.4.6 Coeficiente Dice (DSC)

O coeficiente de similaridade Dice é utilizado para a comparação de duas amostras e é mensurado conforme indica a Equação (5).

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

Em que X e Y são os dois elementos a serem comparados, que no caso desse estudo são as máscaras de segmentação obtidas pela predição da rede e as máscaras de segmentação feita por radiologistas. Os valores para essa métrica podem ir de 0 até 1, em

que 0 indica nenhuma semelhança e 1 indica maior semelhança. O valor desse coeficiente é determinado através da contagem de *pixels* que possuem o mesmo valor atribuído tanto na predição quanto no resultado esperado.

3.4.7 Intersecção sobre a união (IoU)

A intersecção sobre a união (IoU, do inglês *Intersection over Union*), também conhecida como índice de Jaccard ou índice de Tanimoto, é uma métrica utilizada para verificar quanto uma amostra é condizente com o valor esperado, e é calculada através da Equação (6):

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} \quad (6)$$

Em que X e Y são os dois elementos a serem comparados, que no caso desse estudo são as máscaras de segmentação obtidas pela predição da rede e as máscaras de segmentação feita por radiologistas.

3.4.8 Distância de Hausdorff (HD)

A distância de Hausdorff (HD, do inglês *Hausdorff Distance*) é uma métrica de avaliação para encontrar a similaridade do formato dos contornos de duas máscaras (HUTTENLOCHER et al., 1993), que é definida através da Equação 7, apresentada a seguir:

$$H(A, B) = \max[h(A, B), h(B, A)] \quad (7)$$

Em que, a distância de Hausdorff direta de A para B é dada pela Equação 8 que pega um ponto a do conjunto A que esteja mais distante de todos os pontos do conjunto B e mede a distância desse ponto a até o ponto b do conjunto B mais próximo, utilizando-se uma métrica de distância, como a norma Euclidiana $\|a - b\|$ (HUTTENLOCHER et al., 1993).

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (8)$$

Um exemplo de como a HD é calculada para as máscaras de segmentação em relação à anotação *ground truth* está apresentado na Figura 7.

No entanto, para a aplicação desse estudo, não é possível calcular a distância total de Hausdorff entre as máscaras considerando-se todos os *voxels*, porque *clusters* não correspondentes, isto é, conjuntos de *voxels* que são encontrados na predição mas que não possuem um conjunto de *voxels* correspondente na máscara de segmentação

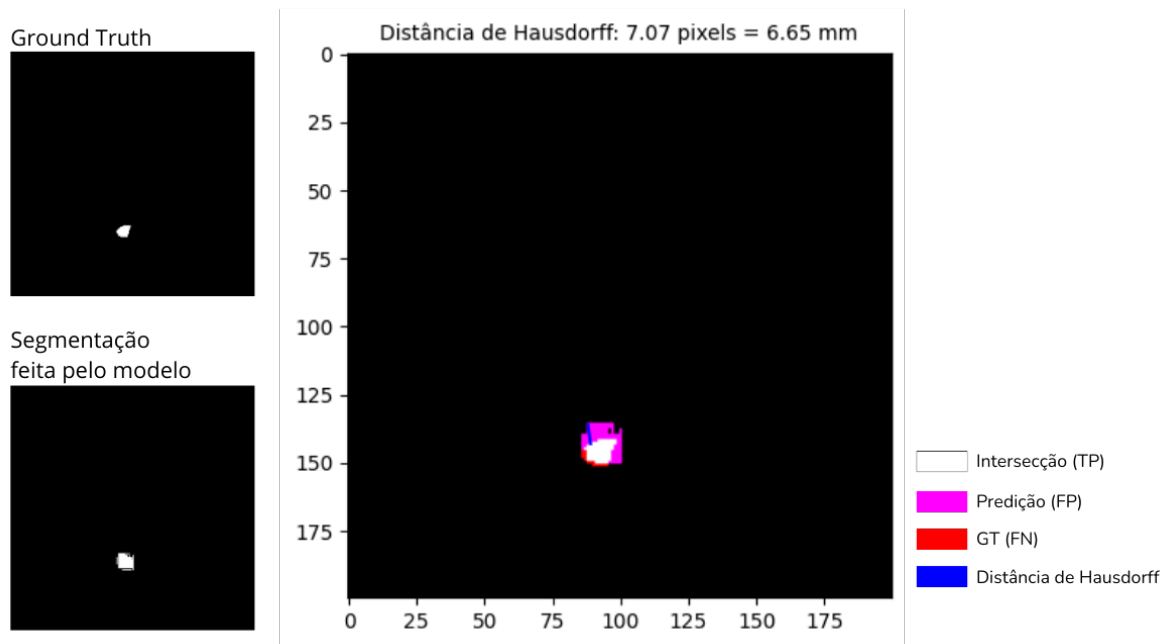


Figura 7 – Exemplo do cálculo da distância de Hausdorff entre a máscara de segmentação de referência (*Ground Truth*) e a segmentação feita pelo modelo, ou predição. Os números na imagem ao centro indicam a posição em *pixels*.

Fonte: Elaborado pelo autor (2024).

(falsos positivos) e vice-versa (falsos negativos), resultam em distâncias de Hausdorff altas quando há múltiplos volumes tumorais em um mesmo exame, mesmo em situações em que a segmentação é exatamente a predição para um determinado conjunto de *voxels*, ou *cluster*.

Além disso, não se pode calcular a média de todas as distâncias de Hausdorff, pois esse resultado é dependente da quantidade de volumes tumorais e traz pouca informação sobre a performance geral do modelo. Além disso, erros grotescos não teriam tanta influência em casos de muitos volumes tumorais, o que pode ocasionar problemas graves para a aplicação clínica.

3.5 Funções de Perda

As funções de perda, comumente chamadas de *loss functions* ou simplesmente *loss*, são funções matemáticas que permitem a medição da diferença entre o resultado que se obteve e o resultado esperado, e são extremamente importantes para realizar a tarefa desejada (CHOLLET, 2017). É através dos valores de *loss* encontrados ao final do treino de uma época do modelo que se obtém o *feedback* se aquela época melhorou ou não a eficiência do modelo em realizar a tarefa para o qual foi treinado.

Em modelos de segmentação, algumas funções perda podem ser bastante úteis, enquanto outras funções perda não irão realizar a tarefa desejada. Entre os tipos de

funções perda utilizadas nas tarefas de segmentação por redes neurais convolucionais, podemos considerar que existem quatro categorias principais (SUGINO et al., 2021), que estão embasadas em:

- Distribuição - mede a dissimilaridade com base na entropia cruzada, como por exemplo a *Cross-Entropy Loss*.
- Região (volume ou área) - quantifica a incompatibilidade das intersecções das regiões, como por exemplo a *Dice Loss*.
- Bordas - Leva em consideração à distância entre as bordas da predição e o esperado, como por exemplo a *Boundary Loss*, ou as formadas pelas distâncias Euclidiana ou de Hausdorff.
- Composições - Podem ser feitas composições entre essas três categorias supracitadas.

Além disso, ao comparar os resultados da segmentação de quatro tipos de estruturas cerebrais, o estudo de (SUGINO et al., 2021) avalia o uso de doze combinações de funções perda com *weighting* e classifica as melhores combinações das funções perda *Cross Entropy Loss* com *Distance Transform Map-based weighting* (DTM) e a *Dice Loss* com *Focal Weighting* para a tarefa de segmentação binária.

Os modelos treinados nesse estudo usaram funções perda individuais e compostas, e um detalhamento sobre cada função perda bem como suas justificativas para uso estão apresentadas nas subseções a seguir.

3.5.1 *Cross-Entropy Loss*

A função perda de entropia cruzada *Cross-Entropy Loss*, também chamada de perda logarítmica ou *log loss*, é uma equação que realiza a otimização de regras de aprendizado através da minimização das suas variâncias para a predição de eventos raros (BOER et al., 2005) e pode ser definida para o uso em *Deep Learning* através da Equação 9:

$$CE(y, \hat{p}) = \begin{cases} -\log(\hat{p}), & \text{se } y = 1 \\ -\log(1 - \hat{p}), & \text{se } y = 0 \end{cases} \quad (9)$$

Em que y representa a classe real do *pixel*, nesse caso binário, e \hat{p} representa a predição da rede. Essa equação pode ser escrita na forma de função de perda, dada pela Equação 10:

$$\mathcal{L}_{CE}(y, \hat{p}) = -y \log(\hat{p}) - (1 - y) \log(1 - \hat{p}) \quad (10)$$

3.5.2 *Balanced Cross-Entropy Loss*

A equação *Balanced Cross-Entropy Loss* é uma variação da *Cross-Entropy Loss* com a adição de um parâmetro α , que realiza o balanceamento de classes e é obtido através da razão entre a quantidade de classe em relação ao conjunto inteiro.

$$BCE(y, \hat{p}) = \begin{cases} -\alpha \log(\hat{p}), & \text{se } y = 1 \\ -\alpha \log(1 - \hat{p}), & \text{se } y = 0 \end{cases} \quad (11)$$

Que pode ser escrita na forma da Equação 12:

$$\mathcal{L}_{BCE}(y, \hat{p}) = -\alpha y \log(\hat{p}) - \alpha(1 - y) \log(1 - \hat{p}) \quad (12)$$

3.5.3 *Dice Loss*

O coeficiente Sørensen–Dice, também chamado de Dice, índice de similaridade Dice ou DSC, do inglês *Dice Similarity Coefficient*, ou mesmo *F1-Score*, é um índice que mede a similaridade entre duas amostras com base na razão entre a intersecção das duas amostras e a soma dessas duas amostras, que é dada pela Equação 13:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2|X \cap Y|}{|X \cap Y| + |X \cup Y|} \quad (13)$$

Em redes neurais, durante o treino em uma época, o coeficiente Dice pode ser utilizado como função perda para a atualização dos parâmetros de treino da rede. Para aplicações Booleanas, a função perda *Dice Loss* (\mathcal{L}_{Dice}) pode ser dada em termos de verdadeiros positivos (tp , do inglês *true positive*), falsos negativos (fn , do inglês *false negative*) e falsos positivos (fp , do inglês *false positive*) pela Equação 14:

$$\mathcal{L}_{Dice}(tp, fp, fn) = \frac{(1 + \beta^2) \cdot tp}{(1 + \beta^2) \cdot tp + \beta^2 \cdot fn + fp} \quad (14)$$

Em que o fator β é um parâmetro de peso para a sensibilidade em relação à precisão, pois a depender do caso uma das métricas pode ser mais relevante do que a outra, geralmente tem seu valor padrão como sendo $\beta = 1$. No entanto, apenas a função perda com base no coeficiente Dice não é suficiente para realizar a tarefa de segmentação, e o uso combinado dessa função com a BCE é uma das soluções para melhorar o resultado da segmentação (SUGINO et al., 2021). Além disso, o fato de haver um grande desbalanceamento de classes entre os *pixels* que são partes de metástases e os *pixels* que fazem parte do *background* faz com que a rede neural ignore os *pixels* classificados como metástases durante o treino, o que pode ser resolvido com a variação da função Dice, como a *Generalized Dice Loss*.

3.5.4 Generalized Dice Loss

Uma estratégia de ponderação ou pesagem, para mitigar o efeito do desbalanceamento das classificações dos *pixels* para segmentação é o uso da função perda *Generalized Dice Loss* (GDL), que leva em consideração um fator peso que é inversamente proporcional ao quadrado da frequência (SUGINO et al., 2021) e é dado pela Equação 15 (SUDRE et al., 2017):

$$\mathcal{L}_{\text{GenDice}} = 1 - 2 \frac{\sum_{l=1}^2 w_l \sum_n y_{ln} \cdot \hat{p}_{ln}}{\sum_{l=1}^2 w_l \sum_n y_{ln} + \hat{p}_{ln}} \quad (15)$$

Em que $w_l = \left(\frac{1}{\sum_{n=1}^N y_{ln}} \right)^2$ é o fator peso inversamente proporcional com o tamanho da região, e diminui a correlação entre o DSC e a área. Em um treino realizado com imagens do BRATS para a segmentação de tumores cerebrais, a GDL foi comparada com as funções perda *Weighted Cross-Entropy Loss*, *Dice Loss* e *Sensitivity-Specificity Loss*, resultando em um maior valor para a métrica DSC para cinco de nove configurações de treino com a rede U-Net, e em quatro de nove configurações de treino com a rede TwoPathCNN. Nessas duas arquiteturas os resultados para a DSC encontrado com a GDL foram sempre maiores do que nas outras duas arquiteturas estudadas, a DeepMedic e HighResNet (SUDRE et al., 2017).

3.5.5 Binary Focal Loss

Um problema enfrentado por redes que fazem detecção de objetos e que está relacionado às redes de segmentação é o fato de haver um grande desbalanceamento de classes, já que os *pixels* que não fazem parte das metástases são considerados como *background* e são predominantes. Esse desbalanceamento pode ser contornado utilizando algumas estratégias como a *hard negative mining*, que adiciona exemplos negativos no treino para diminuir a quantidade de falsos positivos, ou pelo uso da função perda *Focal Loss* (LIN et al., 2020), que é definida pela Equação 16.

$$\mathcal{L}_{\text{Focal}}(y, \hat{p}) = -\alpha y (1 - \hat{p})^\gamma \log(\hat{p}) - \alpha (1 - y) \hat{p}^\gamma \log(1 - \hat{p}) \quad (16)$$

Em que y é o *label* da classe binária 0 ou 1, \hat{p} é a probabilidade da estimativa estar correta, γ é o parâmetro focal que especifica o quanto as predições corretas pesam para a contribuição da estimativa da *loss*, geralmente definido como igual a 2 quando usado em combinação com a *Cross Entropy Loss* ou igual a 2 quando usado em combinação com a *Dice Loss*, e α é o hiperparâmetro que equilibra a relação entre precisão e sensibilidade, que por *default* é igual a 1 e indica que não há preferência para uma entre essas duas (LIN et al., 2020). Essa equação uma variação da função perda *Balanced Cross Entropy* (*BCE*), que é dada pela Equação 11, com a adição do parâmetro focal \hat{p}^γ e $(1 - \hat{p})^\gamma$, que

coloca ênfase em *pixels* cuja classe é de difícil classificação durante o treino, tendo como base o valor de \hat{p} (SUGINO et al., 2021).

3.5.6 Boundary Loss

Uma função perda que leva em consideração a borda da segmentação é denominada *boundary loss* e foi proposta por Kervadec et al (2021) na tarefa de segmentação de hiperintensidades da substância branca em imagens de ressonância magnética cerebrais ponderadas em T1 e FLAIR (KERVADEC et al., 2021). Assim como para o caso de metástases cerebrais, essas regiões hiperintensas podem se apresentar em pequenos tamanhos e a função *boundary loss* mostrou-se mais eficiente na tarefa de segmentação dessas regiões quando comparada à *Generalized Dice Loss*, e é definida pela Equação 17:

$$\mathcal{L}_{\text{boundary}} = \frac{1}{N} \sum_n^N \hat{p}_n \cdot (d_{\text{borda}}(fp) - d_{\text{borda}}(tp)) \quad (17)$$

Em que N é o número total de *pixels*, \hat{p}_n é a predição da rede e d_{borda} é a distância do *pixel* até a borda da segmentação. Note que esse valor é minimizado com os acertos da região interna à borda, dados pelos verdadeiros positivos (tp), e maximizado com os erros da rede fora da região da segmentação, determinados pelos falsos positivos (fp). Assim, como durante a otimização a função perda será minimizada, essa função *boundary loss* mostra-se útil para a tarefa de segmentação.

3.5.7 Tversky Loss

A função perda de Tversky (TL, do inglês *Tversky Loss*) consiste numa equação que mede a similaridade de dois conjuntos, e pode ser usada no lugar da função perda Dice em casos de grande desbalanceamento de dados, para que a rede neural possa ponderar a influência dos falsos negativos e falsos positivos durante o treino do modelo (SALEHI et al., 2017). Essa função, definida por Salehi et al. (2017), é dada pela Equação 18.

$$TL(\eta, \psi) = \frac{\sum_{i=1}^N p_{0_i} g_{0_i}}{\sum_{i=1}^N p_{0_i} g_{0_i} + \eta \sum_{i=1}^N p_{0_i} g_{1_i} + \psi \sum_{i=1}^N p_{1_i} g_{0_i}} \quad (18)$$

Em que o somatório é feito para todos os *voxels* i das imagens, p_{0_i} representa a probabilidade do *voxel* i ser da classe 0 e p_{1_i} a probabilidade de ser da classe 1, enquanto que os termos g_{0_i} e g_{1_i} representam as probabilidades *Ground Truth* de serem essas mesmas classes, que assumem valores de 0 (0%) ou 1 (100%). Os hiperparâmetros η e ψ são os pesos para ponderar os falsos positivos e falsos negativos.

Para o treino do modelo, foi utilizada uma adaptação dessa equação, que não leva em conta as probabilidades dos *voxels* serem uma determinada classe antes de passar pelo

threshold que define sua classe. Além disso, assim como para a função perda Dice, foi adicionado o fator β para ponderar a sensibilidade em relação à precisão, apresentado inicialmente na Equação 14. Essa adaptação em termos de verdadeiros positivos, falsos positivos e falsos negativos é dada pela Equação 19.

$$TL(tp, fp, fn) = \frac{(1 + \beta^2) \cdot tp}{(1 + \beta^2) \cdot tp + \beta^2 \cdot \eta \cdot fn + \psi \cdot fp} \quad (19)$$

Nota-se que para o caso em que η e ψ são ambos iguais a 1, a função perda de Tversky se torna idêntica à função perda Dice e, para β igual a 1 e η e ψ iguais a 2, a equação da função se torna análoga à Equação 6 da métrica IoU (SALEHI et al., 2017). Salehi *et al.* apontam para uma melhoria no resultado da segmentação de lesões cerebrais ao aumentar o valor do hiperparâmetro η , associado aos falsos negativos.

3.6 Recomendações para publicação de trabalhos relacionados à Física Médica com uso de Inteligência Artificial

De modo a tornar a pesquisa envolvendo a IA na área da Física Médica mais replicável e que outros pesquisadores possam ter maiores informações sobre o rigor, a qualidade e a generabilidade de métodos e modelos desenvolvidos em pesquisas, algumas diretrizes foram definidas e também se encontram em desenvolvimento. De acordo com a Agência Internacional de Energia Atômica (IAEA, do inglês *International Atomic Energy Agency*), existem 10 diretrizes publicadas para as aplicações de IA em estudos clínicos (IAEA, 2023). Neste trabalho, duas diretrizes que estão associadas à avaliação de imagens médicas e física médica foram escolhidas e serão utilizadas para avaliação: CLAMP (NAQA et al., 2021) e CLAIM (MONGAN et al., 2020). Assim, foi verificado se os itens descritos pelos autores estavam presentes ou não nas seções especificadas das publicações envolvendo os modelos avaliados e, em caso de estarem parcialmente apresentados, foram marcados como não adequados. Ao final, espera-se que publicações estejam com pelo menos dois terços de adequação aos *checklists* para terem sua qualidade avaliada como adequada.

3.6.1 CLAMP

A American Association of Physicists in Medicine possui uma lista de diretrizes que são consideradas essenciais para publicação de artigos na temática de Inteligência Artificial na Medicina. O *Checklist for AI in Medical Physics* (CLAMP) apresentado por (NAQA et al., 2021) consiste em 26 itens que foram levados em consideração, com informações sobre as imagens adquiridas, o método do treino, a apresentação dos resultados, a metodologia empregada e materiais utilizados para garantir uma maior qualidade dos dados e maior

relevância dos resultados obtidos. Esse *checklist* separa os itens por seção da publicação, que estão descritos na lista a seguir.

1. Resumo

- 1) Proposta, novidade e significância.
- 2) Tamanho do dataset e como ele foi particionado para o treino.
- 3) Resultados com análise estatística.

2. Introdução

- 4) Proposta e justificativa.
- 5) Contribuição para a física médica.
- 6) Estágio da pesquisa.

3. Materiais

- 7) Característica do conjunto de dados, incluindo tamanho e local de obtenção.
- 8) Descrição dos aparelhos utilizados na obtenção dos dados e informações sobre o pré-processamento de dados.
- 9) Modalidade de aquisição das imagens, protocolos e parâmetros em detalhes.
- 10) Dados da população utilizados, representatividade e prevalência da doença.
- 11) Tipo de fantasmas utilizados e métodos para gerar dados sobre os fantasmas, quando aplicável.
- 12) Composição de adequação dos dados para o uso em ML.
- 13) Descrição da *Ground Truth* e como essas informações foram anotadas, bem como os níveis de subjetividade e incerteza.
- 14) Partição dos dados em treino, validação e teste. Assim como formas de mitigar os vieses e justificativa do tamanho da amostra.
- 15) Validação final utilizando um *dataset* público, ou tornando público o *dataset* utilizado.

4. Métodos

- 16) Metodologia replicável, com informações sobre arquiteturas utilizadas, hiperparâmetros, *inputs* e suas dimensionalidades, técnicas de pré-processamento e tipo de *output*.
- 17) Informações sobre o treino e otimização, como a função perda utilizada, técnicas de regularização, maneiras de reduzir *overfitting*.

- 18) Disponibilidade pública do código utilizado no treino.
- 19) Métodos de avaliação de performance e pós-processamento do *output*.
- 20) Métodos para estimar a incerteza, bem como seus intervalos de confiança de 95%.
- 21) Análises de subgrupos para alguns subgrupos relevantes, como sexo, idade, tamanho da metástase, etc.
- 22) Significância dos resultados obtidos comparados com a hipótese nula ou com alguma métrica de referência.
- 23) Resultados demonstrativos do treino, validação e teste.

5. Discussão

- 24) Conclusões suportadas pelos resultados.
- 25) Limitações do estudo.
- 26) Inovação, significância e contribuições para a física médica.

3.6.2 CLAIM

Além do CLAMP, um estudo publicado na revista *Radiology* por (MONGAN et al., 2020) indica outros itens que também devem estar presentes em estudos envolvendo o uso de IA em aplicações nas imagens médicas. O *Checklist for Artificial Intelligence in Medical Imaging* (CLAIM) é constituído de recomendações que visam promover transparência e reprodutibilidade no contexto de pesquisas na área de IA com aplicação médica. Esse conjunto de recomendações foi processado através método Delphi para consenso envolvendo um grupo de 78 especialistas em imagens médicas, cientistas de IA, editores de jornais científicos e estatísticos. O CLAIM passou por atualização para se tornar mais adequado de acordo com os avanços na área da inteligência artificial (TEJANI et al., 2024) e atualmente conta com 44 itens que também estão divididos por seção da publicação, descritos na lista a seguir:

1. Título ou Resumo

- 1) Identificação da categoria de IA utilizada, como por exemplo o *Deep Learning*.

2. Resumo

- 2) Resumo estruturado com desenho do estudo, métodos, resultados e conclusão.

3. Introdução

- 3) Contexto científico e clínico da aplicação, a função do algoritmo e como é a prática atual.
- 4) Objetivos do estudo e hipóteses.

4. Métodos

a) Desenho de pesquisa:

- 5) Prospectivo ou retrospectivo.
- 6) Objetivo do estudo.

b) Dados:

- 7) Fonte dos dados.
- 8) Critérios de elegibilidade dos participantes.
- 9) Etapas de pré-processamento.
- 10) Escolha dos subconjuntos de dados.
- 11) Métodos de desidentificação.
- 12) Formas de contornar os dados faltantes.
- 13) Protocolo de aquisição de imagem.

c) Referência Padrão:

- 14) Definição do método para encontrar a referência padrão, com detalhes para permitir replicação.
- 15) Motivos para utilizar o padrão de referência em caso de existirem alternativas.
- 16) Fonte das anotações da referência padrão, as qualificações e preparações dos anotadores.
- 17) Ferramentas utilizadas na anotação dos dados.
- 18) Medidas de variabilidade e discrepância entre os anotadores e métodos para mitigar essa variabilidade.

d) Partições dos dados:

- 19) Como os dados foram particionados e suas proporções.
- 20) Nível em que os dados estão desagrupados, por exemplo por instituição, pacientes, etc.
- 21) Tamanho amostral pretendido.

e) Modelo:

- 22) Descrição completa do modelo, *inputs*, *outputs*, camadas intermediárias e conexões.
- 23) Bibliotecas, *frameworks* e pacotes utilizados no *software*.

- 24) Parâmetros de inicialização do modelo, como randomização ou *transfer learning*.

f) Treino:

- 25) Detalhes do treino, como *data augmentation*, hiperparâmetros e número de modelos treinados.
- 26) Método de seleção do modelo final.
- 27) Técnicas de *ensemble*, quando utilizadas.

g) Avaliação:

- 28) Métricas de performance do modelo.
- 29) Medidas estatísticas e incertezas associadas.
- 30) Análise de robustez ou sensibilidade.
- 31) Métodos de explicabilidade ou interpretabilidade dos resultados e como é feita a validação.
- 32) Avaliação em dados internos.
- 33) Avaliação em dados externos.
- 34) Registro de ensaio clínico, se aplicável.

5. Resultados

a) Dados:

- 35) Fluxo de participantes ou de casos, com fluxogramas representando os critérios de seleção.
- 36) Características clínicas e demográficas dos participantes em cada partição.

b) Performance do modelo:

- 37) Métricas da melhor performance para cada partição estudada.
- 38) Estimativa da acurácia e precisão do modelo e seus intervalos de confiança.
- 39) Análise de falha para os falsos positivos (FP) e falsos negativos (FN).

6. Discussão

- 40) Limitações do estudo, incluindo vieses, incertezas estatísticas e generalização.
- 41) Implicações práticas e função na clínica.

7. Outras informações:

- 42) Referência ao protocolo do estudo ou outros detalhes técnicos.

- 43) Disponibilidade do software, modelo treinado ou dataset.
- 44) Fontes de financiamento e suporte, bem como o papel das fontes na pesquisa.

Para esse *checklist*, cada um dos itens possui uma subseção com descrições mais detalhadas, que pode ser encontrada nas publicações original e atualizada (MONGAN et al., 2020; TEJANI et al., 2024). Essas descrições foram tomadas como referência na hora de avaliar o *checklist* e, no caso de não adequação ou adequação parcial, o item foi marcado como não adequado.

4 Revisão da Literatura

Neste capítulo, será apresentada a revisão da literatura que permite identificar estudos primários sobre o uso de inteligência artificial baseada em redes neurais na segmentação de metástases cerebrais, para responder à pergunta de pesquisa: "Como a inteligência artificial baseada em redes neurais é utilizada na segmentação de metástases cerebrais?".

Essa revisão é similar à publicada por CHO *et al.* (2020), o qual apresenta uma revisão sistemática com meta-análise e compara diferentes técnicas de aprendizado de máquina envolvendo metástases cerebrais e aponta para a escolha de técnicas em *Deep Learning* para a detecção, que se mostraram superiores às técnicas clássicas de *Machine Learning* (CHO *et al.*, 2020). A revisão apresentada nesse trabalho se difere da apresentada por Cho *et al.* (2020), por possuir uma abordagem mais específica na tarefa de segmentação de tumores metastáticos com o uso de aprendizado profundo, e inclui estudos que foram publicados posteriormente a sua publicação.

Sendo assim, foi realizada uma busca nas bases de dados PubMed, IEEE Xplore e Epistemonikos para a busca de publicações relacionadas com a pesquisa de *Deep Learning* na tarefa de segmentação de metástases cerebrais publicadas desde o ano de 2017 até o dia 10 de Junho de 2024. A busca utilizando os descritores (brain AND "deep learning" AND segmentation AND metas*) retornou um total de 212 artigos nessas três bases, dos quais foram excluídos 8 duplicatas, 36 artigos de revisão, 28 artigos que não tinham tumores cerebrais como foco, 38 artigos que não realizavam a segmentação de tumores cerebrais e 50 artigos que realizavam a segmentação de tumores cerebrais mas não faziam a discriminação de tumor metastático dos outros tipos de tumores, como gliomas ou meningiomas. Além disso, foram adicionados dois artigos (HU *et al.*, 2019) e (RUDIE *et al.*, 2021) de outras fontes. Assim, foram selecionados 46 artigos para a leitura do texto completo. O fluxograma do protocolo PRISMA (PAGE *et al.*, 2021) para a seleção dos artigos está apresentado na Figura 8.

A distribuição dessas publicações ao longo dos anos pode ser vista na Figura 9 e, na Figura 10, estão destacados os países envolvidos nas pesquisas selecionadas para a revisão.

Percebe-se que há um aumento progressivo nos estudos com o passar dos anos, o que pode ser explicado pelo aumento do interesse e desenvolvimento na área da aplicação de aprendizado de máquina em áreas relacionadas à saúde. Além disso, as pesquisas nesta área estão restritas ao hemisfério norte, principalmente Estados Unidos, Canadá e China, com algumas publicações no oriente médio, Europa e Rússia. Isso indica a necessidade de

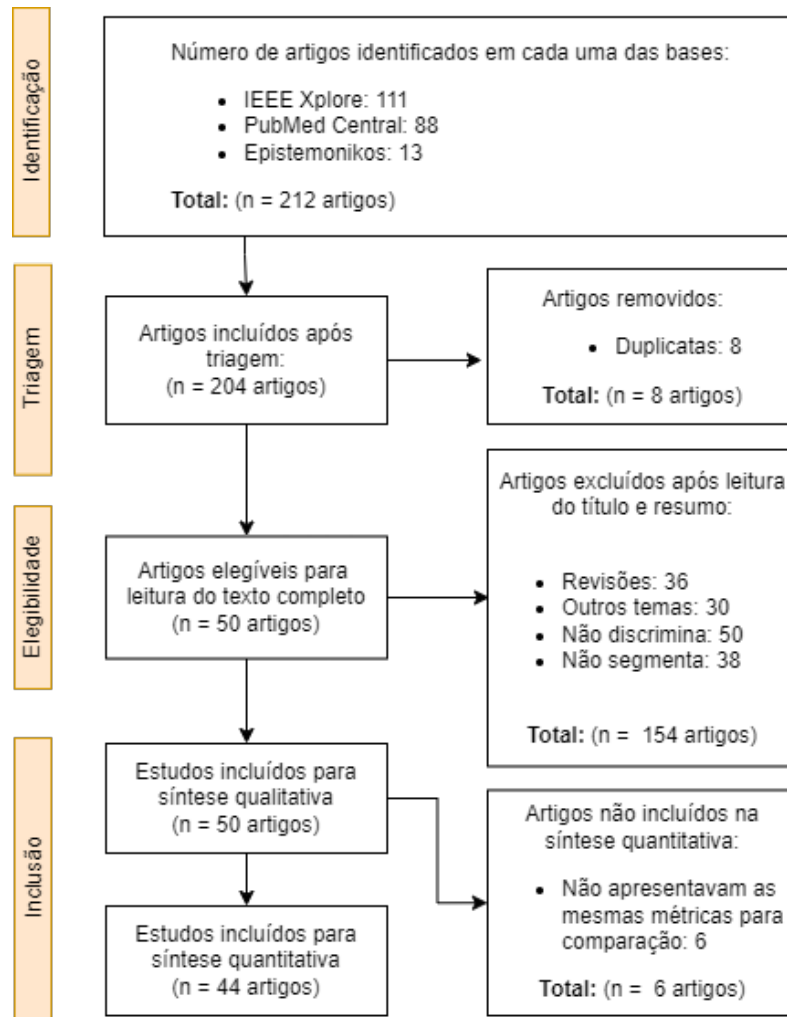


Figura 8 – Fluxo de inclusão dos estudos seguindo o protocolo PRISMA para revisões sistemáticas.

Fonte: Autoria própria.

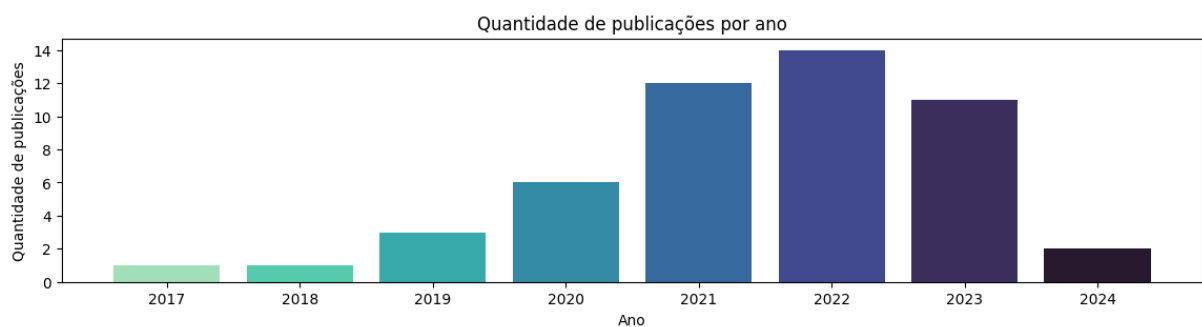


Figura 9 – Publicações ao longo dos anos dos artigos encontrados nas bases PubMed, IEEE Xplore e Epistemonikos que foram selecionados para a leitura completa

Fonte: Autoria própria.

serem realizadas mais pesquisas nessa área no hemisfério sul do globo, para mitigar os efeitos de viés por localidade.

Uma comparação entre os resultados destes estudos pode ser visto na Tabela 1.

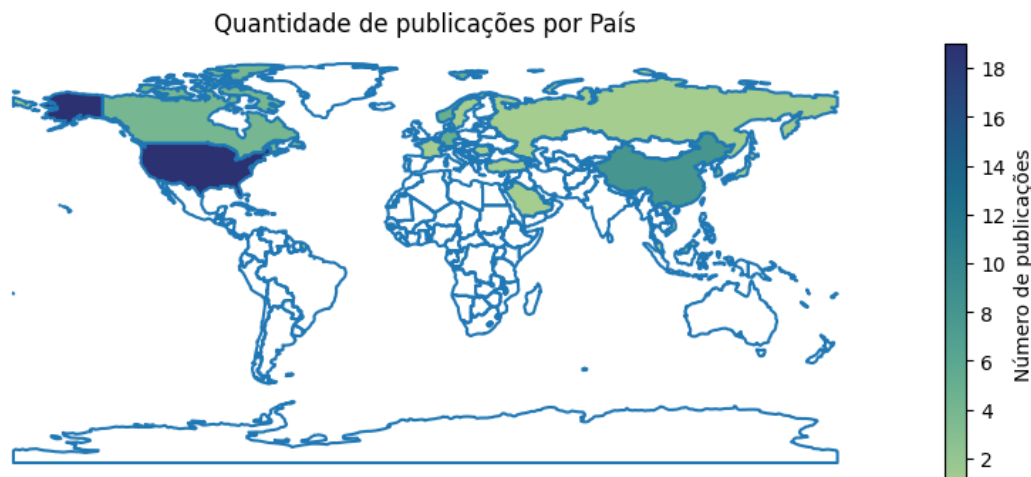


Figura 10 – Quantidade das publicações incluídas na revisão por país
Fonte: Autoria própria.

A seguir, encontra-se uma descrição com a síntese dos métodos e resultados dos trabalhos que foram selecionados para a inclusão na revisão.

(GRØVIK et al., 2020) utilizaram um dos datasets que foi utilizado neste estudo, o *BrainMetShare*, para realizar a identificação das metástases cerebrais com o uso de redes neurais com arquitetura de rede 2.5D baseada na GoogLeNet, e avaliaram as métricas de precisão, sensibilidade, F1 Score, coeficiente Dice e AuROCC. Entre os resultados obtidos, obtiveram valores para a AuROCC de $0,98 \pm 0,04$ e, para precisão, sensibilidade e coeficiente Dice, encontraram os valores de, respectivamente, $0,79 \pm 0,20$, $0,52 \pm 0,22$ e $0,79 \pm 0,12$.

Um outro estudo realizado por (HU et al., 2019) aplica uma função perda com pesos diferentes de acordo com o volume da metástase cerebral, discorrendo sobre o uso da função perda Dice ser ideal para conjuntos de dados com metástases de tamanhos semelhantes, o que pode trazer resultados superestimados quando o modelo é aplicado em casos com metástases de tamanhos variados. Isso pode ser um problema, já que é comum casos em que coexistem metástases de tamanhos diversos. Esse risco se mostra ainda mais relevante ao considerar-se que a probabilidade de um radiologista encontrar uma metástase é menor conforme menor for o seu tamanho. Nesse estudo, (HU et al., 2019) utilizaram um conjunto de dados formado por imagens de RM e imagens de TC, visto que cada técnica possui suas vantagens e desvantagens, e a combinação de ambas possibilita uma maior confiabilidade diagnóstica. Para o treino, foi utilizada um *ensemble* de duas arquiteturas, a 3D U-Net e a DeepMedic. Para o treino da rede 3D U-Net, foi utilizado o otimizador *RMSprop*, com LR de 10^{-3} , BS de 1 e durante 300 épocas. Já para a rede DeepMedic, foram utilizados três caminhos paralelos, um que treina a rede com as imagens em sua máxima resolução e outro que utiliza *patches*, ou pedaços, que são subamostras dessas imagens para o treino. Os resultados obtidos se mostraram promissores quando aplicado

Tabela 1 – Comparação do número N de casos/pacientes no treino o *split* em treino, validação e teste (T/V/T), o coeficiente Dice (DSC), a precisão (PPV), a sensibilidade (TPR) e a quantidade de falsos positivos por paciente (FPP) dos estudos encontrados em literatura para a segmentação de metástases cerebrais. Os melhores resultados estão destacados em negrito.

Autores	N	T/V/T	DSC	PPV	TPR	FPP
(LIU et al., 2017)	490	5-fold	0,67	-	-	-
(CHARRON et al., 2018)	182	146/18/18	0,79	-	0,98	4,4
(HU et al., 2019)	305	245/30/30	0,74	0,78	0,803	-
(GRØVIK et al., 2020)	156	100/5/51	0,8	0,8	0,5	8,3
(DIKICI et al., 2020)	158	5-fold	-	-	0,9	9,12
(XUE et al., 2020)	1201	4-fold	0,85	-	0,96	-
(JALALIFAR et al., 2020)	106	90/6/10	0,9	-	-	-
(BOUSABARAH et al., 2020)	509	469/0/40	0,71	0,96	0,71	0
		469/0/40	0,70	1,00	0,71	0,05
		257/233/0/24	0,27	0,95	0,53	0,05
(ZHOU et al., 2020)	934	748/186	0,8	0,6	0,9	3
(RUDIE et al., 2021)	563	413/50/100	0,75	0,915	0,7	0,46
(LU et al., 2021)	570	-	0,836	-	0,957	0,5
(NOMURA et al., 2021)	22	17/5	0,727	-	0,766	-
(PENNIG et al., 2021)	69	5-fold	0,75	0,74	0,88	0,71
(JÜNGER et al., 2021)	98	66/17/15	0,72	-	0,851	1,8
(PARK et al., 2021b)	188	188/0/94	0,8	0,931	0,907	0,59
(YI et al., 2021)	156	-	0,85	0,68	0,92	-
(YOO et al., 2021)	442	341/36/45	0,62	0,75	0,74	0,53
(GRØVIK et al., 2021)	165	100/10/55	0,8	0,8	0,7	12
(HSU et al., 2021)	511	5-fold	0,76	-	0,95	2,4
(PARK et al., 2021a)	330	290/40	0,66	-	0,877	1,9
(CHO et al., 2021)	147	127/20/35	0,7	-	0,877	1,9
		127/20	0,7	-	0,751	0,8
		127/20/12	0,8	-	0,947	0,5
(SUN et al., 2021)	450	285/72/93	-	-	-	-
(HUANG et al., 2022)	176	135/-/32	0,81	0,987	0,975	-
(YOO et al., 2022)	65	53/-/12	0,75	-	0,97	1,25
(BOUGET et al., 2022)	396	5-fold	0,9	0,9	0,9	0,061
(LIANG et al., 2022)	407	326/78/81	0,73	-	0,91	1,9
(SHIROKIKH et al., 2022)	778	3-fold	-	-	0,9	3,8
(CEAUŞESCU; ALEXE, 2022)	105	63/21/21	0,9	-	0,91	-
(LYU et al., 2022)	1399	10-fold	0,9	-	0,9	-
(LIEW et al., 2022)	156	84/51/21	-	-	0,811	2,952
		144/121/23/0	-	-	0,74	3,131
(CHARTRAND et al., 2022)	530	383/50/97	0,73	0,81	0,91	0,66
(JALALIFAR et al., 2022)	124	89/10/25	0,90	-	-	-
(LEE et al., 2022)	574	474/0/100	0,82	0,85	0,87	-

Continua

Tabela 2 – Comparação do número N de casos/pacientes no treino o *split* em treino, validação e teste (T/V/T), o coeficiente Dice (DSC), a precisão (PPV), a sensibilidade (TPR) e a quantidade de falsos positivos por paciente (FPP) dos estudos encontrados em literatura para a segmentação de metástases cerebrais. Os melhores resultados estão destacados em negrito.

Autores	N	T/V/T	DSC	PPV	TPR	FPP
			0,93	-	0,88	1
(OTTESEN et al., 2023)	221	100/10/116	0,93	-	0,86	0,4
			0,94	-	0,78	0,1
(LI et al., 2023)	649	295/99/255	0,81	0,56	0,9	5
(JALALIFAR et al., 2023)	116	90/-/20	0,92	-	-	-
(QU et al., 2023)	1000	850/150	0,9	0,93	0,88	1,01
(WANG et al., 2023)	100	-	0,723	-	0,997	0,72
(XUE et al., 2023)	121	4-fold	0,836	0,848	0,836	-
(SHU et al., 2023)	1000	800/200	0,896	0,892	0,9	-
(BUCHNER et al., 2022)	339	239/100	0,96	0,96	-	-
(ZHOU et al., 2023)	193	155/38	0,89	0,91	0,9	-
(SON et al., 2023)	186	128/58	0,66	0,796	0,869	1,76
(HAMMER et al., 2024)	226	169/57	0,90	0,76	0,94	-

Fonte: Autoria Própria.

o *ensemble* dessas duas redes, alcançando um DSC de 0.740 ± 0.022 , uma precisão de 0.779 ± 0.010 e uma sensibilidade de 0.803 ± 0.001 . Assim, os autores encontraram uma maior eficiência da rede em detectar metástases pequenas ao utilizar a função perda dependente do volume da metástase, que chamaram de *volume-aware Dice loss*.

(RUDIE et al., 2021) realizaram um estudo retrospectivo com 413 pacientes que foram submetidos à radiocirurgia. Foram utilizadas 563 imagens de ressonância magnética ponderadas em T1 pós contraste, separadas em 413 para treino, 50 para validação e 100 para teste. Em uma etapa do pré-processamento, foi realizada a subtração da imagem ponderada em T1 pré-contraste da imagem ponderada em T1 pós-contraste, de modo a ter um maior enfoque nas regiões de acúmulo de contraste. O treino foi realizado utilizando uma rede 3D U-net, com regularizador de normalização por *batch* (BN, do inglês *Batch Normalization*), otimizador Adam e com *learning rate* de 10^{-4} . A avaliação foi feita pelo coeficiente Dice, precisão e sensibilidade, que resultaram em valores de, respectivamente, 0,75, 91,5% e 70,0%. Para metástases com tamanhos maiores do que 6 milímetros, a sensibilidade da rede foi de 96,4%, indicando uma correlação entre o tamanho da metástase e a capacidade da rede em identificá-la, assim como aponta (HU et al., 2019). Além disso, os autores, avaliaram a confiabilidade do formato da metástase encontrada pela rede através do cálculo da distância de Hausdorff. Por fim, as mesmas métricas foram avaliadas entre dois diferentes radiologistas com 4 e 5 anos de experiência, que levaram cerca de 15 a 20 minutos para realizar a segmentação e obtiveram DCS de 0,85, sensibilidade de 87,9% no geral e 98,4% para metástases maiores do que 6 mm, superando os resultados

encontrados pela rede neural, que levou cerca de 45 segundos para realizar a mesma tarefa.

Algumas considerações feitas pelo estudo de (LIU et al., 2017) apontam para a dificuldade de realizar a segmentação de metástases pequenas, e como a formação de agrupamentos de metástases podem dificultar o delineamento automático, além de relatarem que é comum uma única aquisição de imagens em RM ponderadas em T1 com contraste, dificultando a aplicação de redes neurais treinadas com conjuntos de dados em outras ponderações ou sequências de pulso. Nesse estudo, os autores utilizaram a arquitetura de rede DeepMedic com o acréscimo de um caminho paralelo adicional, que realizou uma subamostragem de menor resolução para encontrar características globais, com um *kernel* de convolução maior do que o presente nos outros caminhos paralelos. Em relação ao conjunto de dados, foram utilizados dois conjuntos, o primeiro foi o *dataset* Brain Tumor Segmentation (BraTS) (MENZE et al., 2015), com imagens em RM com contraste ponderadas em T1 e em FLAIR, para um total de 240 pacientes, e o segundo foi um conjunto de imagens de ressonância em T1 com contraste, provenientes do *University of Texas Southwestern Medical Center*. Para o treino, foi utilizado um *batch size* de 10 imagens durante 35 épocas para a validação com o BraTS e durante 15 épocas para as imagens obtidas do centro médico. Nota-se que para a validação, o *batch size* foi alterado para 48. Os autores encontraram um DCS de $0.67 \pm 0,03$ e uma AuRROC de $0,98 \pm 0,01$, apontando para uma segmentação aceitável de metástases maiores do que $1,5 \text{ cm}^3$. Assim como (HU et al., 2019; RUDIE et al., 2021) a eficiência da segmentação decai consideravelmente conforme menores os tamanhos das metástases.

A utilização de imagens em RM multimodal foi estudada por (CHARRON et al., 2018) na tarefa de segmentação de metástases cerebrais, em que foi utilizado um *dataset* de imagens ponderadas em 3D T1 CC, FLAIR, e 2D T1, para um número de 182 pacientes, em um treino dividido em 80% para treino, 10% para validação e 10% para teste. Um mapa da distribuição espacial da ocorrência de metástases indicou maior ocorrência nas regiões do cerebelo e lobo frontal. Além disso, foram criados novos casos de 1200 pacientes virtuais para o teste. Os melhores hiperparâmetros para a realização do treino foram investigados, resultando em um BS de 10, com amostras de tamanho de $24 \times 24 \times 24 \text{ mm}^3$, durante 30 épocas, com 20 sub-épocas, para 1000 segmentos e realizado para 50 pacientes por época. As métricas de avaliação utilizadas foram a sensibilidade e o número de falsos positivos para a tarefa de detecção. Para a tarefa de segmentação a métrica utilizada foi o DSC, a sensibilidade e a precisão. Os autores apontam para o cuidado em se realizar o *downsampling* das imagens, uma vez que a compressão dos voxels podem truncar algumas informações radiológicas, como os vasos sanguíneos que aparecem hiperintensos na imagem, de modo que a rede neural os confunda com regiões metastáticas, resultando em um maior número de falsos positivos. Um outro achado do estudo foi a utilização de *batch sizes* pequenos para encontrar melhores valores para DSC, cuja razão é devida à inhomogeneidade do *dataset* e à menor probabilidade de se

ter uma alta variedade de tamanhos de tumores em um menor conjunto de dados, o que permite o treino de um caminho neural adaptado para um tipo de metástase em vez de buscar encontrar um caminho global que generalize o resultado para qualquer tamanho de metástase, que é o que aconteceria em um *batch size* maior, e que geralmente é utilizado em dados mais homogêneos. Entre as descobertas do estudo, destaca-se a utilização multimodal de imagens para o treino da rede, que resultou em uma sensibilidade de 98% para a identificação de metástases, enquanto que o acréscimo de pacientes virtuais não ocasionou melhorias significativas.

(DIKICI et al., 2020) utilizaram uma abordagem de seleção através do Laplaciano da Gaussiana, utilizando essa aproximação ao determinar possíveis candidatos para serem inseridos no modelo de segmentação e, após essa seleção, foram aplicadas deformações elásticas aleatórias e correções gama aleatórias nas imagens para *data augmentation*. A rede utilizada foi a CropNet, que tem como *input* regiões volumétricas isotrópicas de tamanho de 1 mm^3 para cada *voxel*, e o resultado da rede para a sensibilidade foi de 0,90, com uma média de falsos positivos de 9,12 por paciente.

Com o uso de uma rede com base em uma 3D *Fully Connected Network*, (XUE et al., 2020) propuseram a rede BMDS Net para a segmentação de metástases cerebrais, que possui dois estágios, o primeiro para identificação e o segundo para segmentação. Para a identificação, a imagem original de tamanho $256 \times 256 \times 120 \text{ voxels}$ foi diminuída em quatro vezes o tamanho original e, após a detecção, a imagem era redimensionada para seu tamanho original para realizar a segmentação, que consistia no isolamento da região detectada do resto da imagem, através do uso de *bounding-boxes* que seriam os *inputs* da rede de segmentação. Os hiperparâmetros utilizados no treino incluíram uma LR inicial de 0,002 com decaimento de 0,1 por 500 épocas, o valor de *momentum* igual a 0,9 e foi utilizado o inicializador Xavier. O resultado obtido para a acurácia de detecção foi de 100%, e para a segmentação os valores de DSC, sensibilidade e especificidade foram de, respectivamente, $0,85 \pm 0,08$, $0,96 \pm 0,03$ e $0,99 \pm 0,0002$.

(BOUSABARAH et al., 2020) realizaram modificações em uma U-Net para o treino de metástases separadas por tamanho, e compararam os resultados com uma U-Net convencional. Após os treinos, foi comparado os métodos de *ensemble* por soma e por voto de maioria entre as três redes, que foram treinadas por 450 épocas, com LR inicial de 0,0001 e taxa de decaimento de 1% por época, otimizador Adam com regularização L2, BS de 2 e validação cruzada em 4 *folds*. Os autores concluíram que o uso de *ensemble* por soma é melhor para as métricas de sensibilidade (0,82) e DSC (0,74), enquanto que o *ensemble* por voto de maioria apresenta melhores precisão (0,96), taxa de falsos positivos (0,08), *F1-Score* (0,85), e DSC médio (0,71). Além disso, a rede treinada apenas com metástases cerebrais pequenas foi até 50% melhor na segmentação de metástases pequenas quando comparada às redes treinadas com todos os tamanhos de metástases.

Yang *et al.* (2020) desenvolveram uma plataforma *web-based* que realiza a segmentação de metástases cerebrais, e que também realiza operações de pós processamento de imagens (YANG *et al.*, 2020). O modelo é composto por uma arquitetura de rede EnDeepMedic 3D, e os hiperparâmetros não foram informados. Para avaliar o modelo, aplicaram-no em 10 casos de pacientes com metástases cerebrais, e mediram as métricas de média de desvio de centro de massa ($1,55 \pm 0,36$ mm), distância de Hausdorff ($2,98 \pm 0,63$ mm), média da distância entre superfícies ($1,06 \pm 0,31$ mm), a média de falsos positivos sobre a união ($0,43 \pm 0,19$) e a taxa de falsos negativos ($0,15 \pm 0,10$). Como essas métricas de avaliação desse modelo estão muito distoantes das métricas de outros estudos, não permitindo a comparação, essa publicação não foi incluída na análise quantitativa.

(ZHOU *et al.*, 2020) propuseram uma rede chamada MetNet em dois estágios, sendo o primeiro uma rede *single-shot detector* (SSD) e o segundo uma rede FCN. A rede SSD foi responsável pela identificação da metástase enquanto a rede FCN foi utilizada para a segmentação 2D das metástases em uma aplicação parecida com a realizada por (XUE *et al.*, 2020), que realizou a mesma estratégia em três dimensões. No estudo, também foram investigadas três variações da função perda de Tversky e uma Dice e os *inputs* da rede de segmentação possuíam tamanho de 64×64 *pixels*, outros hiperparâmetros do treino não foram informados pelos autores. Os resultados gerais para a precisão, sensibilidade, FPP e DSC foram, respectivamente, $0,58 \pm 0,25$, $0,88 \pm 0,19$, 3 ± 3 e $0,85 \pm 0,13$. Os autores também realizaram testes para metástases de tamanhos distintos, cujos resultados foram, em alguns casos, mais do que 2 vezes melhores para as metástases de tamanhos maiores do que 6 mm quando comparados com os resultados para metástases de tamanhos menores do que 3 mm.

O uso da rede DeepMedic 3D por (PENNIG *et al.*, 2021) com *inputs* de tamanho $25 \times 25 \times 25$ *pixels* formados por *patches* constituintes por regiões cerebrais com e sem metástases balanceados em 50% cada, trouxe resultados para DCS de 0,75, precisão de 0,74, sensibilidade de 0,88 e falsos positivos por paciente de 0,71. Os hiperparâmetros de treino incluíram BS de 10, durante 35 épocas com função perda do tipo *Dice loss*, a LR teve decaimento de 50% com paciência de 3 épocas e seu valor inicial não foi informado. A DeepMedic com *patches* de mesmo tamanho e mesmos hiperparâmetros de rede também foi utilizada por (JÜNGER *et al.*, 2021) para a segmentação de metástases, chegando em resultados de 0,72 para o DCS, 0,851 para a sensibilidade e 1,8 para a FPP, a única diferença em relação ao estudo de (PENNIG *et al.*, 2021) foi o número de pacientes, que antes era 69 e aumentou para 98.

Lu *et al.* (2021) realizaram um estudo para avaliar o impacto do uso de IA na assistência da radioterapia estereotáxica em casos de pacientes com metástases cerebrais (LU *et al.*, 2021). A arquitetura utilizada foi um ensemble da DeepMedic com a U-Net 3D, em que cada uma das arquiteturas possuía um objetivo distinto. A DeepMedic estava focada

em otimizar a sensibilidade para tumores de tamanho pequeno, definido pelos autores em volumes menores do que 0,7 mL ou 1,1 cm de diâmetro, que representa a mediana dos volumes dos tumores de todos os pacientes do banco de dados, composto por 570 pacientes com metástases cerebrais (1790 volumes tumorais), 353 pacientes com neuromas acústicos (363 volumes tumorais) e 365 pacientes com meningiomas (413 volumes tumorais). Por outro lado, a U-Net 3D estava focada em otimizar a especificidade da segmentação de tumores de modo geral, em casos de metástase, meningiomas e neuromas acústicos. O foco do estudo era demonstrar que o uso dessa tecnologia combinada com os profissionais da área da saúde aumentava a concordância entre diferentes observadores, o que foi comprovado em um número de 9 anotadores com diferentes graus de experiência, em que o valor do DSC mediano entre os observadores aumentou de 0,862 sem o uso do modelo para 0,900 com o uso do modelo. Também houve a melhoria no tempo de realização da tarefa de delineamento para a SRS, indo de 11,2 minutos sem o uso da tecnologia para 7,3 com o uso dela. Já para o modelo em si, em uma análise envolvendo 10 casos, sendo 5 de metástases totalizando 16 volumes tumorais, 2 de neuroma acústicos solitários e 3 de meningiomas totalizando 5 volumes tumorais. Sendo assim, sem realizar uma distinção na apresentação dos resultados para cada tipo de tumor em específico, o sistema obteve uma sensibilidade de 95,7% para a detecção da lesão, com uma sensibilidade de 88,9% para tumores pequenos. Além disso, os autores reportam uma quantidade de falsos positivos de 0,5 por paciente, um DSC com mediana de 0,836 e uma AuROCC *pixel-wise* de 0,995.

Nomura *et al.* (2021) utilizaram uma rede U-Net 3D com *inputs* definidos em $32 \times 32 \times 32$ *voxels* e redimensionados de modo a serem representativos de volumes em forma de cubos de 1 mm^3 , sendo que foram comparados os resultados obtidos tanto para os *inputs* formados pelos volumes originais quanto pelos redimensionados (NOMURA *et al.*, 2021). Essa rede foi utilizada em três bancos de dados distintos, dois em TC para nódulos pulmonares e lesões no fígado, e outro em RM para metástases cerebrais. Os hiperparâmetros foram escolhidos a partir de uma série de variações realizadas em um supercomputador, que envolveu a variação da profundidade da rede U-Net de 3 a 5, o número de filtros na primeira cada de 2 a 32, a presença ou ausência da unidade residual, o *batch size* de 1 a 16, a função perda como sendo a Dice ou a Tversky focal e dois parâmetros para o otimizador RAdam. Além disso, foram avaliadas diferentes combinações de técnicas de *Data Augmentation*. Dessa forma, os autores encontraram diferentes composições de hiperparâmetros para as diferentes aplicações, e o modelo que teve um melhor desempenho para a segmentação de metástases cerebrais foi a U-Net com profundidade de 4 blocos de camadas com unidades residuais, 18 filtros na primeira camada, *batch size* de 20, função perda Dice e parâmetros $\alpha = 0,001229$ e $\beta_2 = 0,997552$ do otimizador RAdam, que resultaram em um coeficiente Dice de $0,727 \pm 0,115$ e sensibilidade de $0,766 \pm 0,143$ para o caso dos volumes redimensionados, e a menor taxa de falsos positivos de $0,152 \pm 0,168$ foi obtida para o caso dos volumes em seus tamanhos originais e, em ambos os

casos, a técnica de *Data Augmentation* com melhor performance foi a composição de aguçamento, borramento, ruído, brilho, contraste, perturbação na intensidade, rotação, redimensionamento e deformação, somada à inversão monocromática aleatória.

(PARK et al., 2021b) treinaram uma rede 3D U-Net com *inputs* de tamanho $184 \times 184 \times 152$ *pixels* e combinaram dois modos de aquisição de imagem em RM, Gradiente Eco e Sangue Escuro, para comparar com os treinos realizados apenas com um dos modos. Foi realizada validação cruzada em 5 *folds* com otimizador Adam, durante 500 épocas com LR de 0,0001. Com exceção da métrica de precisão para todas as metástases, o modelo utilizando a combinação dos dois modos de aquisição de imagem foi melhor, alcançando sensibilidade de 0,907, precisão de 0,931 e DSC de $0,822 \pm 0,152$.

(YI et al., 2021) investigaram a performance de uma rede 2.5D DeepLabv3 ao remover uma das sequências de pulso das imagens de RM do *input*, que constou com imagens 2d de tamanho de 256×256 *pixels* para quatro tipos de sequências diferentes das quais foram incluídas 5 *slices* de cada no canal do *input*. Cada sequência de pulso foi removida aleatoriamente, e os valores dos 5 *slices* foram definidos como 0 para o treino. A duração do treino foi de 10 épocas e outros hiperparâmetros não foram informados pelos autores. O resultado obtido para a DSC foi de 0,85, enquanto que para precisão e sensibilidade foram, respectivamente, 0,68 e 0,92.

O estudo realizado por (YOO et al., 2021) utilizou redes U-Net 2.5D e 3D na tarefa de segmentar metástases cerebrais, alcançando um DCS, precisão e sensibilidade de, respectivamente, 0,62, 0,75 e 0,74. Os autores realizaram uma análise de acordo com o número de metástases e de acordo com o volume total de metástases por paciente, e concluíram que a rede 3D é melhor para a segmentação de metástases pequenas, enquanto que a rede 2.5D é superior para a segmentação de metástases maiores. Além disso, foi comparada a performance das redes U-Net com redes *Fully Convolutional One-Stage Monocular 3D*, e ambas as redes U-Net foram superiores na tarefa de segmentação. O treino da rede U-Net 2.5D foi realizado com o otimizador Adam, com regularização por decaimento de pesos desacoplados, tamanho de *mini-batch* igual a 100 e LR com valor de 0,001. Além disso foram utilizados *Dropouts* nas camadas do *encoder* e *decoder* com probabilidade de 0,2 e a função perda utilizada nos *outputs* do *decoder* foi a *cross-entropy*. Por outro lado, o treino da rede U-Net 3D contou com a subamostragem em *patches* formados por *voxels* de tamanho de $96 \times 96 \times 96$, sendo amostrados 8 *voxels* por volume tanto para regiões com lesão quanto para regiões sem lesão, totalizando 40% dos *voxels* obtidos com lesão e 60% dos *voxels* sem lesão. O *dropout* utilizado também foi de 0,2 com a função perda de Dice binária, otimizador Adam, LR de 0,005 e tamanho de *mini-batch* igual a 30.

Park et al. (2021) também empregaram a arquitetura U-Net 2D com base na DenseNet201 com *encoder* pré-treinado no banco de imagens ImageNet, em que utilizaram 3000 imagens de RM cerebrais, e aplicaram *Data Augmentation* de 16 vezes, envolvendo

rotação, inversão, conversão paralela e ajuste de escala (PARK et al., 2021a). Além da 2D, os autores utilizaram a U-Net 3D em um primeiro passo para a localização das metástases, com a função perda Dice para detectar metástases pequenas, em *inputs* de tamanho $192 \times 192 \times 192$ *voxels* em treinos de 200 épocas de duração, com *batch size* de 1 e *learning rate* de 0,001, com redução para 0,0001 a partir da época 100 e 0,00001 a partir da época 150. O resultado obtido pela rede 3D das regiões com presença de metástases, foram então recordados da imagem originais e inseridas na rede 2D, para realizar a segmentação das metástases cerebrais nos *slices* de tamanho 512×512 *pixels*, onde a região com a metástase é redimensionada de modo a caber nessa imagem. Dessa maneira, os autores encontraram uma sensibilidade de 0,58 para a detecção, com DSC médio de $0,67 \pm 0,23$ e 2,5 falsos positivos por exame em um banco de dados parecido com o de treino, que estava separado temporalmente. Para o banco de dados externo, a sensibilidade de detecção foi de 0,76, com DSC médio de $0,66 \pm 0,22$ e 7,6 falsos positivos por exame e, ao realizar a separação para casos de metástases maiores do que 5 milímetros, os resultados foram melhores, sendo 0,877 para sensibilidade de detecção, $0,68 \pm 0,20$ para o DSC médio e 1,9 falso positivo por exame.

(GRØVIK et al., 2021) realizou um novo estudo utilizando uma rede baseada na DeepLabv3, nomeada *Input-Level Dropout*, para a segmentação de metástases cerebrais com a remoção de uma das quatro sequências de pulso obtidas em RM, como uma maneira de lidar com dados faltantes no dataset. Foram utilizados cinco *slices* de cada modalidade de imagem para o treino, que foram acumulados no canal de cor do *input array* que continha dimensão de $256 \times 256 \times 20$. O chamado *Input-Level Dropout* foi aplicado nos 5 *slices* de uma determinada sequência de pulso, cujos valores do tensor foram definidos como 0's. Os resultados obtidos com essa rede para o coeficiente Dice ($0,795 \pm 0,105$), a IoU ($0,561 \pm 0,225$), a sensibilidade ($0,671 \pm 0,262$) e precisão ($0,790 \pm 0,158$) foram superiores quando comparado com a rede DeepLabv3 original, indicando a relevância do método aplicado.

No estudo feito por (HSU et al., 2021) em imagens tanto de RM quanto de TC para a identificação e segmentação de metástases cerebrais, foi utilizada uma rede 3D V-Net, arquitetura parecida com a U-Net, com convoluções de *kernel* $3 \times 3 \times 3$ tanto na parte de *downsampling* quanto na parte de *upsampling*, com *stride* de 2 e função de ativação *parametric rectified linear unit* (PReLU), com *batch normalisation* após cada camada de convolução e pesos dos *kernels* regularizados com o iniciador Xavier. Para o treino em questão, o volume de *input* foi reduzido para 48 mm^3 e o número de *feature channels* foi aumentado para 24, bem como o número dos *strides* que passaram a ser 3, 3, 3, 3 e 4 conforme aumentava-se o *downsampling*, otimizador Adam com LR de 0,001 com *Early stopping*. Além disso, foi utilizado *data augmentation* e transformações nas imagens envolvendo a equalização do histograma com base em outro paciente aleatório, deformação *B-Spline* aleatória e rotação 3D aleatória de até 20° . Os autores apontaram

para o uso da função perda *soft Dice*, uma vez que os dados estavam muito desbalanceados com menos do que 0,1% dos voxels contendo metástases, e indicaram que nesse cenário a função perda *cross-entropy* não deveria ser utilizada.

A utilização de uma mistura de redes U-Net 2D e 3D também foi utilizada por (CHO et al., 2021) para a detecção e segmentação de metástases cerebrais. Nesse modelo, a rede 3D foi usada para a identificação e aumento da dimensionalidade das metástases identificadas para que a rede 2D fizesse a segmentação. Além disso, o treino foi dividido em três tipos de validação, a primeira com pacientes deslocados temporalmente entre os anos de 2015 e 2016, a segunda com pacientes de outra instituição e a terceira com pacientes dos anos entre 2017 e 2020, com deslocamento temporal maior do que o primeiro conjunto. Dessa forma, as métricas de avaliação foram melhores para os pacientes do terceiro conjunto, com valores de sensibilidade de 0,947, a quantidade de falsos positivos encontrada de 0,5 por paciente e o DCS de $0,82 \pm 0,20$. Vale notar que a análise foi realizada para metástases de tamanhos maiores do que 5 mm e os parâmetros utilizados para o treino da rede de identificação, a 3D U-Net, incluíram a função perda *Dice Loss* para *inputs* de tamanho $192 \times 192 \times 192$ com otimizador Adam, BS de 1 e LR com valor inicial de 0,001 durante 300 épocas. Nota-se que entre as épocas 100 e 250, a LR utilizada foi de 0,0001 e após a época 250 a LR foi de 0,00001. Já para o treino da rede de segmentação, a 2D U-Net, os parâmetros utilizados foram *input* de 512×512 pixels com *data augmentation* de 16 vezes, e os outros hiperparâmetros não foram mencionados na publicação.

(SUN et al., 2021) realizaram uma comparação entre os resultados obtidos entre as redes U-Net 2D e 3D para a segmentação de metástases cerebrais em um banco de dados próprio de 450 pacientes. As imagens utilizadas passaram por *data augmentation* e regularização L2 para as duas arquiteturas. A métrica estudada foi a AuROCC que se mostrou maior para a rede 3D com regularização para todos os tamanhos de metástases (0,876), seguida da rede 2D com regularização (0,815). A rede sem regularização 3D foi melhor para a segmentação de metástases com volume acima de $1,37 \text{ cm}^3$ (AuROCC = 0,905) quando comparada com a rede sem regularização 2D U-Net (AuROCC = 0,88), essa que se mostrou superior para a segmentação de metástases menores (AuROCC = 0,759) contra a AuROCC de 0,714 da rede 3D. Dessa maneira, aponta-se para a importância da técnica de regularização L2 para uma melhor AuROCC na tarefa de segmentação de metástases cerebrais, bem como a relevância das características espaciais obtidas pela rede 3D para uma melhor segmentação.

Com o intuito de diminuir a quantidade de falsos positivos na segmentação de metástases cerebrais, Park *et al.* (2022) realizaram um *ensemble* entre um modelo de *Deep Learning* e um classificador radiômico, consistindo de uma rede En-DeepMedic para a segmentação e uma rede neural siamesa junta de um classificador SVM radiômico para a

classificação (YANG et al., 2022). Esse *ensemble* foi treinado em um conjunto de 242 MCs com imagens em T1 pós contraste, e os autores reportaram valores de 0,13 para a taxa de falsos negativos e 0,09 para a métrica de falsos positivos sobre a união (FPoU), essa que foi reduzida em cerca de seis vezes ao se aplicar o modelo de classificação adicionado à rede En-DeepMedic, que não sofreu alterações em seu formato padrão. Assim, os autores tiveram ênfase no estudo da rede siamesa e afirmam que seu uso em conjunto com um classificador SVM podem diminuir a quantidade de falsos positivos em segmentações de metástases cerebrais.

Uma alteração na arquitetura *DeepMedic* proposta por (HUANG et al., 2022) aumentou significativamente as métricas de sensibilidade, de 0,853 para 0,975 e de precisão, de 0,691 para 0,987. A alteração foi o uso de uma função perda específica para o caso de metástases cerebrais, chamada de *Volume-level Sensitivity-Specificity* (VSS), e foi empregada no lugar da *Binary Cross-Entropy* para a tarefa de identificação. Essa função perda foi criada para mitigar a variabilidade dos resultados obtidos para metástases de tamanhos diversos, uma vez que a função perda para metástases pequenas não possui grande impacto no conjunto todo. Assim, foi realizada o *ensemble* das duas funções perdas para a tarefa de segmentação, que contou com um número de 60 épocas, com LR inicial de 0,001 e otimizador *Root Mean Squared Propagation* (RMSProp) com momento de Nesterov ($m = 0,6$, $\rho = 0,9$ e $\epsilon = 0,0001$). Os dados de input possuíam tamanho de $37 \times 37 \times 37$ com 9 camadas de convolução de *kernels* de tamanho $3 \times 3 \times 3$ sem *padding*, de modo que o *output* do modelo tinha o tamanho de $19 \times 19 \times 19$ *pixels*, e os segmentos passaram por *data augmentation* envolvendo mudança aleatória de intensidade, inversão e rotação.

Para a segmentação de metástases pequenas, com volumes menores do que $0,04 \text{ cm}^3$, uma rede foi proposta por (YOO et al., 2022) com base nas redes U-Net 2.5D e 2D. As imagens de tamanho original de 1024×1024 *pixels* foram recortadas em *patches* de 128×128 *pixels* com sobreposição de *patches* e um espaçamento de 64 *pixels* entre as sobreposições. Técnicas de *data augmentation* incluíram inversão vertical e horizontal, bem como rotação e borrimento aleatórios. A rede U-Net teve alteração na função de ativação da habitual ReLU para a *Exponential Linear Unit* (ELU), para evitar o gradiente nulo, com *Batch Normalisation* antes da função de ativação ELU. A LR inicial foi de 0,0001 com redução de 50% para cada vez que a métrica acompanhada não mudasse em duas épocas consecutivas, em um total de 30 épocas, com BS de 256, otimizador Adam e regularização L2 ($\beta = 0,00001$). Os resultados obtidos para a segmentação teve uma sensibilidade de 0,966, com coeficiente Dice de 0,55 e um número médio de FP de 1,25 por paciente. Vale ressaltar que o conjunto de dados utilizados consistia majoritariamente de metástases de tamanhos menores do que $0,1 \text{ cm}^3$, e a discussão desse trabalho é relevante para o estudo de metástases de tamanhos pequenos, cuja identificação é um desafio para as pesquisas realizadas nessa área.

Para a realização independente de segmentação de quatro tipos de tumores cerebrais: glioblastomas, gliomas de baixo grau, meningiomas e metástases, (BOUGET et al., 2022) utilizaram uma arquitetura de rede com base na AGU-Net e analisaram a correlação de 25 métricas distintas para avaliar a performance dos modelos, apontando para a relevância das métricas de diferença de volume e distâncias de Hausdorff, Mahalanobis e de superfície simétrica média do objeto. Em relação à rede, foi utilizada a função perda *class-averaged Dice loss*, com otimizador Adam e BS de 32, paciência de 30 épocas para *early stopping* e LR de 0,001, ou LR de 0,0001 quando aplicado *transfer learning* para um *fine-tuning* do melhor modelo encontrado para a segmentação de glioblastomas. Foi utilizado *data augmentation* com probabilidade de 50% inversão vertical e horizontal, rotação aleatória de 20° e translação de até 10% nos eixos. A eficiência da rede se mostrou melhor para a segmentação de metástases e glioblastomas para todas as métricas analisadas. Para metástases, a rede alcançou um valor de $0,8773 \pm 0,1894$ para o DSC e $0,8156 \pm 0,2042$ para IoU, sendo ambas as melhores métricas encontradas entre os diferentes tipos de tumor.

No estudo realizado por Ceaulescu e Alexe (2022), o *BrainMetShare* foi utilizado em uma aplicação em dois estágios (CEAUŞESCU; ALEXE, 2022). No primeiro estágio, foi utilizada técnica de *transfer learning* com os pesos de uma DR-UNet104 obtidos em um treino realizado com imagens de gliomas do BraTS e foi feito o *fine tuning* com imagens de metástases cerebrais. No segundo estágio, foi utilizado o GrabCut para refinar as máscaras de segmentação obtidas no primeiro estágio. Sendo assim, os autores encontraram valores para o coeficiente Dice em quatro composições de parâmetros distintas que foram de 85,03% para o modelo padrão, 88,24% para o modelo pré-treinado, 89,23% para o modelo pré-treinado com o uso do fator $p = 1$ para o GrabCut e 90,07% nessas mesmas condições com o fator p otimizado em seu valor de 1,22. Dessa forma, os autores concluem que o uso desses dois estágios melhora a performance de segmentação quando comparado ao treino da DR-UNet104 realizado apenas com o banco de dados BMS.

Shirokikh et al. (2022) avaliaram uma rede U-Net 3D na tarefa de segmentação de metástases cerebrais para aplicação em radiocirurgia. Essa rede passou por alterações nos blocos convolucionais, que foram substituídos por blocos residuais e evitaram o uso de concatenação para os canais, em que realizaram a soma, para diminuir a carga computacional (SHIROKIKH et al., 2022). Além disso os tamanhos e quantidade de filtros foram reduzidos, uma vez que os autores afirmam que mudanças na arquitetura não geram melhorias na qualidade da rede. Para o treino, foi utilizada a função perda BCE com pesagem inversa para dar ênfase em segmentações de volumes pequenos. O treino foi realizado durante 100 épocas, com *learning rate* de 0,01 que diminuiu para 0,001 a partir da época de número 80. Além disso, cada época contava com 100 iterações do otimizador SGD com momento Nesterov de 0,9, para um *batch size* de 16 *patches* de tamanho $64 \times 64 \times 64$ *pixels*. Nesse estudo, os autores encontraram um valor de sensibilidade de

0,9 com 3,8 falsos positivos por exame, além de um DSC por metástase de 0,71. Entre os resultados obtidos, os autores também apontam para uma redução de 48% dos desacordos entre diferentes especialistas ao mesmo tempo que melhora o coeficiente Dice de superfície de 0,845 para 0,871. Além de diminuir a variabilidade entre diferentes especialistas, o uso do modelo acelera o processo de delineamento em duas vezes.

(LIANG et al., 2022) estudaram diversas variantes de 3D-Unets para a tarefa de segmentação de metástases cerebrais, e relatam que um número de *kernels* de 128 com tamanho de *input* de $64 \times 64 \times 64$ *voxels* demonstrou os melhores resultados em uma validação cruzada de 5-*folds*. No conjunto de teste que contava com 327 metástases, o modelo alcançou um DSC de 0,73, com sensibilidade de 0,91 e taxa de falsos positivos de 1,7, e indicam que o modelo possui melhor performance para a segmentação de metástases que possuem mais que 12 mm de diâmetro.

(LYU et al., 2022) realizaram a segmentação de metástases com uma variante da U-Net, e compararam seus resultados com as redes U-Net, AttU-Net e U-Net++, e encontraram melhores resultados para o DSC ($0,878 \pm 0,297$), sensibilidade ($0,932 \pm 0,363$) e especificidade ($0,999 \pm 0,004$) com o modelo que propuseram. O modelo foi treinado com duas modalidades de imagens de ressonância magnética, e foi utilizada uma rede cycleGAN para gerar as imagens faltantes em caso de haver apenas aquisição em uma das modalidades. Foi utilizado o otimizador Adam, com LR de 0,0001, e BS de 12, sendo aplicada a função de ativação ReLU nas camadas de convolução. A *callback* utilizada foi a de *loss decay* com margem de variação de pelo menos 1% após 20 épocas consecutivas, e a função perda analisada foi uma combinação da *cross-entropy* com a Dice *loss*.

Uma outra pesquisa envolvendo o *BrainMetShare*, em conjunto com o BrATS e um *dataset* adquirido localmente, foi realizada por (LIEW et al., 2022) para a segmentação de metástases cerebrais. Eles utilizaram uma arquitetura de rede com base em uma N-Net, que é uma versão modificada da U-Net, e fizeram algumas alterações nas conexões entre as camadas, como a adição de blocos não-locais. O treino foi realizado com *patches* de $96 \times 96 \times 96$ *pixels* com janela de deslocamento de 32 *pixels*, um BS de 4, uma LR adaptativa com valor inicial de 0,001 para as 75 primeiras épocas, aumentando para 0,1 nas 45 épocas seguintes, e 0,01 nas 30 épocas finais, totalizando 150 épocas. A função perda utilizada foi uma combinação da *mean-squared error* e a *binary cross-entropy*. Após realizarem o treino para os três *datasets*, foi aplicado um *framework* iterativo para o *self-training* gradual do modelo, denominado CAVEAT, os resultados obtidos para a rede foram otimizados, e incluíram uma sensibilidade de 0,740 para o *dataset* local, e 0,811 para o *BrainMetShare*, sendo que a quantidade de falsos positivos encontrados por paciente foi de 3,131 e 2,952 para esses mesmos *datasets*, respectivamente.

Ao utilizarem uma combinação de funções perda Dice, *boundary* e *Volume-Aware*, (CHARTRAND et al., 2022) investigaram as diferentes combinações nos resultados para

a segmentação de metástases cerebrais utilizando uma rede de arquitetura 3D U-Net. Para o treino, foram utilizados *patches* de tamanho de $128 \times 128 \times 128$ *pixels*, com uso do otimizador Adam, com LR de 0,001 que era dividida pela metade a cada 50 épocas sem variações, durante um total de 400 épocas. Em relação às transformações nas imagens, foram aplicadas rotações aleatórias, ruído gaussiano e correção gama de contraste. Os valores encontrados para a sensibilidade foram de 0,909, com DSC de 0,73 e 0,66 falsos positivos por paciente. Os autores também apontaram para a piora dos resultados das métricas conforme o tamanho das metástases era menor e indicam a melhora da segmentação quando utilizada uma combinação de função perda *Dice* com *Volume-Aware sampling*, que realiza a amostragem com base em uma probabilidade de escolher a lesão sendo inversamente proporcional ao seu tamanho, para mitigar os efeitos de desbalanceamento de classes para metástases pequenas.

Uma aplicação para a predição de toxicidades na radioterapia feita por (JALALIFAR et al., 2022) a partir da união de radiômica com a segmentação de metástases cerebrais realizadas com uma rede de arquitetura formada pela combinação de duas 2D U-Nets, duas 3D U-Nets e uma MSGA, encontrou resultados que foram melhores do que cada uma das redes individuais. Para a primeira 2D U-Net, os *inputs* possuíam tamanho de 512×512 *pixels* e 256×256 *pixels* para a segunda 2D U-Net. Já para as 3D-U-nets, os *voxels* do *input* possuíam tamanho de $128 \times 128 \times 128$ *pixels*. Os resultados obtidos incluíram um DSC de $0,90 \pm 0,04$ e uma distância de Hausdorff de $2,3 \pm 0,55$ mm para o conjunto de teste. Os autores também publicaram outro estudo (JALALIFAR et al., 2023) considerando apenas as imagens em RM para a predição de efeitos adversos após a radiocirurgia, e utilizaram a mesma arquitetura de rede com valores de BS de 4, 1 e 1 para as redes 2D U-Net, 3D U-Net e MSGA, com respectivas funções perda *Dice*, *cross-entropy* e a soma das duas. A LR de 0,0001 foi a mesma para as três redes e os resultados para a segmentação foram superiores quando comparados com as redes individuais e uma nnU-Net tida como referência, apresentando resultados para DSC de $0,915 \pm 0,037$ e distância de Hausdorff de $2,1 \pm 0,6$ mm, que também foi superior em relação ao primeiro estudo realizado. Quanto à segmentação realizada nos 5 *follow-ups* após a cirurgia, a rede proposta pelos autores foi superior com exceção do 4^o *follow-up*, quando a rede nnU-Net foi menos do que de 1% melhor nas métricas avaliadas. Além disso, os resultados obtidos nesse segundo estudo foram melhores apenas para a métrica do coeficiente Dice quando comparado ao estudo feito com um *dataset* menor (JALALIFAR et al., 2020), que apresentou uma distância de Hausdorff de $1,12 \pm 0,03$ mm ao ser utilizada uma rede composta por U-Nets 2D e 3D.

(LEE et al., 2022) implementaram uma rede CNN *encoder-decoder* de via dupla para realizar o delineamento de três tipos de tumores cerebrais: schwannoma vestibular, metástases e meningioma. O modelo treinado para delinear metástases cerebrais contou com 100 épocas de treino, com otimizador Adam, e LR de 0,001, com imagens formadas

por *patches* de tamanho $256 \times 256 \times 20$ *pixels* como *input* e BS de 1. A função perda utilizada foi a Dice generalizada, com fator $\epsilon = 0,00001$. Ao utilizarem os parâmetros de treinos das outras duas patologias no treino para as metástases, foram encontrados resultados otimizados, cujos valores para DSC de $0,85 \pm 0,18$, sensibilidade de $0,87 \pm 0,29$ e precisão de $0,85 \pm 0,18$. Os resultados obtidos para os outros tipos de tumores foram concordantes no que diz respeito a se utilizar os parâmetros de treino cruzados e podem ser importantes para a generabilidade dos modelos treinados.

(OTTESEN et al., 2023) realizaram uma comparação entre três diferentes CNNs para a segmentação de metástases cerebrais treinadas com o dataset *BrainMetShare* e testadas em um dataset composto de imagens de ressonância magnética em ponderação T_1 *Fast Spin Echo* (SPACE) e FLAIR anotadas por dois radiologistas com 5 e 14 anos de experiência. As arquiteturas de rede utilizadas foram a 2.5D e a 3D HRNetV2, do inglês *high-resolution net V2*, e a nnU-Net, e utilizaram uma mistura de duas funções perda, a *Focal Tversky* e a *binary cross-entropy*. Seus resultados apontam para uma maior eficiência da rede 2.5D em relação à rede 3D e nnU-Net, identificando, respectivamente, 79%, 71% e 65% das metástases. Para a tarefa de segmentação, as redes apresentaram resultados similares, a nnU-Net com DSC de $0,94 \pm 0,05$ e as duas HRNetV2 com DSC de $0,93 \pm 0,04$. Já para o dataset *BrainMetShare*, a eficiência de detecção de metástases foi de, 88%, 86% e 78% com sensibilidades de 0,92, 0,91 e 0,85, respectivamente. Vale ressaltar que a acurácia de detecção de metástases decresce quando essas possuem área de seção menor do que $0,4 \text{ cm}^2$.

(LI et al., 2023) utilizaram duas modalidades de imagens em RM, com e sem contraste, e realizaram a subtração entre essas imagens de modo a formar três canais para o *input* em *patches* de tamanho $128 \times 512 \times 512$ *pixels*, com *data augmentation* envolvendo rotações, inversões e variação na intensidade aleatórias. A rede utilizada possuía dois estágios, o primeiro sendo constituído de uma rede 3D U-Net para a segmentação de metástases e o segundo uma CNN multi-níveis para a supressão de falsos positivos. Foi utilizado o otimizador Adam, com a união das funções perda *Bias Dice* e *cross-entropy*, durante 500 épocas, com LR de 0,001 e BS de 1. Os resultados obtidos por essa rede foram, em geral, superiores quando comparados às redes SSD, nnU-Net e 3D U-net, sendo que essa última provou-se superior para a segmentação de metástases de diâmetros maiores do que 5 mm.

A afirmação de que um protocolo de aquisição de imagens somente com a sequência T1 pós contraste é suficiente para a segmentação automática de metástases cerebrais também foi sustentada pelos achados de Buchner *et al.* (2022), que utilizou diferentes combinações de sequências como *input* da rede neural U-Net 3D e encontrou um resultado de DSC de 0,96, superior aos resultados de 0,70 e 0,73 utilizando apenas as sequências T1 pré-contraste e T2 FLAIR (BUCHNER et al., 2022). Para o treino desse modelo, foi

utilizado *Data Augmentation* envolvendo inversões horizontais, ruído gaussiano e transformações *affine* aleatórias, com a função de perda composta pela soma da Dice e BCE por um total de 500 épocas. Os autores também concluíram que, nos casos em que a região de edema é importante para a segmentação, é recomendado o uso da sequência FLAIR em conjunto com a sequência T1 pós contraste.

Utilizando um banco de dados com imagens em T1 pós contraste de 1000 pacientes diagnosticados com um total de 11686 metástases cerebrais, Qu *et al.* (2023) desenvolveram uma rede *autoencoder* com o uso de uma camada *Atrous Spatial Pyramid Pooling* após a última camada de *Max Pooling* (QU *et al.*, 2023). Também foram utilizadas camadas de concatenação e um módulo de *Attention*. O otimizador utilizado foi o Adam, com *learning rate* inicial de 0,0001, que decaía a cada 8 épocas com uma taxa de 0,98. Além disso foi utilizado um *mini batch size* de 32 em uma duração de 150 épocas. Entre os resultados, os autores encontraram uma sensibilidade de 0,88, DSC de 0,90, precisão de 0,93 e com 1,01 falsos positivos por paciente. Para o teste externo, os resultados foram similares, garantindo a generabilidade do modelo. Além disso, o modelo foi utilizado por quatro radiologistas, que tiveram uma melhoria de até 27% na taxa de detecção de metástases menores do que 5 mm.

A avaliação do algoritmo VBrain realizada por Wang *et al.* (2023) teve como finalidade verificar sua performance em 100 pacientes com diferentes características demográficas (WANG *et al.*, 2023). Nessa avaliação, foram utilizadas imagens em TC e em RM na ponderação T1 pós contraste. Para um número mediano de 2 lesões de tamanhos medianos de 0,112 mLs, e com 56% dos tumores primários sendo de pulmão, os autores encontraram um valor de DSC por lesão médio de 0,723, com HD de 7,34% do tamanho do tumor e com um número de falsos positivos de 0,72 por paciente. Além disso, a sensibilidade encontrada foi de 99,07%, 97,59% e 96,2% para tumores de tamanhos maiores ou iguais a 10 mm, 7,5 mm e 5 mm, respectivamente. Assim, os autores confirmam que esse modelo, já aprovado pela *Food and Drug Administration*, está apto para uso na prática clínica.

Em uma abordagem diferente, ao utilizar um modelo *hypergraph-based neural-like* (HNN) *P system* inspirado nos modelos *autoencoder*, Xue *et al.* (2023) desenvolveram um modelo que realiza a implementação de 3 regras para a segmentação de imagens médicas, que envolvem a atribuição de valores de classe para cada *pixel* e, após inicialização das imagens e *labels*, o HNN P System realiza operações similares às da U-Net, em que um mecanismo de *attention* é utilizado, junto à função de perda CE (XUE *et al.*, 2023). A última regra do modelo é o acúmulo de *spikes* para os *outputs* da rede, até um número previamente definido e, após votação por maioria, o resultado final é obtido. Para esse treino, foi utilizado o otimizador Adam, com *learning rate* de 0,003 e decaimento exponencial após 500 iterações, em um total de 80000 iterações. Foi utilizado *dropout* de 0,7,

um *batch size* de 8 e validação cruzada em 4 *folds*. Dessa forma, os autores chegaram em resultados de DSC de $0,836 \pm 0,054$ para a segmentação de metástases cerebrais, com uma precisão de $0,848 \pm 0,074$ e sensibilidade de $0,836 \pm 0,087$, superando outros modelos da literatura.

Shu *et al.* (2023) desenvolveram uma rede chamada *Deep Slice-Crossed* com o intuito de melhorar a acurácia da segmentação de micrometástases. Nessa rede 2D, foi utilizada uma função de perda denominada *Local Weighted Loss*, que coloca pesos diferentes para tamanhos diferentes de lesões, de modo a aumentar a influência de metástases de menor tamanho (SHU *et al.*, 2023). Sendo assim, os autores realizaram o treino com o otimizador Adam por 200 épocas e com *callback early stopping* com paciência de 20 épocas. O *batch size* foi de 48 e *weight decay* de 5×10^{-4} , esse também foi o valor para a *learning rate* inicial, que teve taxa de decaimento de 0,98 a cada 4 épocas. Foi utilizado *Data Augmentation* de ampliação e inversão aleatórios. O tamanho do *input* foi definido como 256×256 e as imagens foram redimensionadas para esse tamanho antes de entrarem no modelo. De modo geral, o modelo obteve um resultado de 0,896 para o coeficiente Dice, com sensibilidade de 0,90 e precisão de 0,892. Além disso, os autores avaliaram diferentes funções de perda no banco de dados BMS, e concluíram que a função proposta foi superior em até 3% comparada às outras funções utilizadas em outros estudos, como a Dice ou a Tversky.

Combinando as informações do status de mutação do receptor do fator de crescimento epidérmico e segmentação do GTV, Zhou *et al.* (2023) propõem uma rede *multi-scale self-attention encoder* seguida por uma rede híbrida formada por uma CNN e um *Transformer* para a multi-tarefa de segmentar o tumor e classificá-lo em seu status de mutação (ZHOU *et al.*, 2023). A função de perda multi-tarefas também foi empregada, de modo a otimizar as duas tarefas simultaneamente. Essa função consiste na combinação das funções de perda Dice somada à função de entropia cruzada focal com ponderação de 4×10^{-5} para a segmentação e a função perda de entropia cruzada para a tarefa de classificação. Essa combinação possui parâmetros que são ajustados através do método baseado em incerteza, que divide o valor de cada função de perda pelo quadrado dos parâmetros de treino de cada tarefa e soma o valor do logaritmo desses parâmetros multiplicados. Assim, foi utilizado o otimizador Adam com *learning rate* de 3×10^{-4} com *early stopping* de paciência de 20 épocas. Dessa maneira, os autores encontraram um valor para o DSC de 0,89 para a segmentação do GTV no teste interno e 0,85 para o teste externo. Além disso, para o teste interno, os autores encontraram uma HD de 3,58 mm, precisão de 0,91 e sensibilidade de 0,90.

Com a intenção de minimizar o número de falsos positivos, utilizar uma quantidade menor de sequências em RM e dar informações sobre as regiões de necrose tumoral, Son *et al.* (2023) desenvolveram a rede RLK-UNet, que foi treinada com imagens em RM na

ponderação T1 *Black Blood* de 128 pacientes, e foi avaliada em um conjunto de 58 pacientes para as métricas de sensibilidade, precisão e DSC (SON et al., 2023). Esse modelo tem como base uma rede U-Net 3D com a adição de *multiscale highlighting foreground* e pós processamento para a remoção do plexo coroide e vasos sanguíneos, foi utilizada validação em 5 *folds* e . Além disso, o modelo foi capaz de avaliar à resposta ao tratamento, que concordou em 84% com os neurorradiologistas. Sendo assim, a sensibilidade encontrada foi de $86,90 \pm 4,07$, precisão de $79,60 \pm 6,46$, DSC de $0,66 \pm 0,02$ e com $1,76 \pm 0,22$ falsos positivos por exame.

Pareando as imagens separadas temporalmente em diferentes canais da imagem do *input*, Hammer et al. (2024) salientam a importância de se obterem imagens de *follow-up* dos pacientes que realizaram tratamento via SRS. Em seu estudo, os autores utilizaram uma rede U-Net 3D denominada SimU-Net, que possui três canais, utilizando as duas imagens separadas temporalmente em um período de no máximo 12 anos, e a máscara de segmentação do primeiro exame. Os hiperparâmetros de treino não foram informados pelos autores, que encontraram um valor para DSC de $0,90 \pm 0,10$, precisão de $0,76 \pm 0,27$ e sensibilidade de $0,94 \pm 0,16$ (HAMMER et al., 2024).

Prezelski et al. (2024) utilizam um algoritmo denominado METRO para acompanhar as metástases cerebrais de pacientes que realizaram *follow-up* após tratamento por SRS (PREZELSKI et al., 2024). Esse algoritmo foi aplicado em imagens de RM ponderadas em T1 pré e pós contraste de 71 pacientes e tem como base uma CNN 3D. O foco do estudo foi realizar a comparação uma a uma das segmentações do modelo com as segmentações manuais realizadas pelos autores e foram avaliadas a classificação dos volumes tumorais como estáveis, em crescimento, em decréscimo, imperceptíveis e sem resposta. Além das métricas de classificação, não foram apresentadas métricas de performance de segmentação do modelo.

5 Materiais e Métodos

Esse estudo envolve a identificação e segmentação metástases cerebrais através de Modelos de *Deep Learning* com imagens de RM. A Figura 11 ilustra um panorama geral da pesquisa desenvolvida, onde é possível observar o fluxo de trabalho e as análises realizadas.

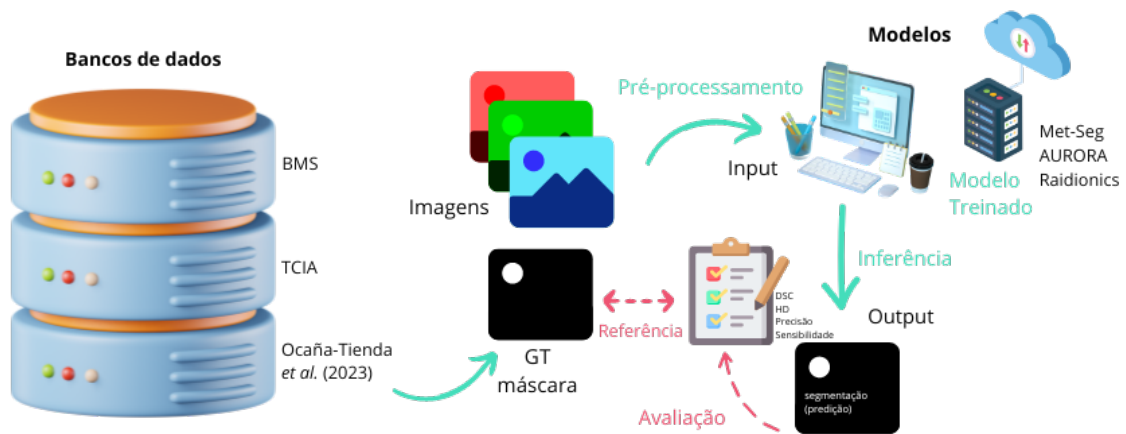


Figura 11 – Fluxo de trabalho e análises realizadas nessa pesquisa.

Fonte: Autoria própria (2024)

Sendo assim, os bancos de dados utilizados nesta pesquisa são constituídos de pacientes diagnosticados com metástases cerebrais submetidos a tratamentos de radioterapia.

5.1 Bancos de dados

O primeiro banco de dados apresentado é o *BrainMetShare* (BMS) (GRØVIK et al., 2020), que foi o único banco de imagens em RM disponível publicamente com ênfase nas metástases cerebrais até a publicação de Ocaña-Tienda et al. (2023) e Ramakrishnan et al. (2024), que são os outros dois bancos de dados utilizados nesse estudo, denominados por Ocaña-Tienda et al. (2023) e TCIA, respectivamente. Cada um dos três bancos de dados será detalhado nas próximas subseções.

5.1.1 *BrainMetShare* (BMS)

A principal base de dados deste trabalho é a *BrainMetShare*, BMS, disponibilizada pela *Stanford University School of Medicine* (GRØVIK et al., 2020), que contém imagens em RM de crânio de 156 pacientes, em quatro modalidades de aquisição, sendo elas: 3D Cube (T1 pré contraste), 3D Cube CC (T1 pós contraste), BRAVO e FLAIR. Todas as imagens foram pré-processadas para extração do cérebro, isto é, remoção da região do

crânio e pele e, para 105 dos 156 pacientes, as máscaras de segmentação dos sítios metastáticos foram realizadas em verificação cruzada por dois neurorradiologistas com 2 e 8 anos de experiência. Para os outros 51 pacientes, as máscaras de segmentação não foram disponibilizadas, para que fossem feitas anotações por outros profissionais para testar o modelo, e não foram utilizadas nesse estudo. Sendo assim, foram utilizadas 72756 imagens relativas aos *slices* da imagem de RM dos 105 pacientes, que é constituído de um conjunto de imagens 2D em sequência através de todo o volume cerebral. Esses *slices* estão todos corrigidos para cada paciente, isto é, a numeração de cada *slice* para cada sequência representa a mesma região da posição espacial real do cérebro, o que é de grande importância para a aplicação do modelo, que necessita que as coordenadas virtuais dos tensores que formam a imagem correspondam às mesmas coordenadas espaciais do cérebro. Um exemplo desses *slices* corrigidos do banco de dados *BrainMetShare* está apresentado na Figura 12, em que o mesmo índice de *slice* para o mesmo paciente representa o mesmo corte transversal cerebral.

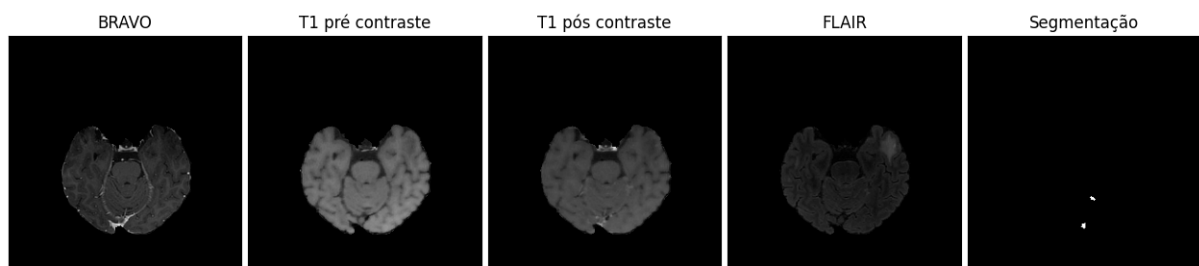


Figura 12 – Exemplos das imagens em PNG do banco *BrainMetShare*.

Fonte: *BrainMetShare*, (2020)

Esse conjunto de dados será analisado e utilizado no desenvolvimento deste trabalho. Ademais, outras variáveis adicionais às imagens podem ser interessantes para o treino de um modelo multimodal, como sexo, origem da metástase, número de nódulos metastáticos, tamanho médio dos nódulos metastáticos, proporção de nódulos maiores do que 10 mm, modelo da máquina de RM, Intensidade em Tesla do campo de RM, resolução planar da RM, espessura da camada de RM em milímetros, idade, presença de sintomas, status da doença primária, presença ou ausência de metástases em outros órgãos, status neurológico do paciente, condição média do paciente, volume estimado do tumor cerebral, número de tumores, histórico de WBRT e histórico de procedimento no cérebro (NIRANJAN et al., 2019; CHO et al., 2020; GRØVIK et al., 2020).

5.1.2 *The Cancer Imaging Archive* (TCIA)

O banco de dados apresentado por Ramakrishnan *et al.* (2024) contém arquivos de imagens em ressonância magnética 3D para 200 pacientes diagnosticados com metástases cerebrais (RAMAKRISHNAN et al., 2024). Esse banco de dados é composto por imagens de RM em formato *Neuroimaging Informatics Technology Initiative* (NIfTI) adquiridas

em seqüências FLAIR, T2, T1 antes e T1 depois da aplicação de contraste, além de possuir as máscaras de segmentação realizadas por um estudante de medicina e validadas por dois neurorradiologistas com mais de 7 anos de experiência. Os autores informam que o número total de metástases desse dataset é 975, que foram adquiridas em diferentes aparelhos, com diferentes intensidades de campo magnético indo de 1 Tesla até 3 Teslas. Nesse banco de dados, são feitas três segmentações que correspondem à região de edema peritumoral, identificado pela seqüência FLAIR, a região de porções do tumor que estão em estado de necrose, que não absorve contraste, e a região da lesão com absorção de contraste, que geralmente envolve a região com necrose, conforme representado na Figura 13. Vale ressaltar que nos arquivos NIfTI, as *labels* 0, 1, 2 e 3 das diferentes classes da segmentação representam, respectivamente o fundo preto (BG), a região do tumor com necrose, a região do edema peritumoral e a região tumoral com absorção de contraste.

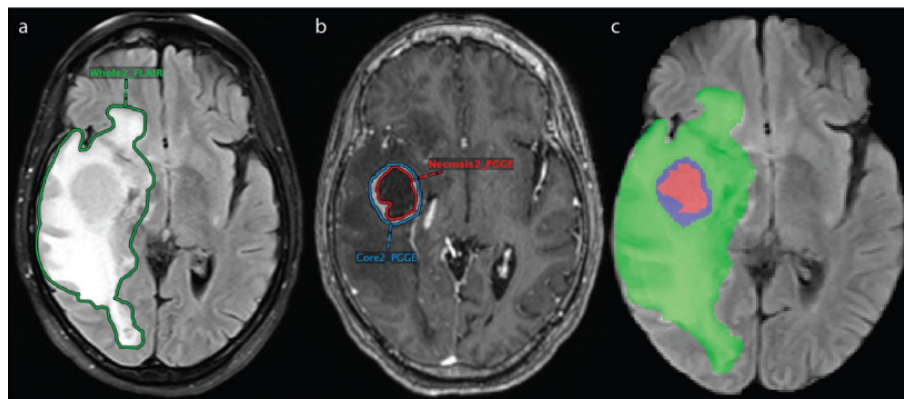


Figura 13 – Exemplos das segmentações encontradas no *dataset* do TCIA, em que (a) é a região com edema peritumoral (em verde) encontrada na seqüência FLAIR, (b) possui as marcações encontradas em T1 pós contraste para a região com necrose (em vermelho) e a região tumoral com absorção de contraste (em azul) e em (c) encontram-se as máscaras de segmentação.

Fonte: Ramankrishnan *et al.* (2024)

5.1.3 Ocaña-Tienda *et al.* (2023)

O conjunto de dados disponibilizado por (OCAÑA-TIENDA *et al.*, 2023) é composto por 373 imagens em RM de 75 pacientes com um total de 260 volumes tumorais, todas em formato NIfTI. Esse número de volumes tumorais são relativos aos 75 pacientes e não para o conjunto de 373 imagens, cuja quantidade de volumes tumorais é de 591. Nesse banco de dados, todas as imagens foram obtidas utilizando-se a seqüência T1 com contraste e as máscaras de segmentação de 154 dessas metástases foram realizadas semi-automaticamente pela autora com auxílio de um *software*, e foram revisadas por três profissionais com mais de 7 anos de experiência em RM, além de serem corrigidas por um dos radiologistas da equipe de pesquisa. Além disso, o banco de dados fornece informações morfológicas, radiômicas e de aquisição da imagem. Um exemplo de um *slice* e uma má-

cara de segmentação do banco de dados de Ocaña-Tienda *et al.* (2023) está apresentado na Figura 14, em que é possível observar que há a presença de outras estruturas que não fazem parte do cérebro e possui um aspecto ruidoso no fundo da imagem.

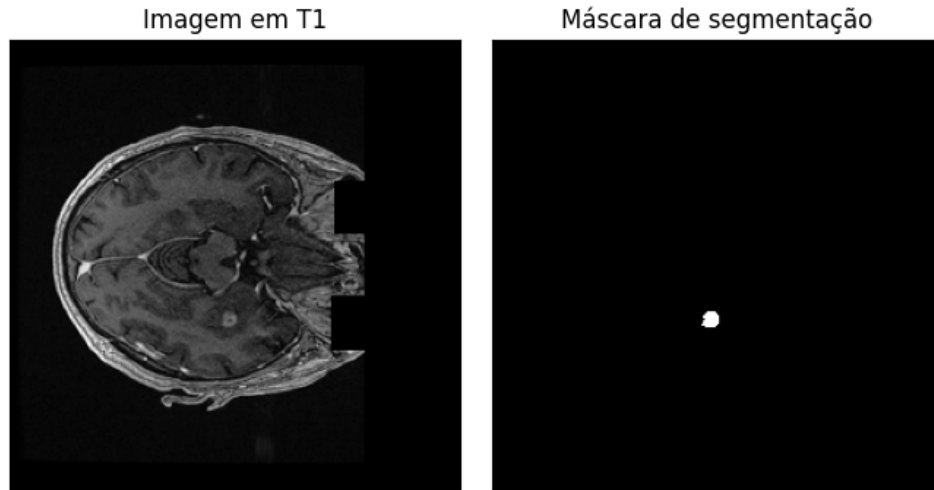


Figura 14 – Exemplo de um *slice* e sua respectiva máscara de segmentação de um paciente do banco de dados de Ocaña-Tienda *et al.* (2023)

Fonte: (OCAÑA-TIENDA *et al.*, 2023)

5.2 Comparação entre os bancos de dados

A diferença entre os três bancos de dados serão descritas a seguir, e na Tabela em que é possível observar a distribuição do número de volumes tumorais por paciente (Figura 15) e o volume tumoral total por paciente (Figura 16). Nota-se que o banco de dados *BrainMetShare* possui um espectro mais amplo de volumes tumorais por paciente, apresentado na Figura 15, porém sem apresentar diferenças para o valor do volume tumoral total quando comparado com os outros bancos, pois se tratam de volumes tumorais de menor volume individualmente, mas que somados possuem volumes tumorais totais comparáveis com os volumes tumorais totais dos outros bancos de dados, como evidenciado na Figura 16.

Já o banco de dados de Ocaña-Tienda *et al.* (2023) não possui uma padronização no tamanho da imagem adquirida, e possui um baixo número de tumores por paciente, com um volume tumoral total comparável com o dos outros bancos, indicando se tratar de metástases maiores. Em relação ao tamanho das imagens, o banco de dados *BrainMetShare* está disponível em duas versões, uma antiga cujo tamanho das imagens não está padronizado e outra mais recente que possui imagens de tamanho $256 \times 256 \times 150$ *voxels*, o banco de dados do TCIA (RAMAKRISHNAN *et al.*, 2024) apresenta imagens de tamanho padronizados em $240 \times 240 \times 155$ *voxels*, enquanto que o banco de dados de Ocaña-Tienda *et al.* (2023), não possui um tamanho padronizado e apresenta imagens até

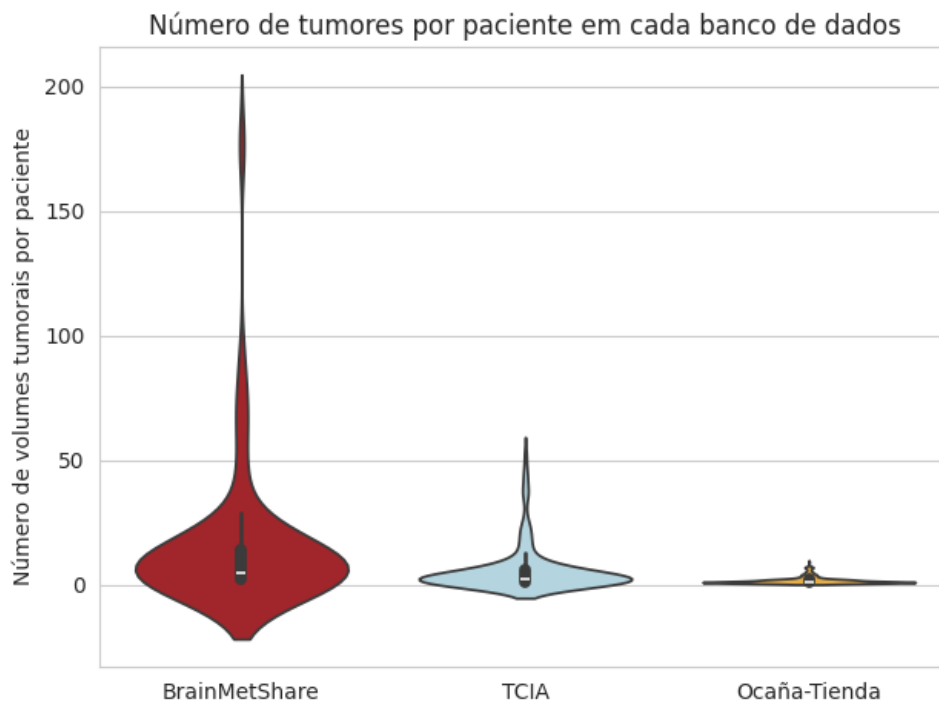


Figura 15 – Plot violino da quantidade de volumes tumorais encontrados por exame, nos diferentes bancos de dados.

Fonte: Elaborado pelo autor (2024)

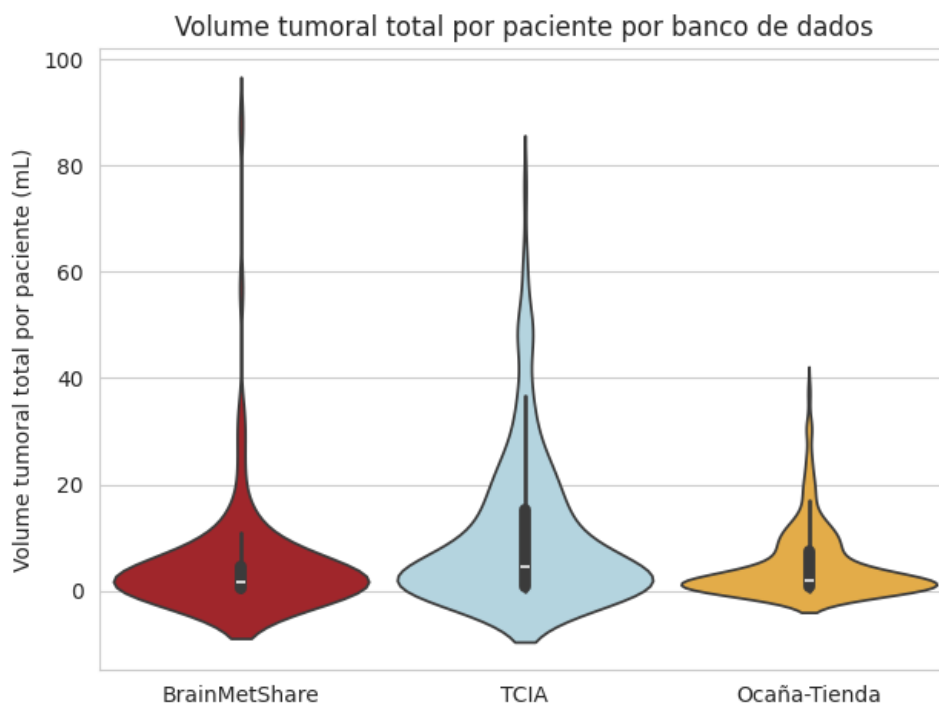


Figura 16 – Plot violino do volume tumoral total por exame, nos diferentes bancos de dados.

Fonte: Elaborado pelo autor (2024)

uma ordem de grandeza maiores do que as imagens de seu próprio banco e dos outros bancos.

Na Tabela 3, encontram-se as informações de aquisição de imagem dos três bancos de dados, sendo possível comparar as diferenças entre os bancos de dados, que se apresentam com características de aquisição distintas.

Tabela 3 – Comparação dos parâmetros de aquisição dos três bancos de dados, em que os parâmetros das 5 primeiras linhas apresentam os valores de mediana acompanhada dos valores mínimo e máximo entre parênteses. TE - tempo de eco, TR - tempo de repetição, N/A - informações não disponíveis. Os números entre parênteses nas intensidades de campo indicam a quantidade de imagens para cada intensidade.

Parâmetros	BrainMetShare	Ocaña-Tienda	TCIA
TE (ms)	N/A (3,24 - 136)	4,76 (1,51 - 16,74)	FLAIR: 92 (10 - 400) T1pos: 3,1 (1,8 - 26,1)
TR (ms)	N/A (8,24 - 6000)	11,00 (5,74 - 600,00)	FLAIR: 9000 (1700 - 12000) T1pos: 1900 (5,9 - 2619,8)
Resolução de pixel (mm)	0,94 (N/A - N/A)	0,5 (0,39 - 1,01)	N/A
Espessura de slice (mm)	1 (N/A - N/A)	1,00 (0,49 - 2,00)	FLAIR: 5,00 (1,0 - 5,5) T1pos: 1,0 (0,9 - 5,0)
Espaçamento de slice (mm)	N/A (1 - 1,6)	1,30 (0,90 - 4,00)	FLAIR: 5,0 (0,0 - 7,5) T1pos: 0,0 (0,0 - 7,0)
Resolução (pixels)	256 x 256 512 x 512	240-640 x 240-640	240 x 240 x 155
Sequências	T1 CUBE pré/pós BRAVO FLAIR	T1 pós	T1 pré/pós T2 FLAIR
Intensidade de Campo (T)	1,5 (18) 3,0 (138)	1,0 (8) 1,5 (550) 3,0 (35)	1,0 (4) 1,5 (113) 3,0 (83)
Fornecedores (quantidade)	GE Siemens	GE (225) Philips (197) Siemens (171)	Siemens (158) General Electric (31) Philips (7) Hitachi (4)

Fonte: Autoria Própria.

5.3 Pré-Processamento

5.3.1 BrainMetShare (BMS)

O antigo banco de dados *BrainMetShare* era composto apenas pelas imagens PNG. Dessa forma, foi necessário realizar estimativas com base em aproximações em torno de uma média de volume do cérebro humano para estimar o volume cerebral da imagem com base nessa média. Para a informação do volume total cerebral, assumimos que todas as imagens dos 105 pacientes possuíam um tamanho padrão, encontrado em literatura como

sendo $1093 \pm 111 \text{ cm}^3$ (AKERET et al., 2021), $1380 \pm 114 \text{ cm}^3$ (FILIFEK et al., 1994), $1290 \pm 148 \text{ cm}^3$ (HAMMERS et al., 2003) e $1180 \pm 112 \text{ cm}^3$ (LANCASTER et al., 2010). A média dos valores e sua incerteza é dada por $1236 \pm 245 \text{ cm}^3$. Esse será o valor de referência utilizado nas estimativas do tamanho das metástases na seção dos resultados para o banco de dados BrainMetShare no formato PNG.

Por outro lado, o banco de dados BMS mais recente, que está em formato NIfTI, possui as informações de volume anotadas no arquivo NIfTI, que serão utilizadas para a definição dos volumes. Apesar de possuírem as imagens cerebrais, tanto o banco antigo quanto o mais recente foram discriminados como dois bancos de dados distintos para fins de comparação da metodologia e para saber se há influência do tipo de arquivo utilizado.

5.3.2 Ocaña-Tienda

O banco de dados disponibilizado por Ocaña-Tienda *et al.* (2023) possui imagens sem a realização da extração cerebral, isto é, contém regiões do crânio, cartilagens e pele dos pacientes. Por isso, foi realizada a operação de extração cerebral neste banco de dados para alcançar uma padronização dos dados, já que todas as imagens dos outros dois bancos de dados, *BrainMetShare* e TCIA, são constituídas de regiões cerebrais extraídas.

Para realizar essa tarefa, foram utilizadas duas estratégias distintas, o *FMRIB Software Library Brain Extraction Tool* (FSL-BET) em suas configurações *Default* e a biblioteca HD-BET em Python (ISENSEE et al., 2019), essa segunda utiliza um modelo de CNN treinado para realizar a extração cerebral e retorna as imagens em formato NIfTI junto das máscaras de segmentação da extração realizada, também em formato NIfTI. Para realizar esse processamento, foi necessário compactar as imagens NIfTI originais do banco de dados para o formato gzip (.gz), uma vez que o HD-BET somente realiza a extração de arquivos NIfTI compactados em formato gzip (.nii.gz), que foi realizado pelo *software* 7-Zip em sua configuração *Default* de compactação. Após utilizar as duas estratégias, foi optado pelo uso da biblioteca HD-BET (ISENSEE et al., 2019), uma vez que os resultados da extração cerebral foram qualitativamente mais coerentes com o esperado, conforme apresentado na Figura 17, que contém um exemplo dos resultados obtidos com as duas estratégias, bem como a imagem original. Nesse exemplo, o paciente foi escolhido arbitrariamente, e o padrão de extração mais agressivo pelo FSL-BET também foi verificado nas imagens de outros pacientes.

5.3.3 The Cancer Imaging Archive (TCIA)

Como as máscaras da segmentação fornecidas por Ramakrishnan *et al.* (2024) possuem diferentes classes de segmentação, foi necessário remover a região com edema peritumoral, identificado pela *label* 2 no arquivo NIfTI. Dessa forma, apenas as regiões

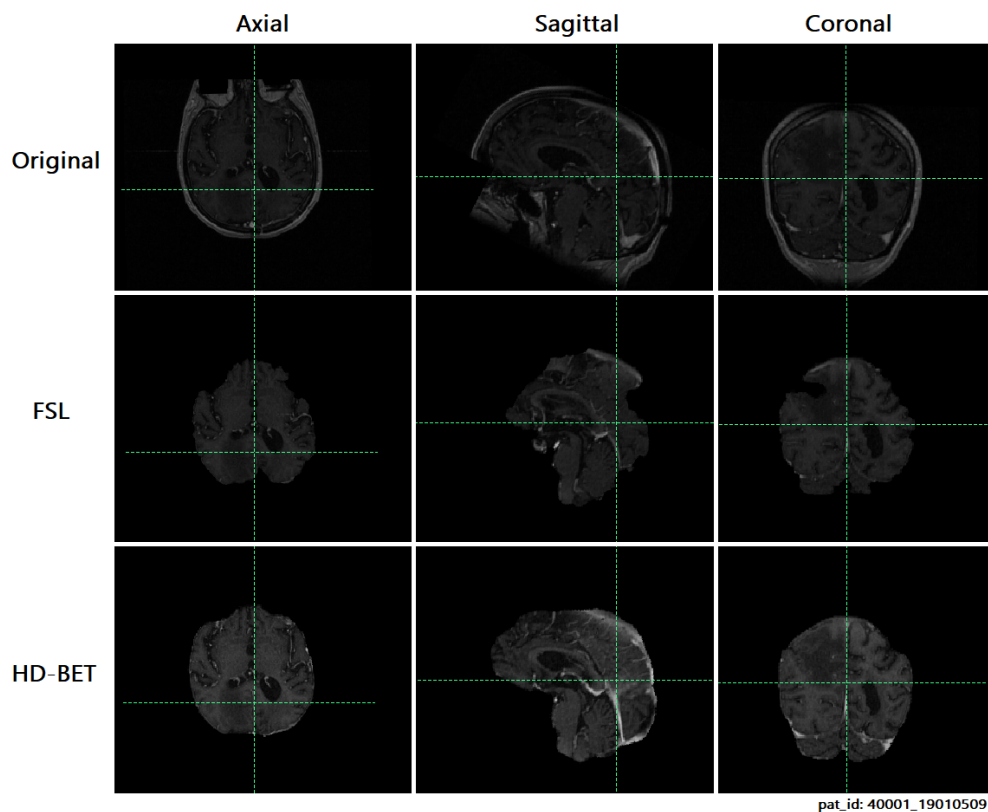


Figura 17 – Planos axial, sagital e coronal originais, com o uso do *software* FSL-BET e com o uso da biblioteca HD-BET para a extração cerebral na etapa de pré-processamento do banco de dados de Ocaña-Tienda *et al.* (2023).

Fonte: Elaborado pelo autor (2024)

tumorais com absorção de contraste e com necrose foram consideradas e unidas em uma mesma classe de segmentação, conforme indicado na Figura 18. Isso foi realizado, porque os outros bancos de dados não fornecem as máscaras de segmentação da região com edema peritumoral, e apresentam as regiões tumorais com necrose e com absorção de contraste como uma única região. Sendo assim, esse procedimento foi realizado para que esse banco de dados pudesse ser compatível com as análises realizadas nos outros bancos de dados que possuem a máscara de segmentação em classes binárias para regiões com presença de tumor e com ausência de tumor.

5.4 Modelos de Segmentação

Nesta seção, serão apresentados os três modelos da literatura utilizados neste estudo: Met-Seg (OTTESEN *et al.*, 2023), AURORA (BUCHNER *et al.*, 2023) e Raidionics (BOUGET *et al.*, 2023).

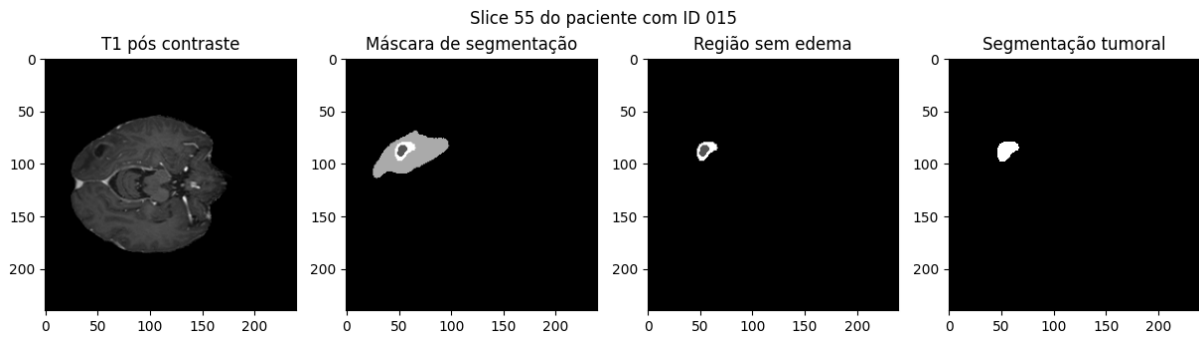


Figura 18 – Exemplo do pré-processamento em que ocorre a remoção da região de edema peritumoral da máscara de segmentação das imagens do banco do TCIA.

Fonte: Elaborado pelo autor (2024)

5.4.1 Met-Seg

O modelo desenvolvido por Ottesen *et al.* (2023) denominado Met-Seg, é resultado de uma adaptação e avaliação de modelos de CNNs com arquiteturas de rede HRNetV2 2.5D e 3D com *decoder* em formato de U-Net na tarefa de segmentação de metástases cerebrais e que envolveu o uso de dois bancos de dados do Hospital da Universidade de Oslo (OUH, do inglês *Oslo University Hospital*) e o *BrainMetShare* (OTTESEN *et al.*, 2023).

Esse modelo foi treinado apenas com imagens de 110 pacientes (100 para treino e 10 para validação) do banco de dados *BrainMetShare* e foi testado no banco de dados OUH, externo ao treino, e os resultados para as duas arquiteturas foram comparadas entre si e com uma outra arquitetura de referência, a nnU-Net (ISENSEE *et al.*, 2021). Nessa comparação, os autores encontraram uma melhor performance para o modelo 3D, que foi disponibilizado no GitHub¹ junto com os parâmetros do modelo treinado e o código para aplicação em linguagem de programação *Python*.

O formato da imagem compatível com esse modelo é a NIfTI (.nii ou .nii.gz) e não possui restrição de tamanho de imagem. Além disso, o modelo possui a opção de inferência com *Dropout* em nível de *Input*, isto é, é possível remover uma ou mais sequências de RM para realizar a segmentação final. No entanto, são limitadas a combinações envolvendo as sequências ponderadas em T1 pré e pós aplicação de contraste, BRAVO e a sequência ponderada em T2 FLAIR, que estão no padrão de aquisição do banco de dados *BrainMetShare*.

5.4.2 AURORA

O modelo AURORA foi desenvolvido por Buchner *et al.* com o intuito de realizar a segmentação de GTVs para o planejamento radioterápico de metástases cerebrais,

¹ <https://github.com/JonOttesen/Met-Seg>

visando acelerar os processos envolvidos no delineamento e minimizar sua variação nos planejamentos radioterápicos (BUCHNER et al., 2023).

Para o treino desse modelo, que tem como base a arquitetura U-Net 3D, foram utilizadas imagens de 348 pacientes (260 para treino e 88 para validação) em ponderação T1 pré e pós aplicação de contraste, e também na ponderação T2 e T2 FLAIR. O modelo treinado foi disponibilizado no GitHub², junto com o código que permite sua aplicação em linguagem de programação *Python*. Para a inferência, os parâmetros definidos pelo usuário do *inferer* foram utilizados da seguinte maneira: o *test time augmentations* igual a *True* com *sliding window batch size* igual a 8 e uso de *cuda devices* igual a "0".

Assim como o Met-Seg, o modelo AURORA também realiza segmentações de algumas combinações das sequências, porém é mais limitado na quantidade de combinações quando comparado ao Met-Seg. O formato da imagem compatível com esse modelo também é a NIfTI (.nii ou .nii.gz) e não possui restrição de tamanho de imagem.

5.4.3 Raidionics

O modelo Raidionics, desenvolvido por Bouget *et al.* (2023) possui uma interface gráfica do utilizador que realiza segmentações automáticas de regiões tumorais pré e pós operatórias nos principais tipos de tumores cerebrais, como glioblastomas, gliomas de baixo grau, meningiomas e metástases (BOUGET et al., 2023). Essa interface é amigável aos profissionais que não possuem experiência em linguagens de programação e facilita a aplicação na prática clínica. O *software* e os pesos de alguns modelos treinados foram disponibilizados no GitHub³, porém não se encontra disponível o algoritmo de treino utilizado pelos autores.

Nessa interface, o usuário seleciona o tipo da sequência da imagem e seleciona qual a sequência de RM utilizada na aquisição, se é uma segmentação pré ou pós operatória e qual o tipo de tumor para segmentação. O formato da imagem compatível com esse modelo é a NIfTI (.nii ou .nii.gz) e DICOM, também sem restrição de tamanho de imagem. Após a inferência, o *software* gera um arquivo NIfTI com o resultado da segmentação.

Para esse modelo em específico, só foi possível realizar as segmentações de metástases nas imagens ponderadas em T1 com contraste, pois ao realizar o processo de predição para as outras modalidades disponíveis pelo *software* (T1, T2 e FLAIR) não eram gerados *outputs* das segmentações. Como o banco de dados de Ocaña-Tienda *et al.* (2023) é composto apenas de imagens T1 com contraste, foi conveniente utilizar apenas as imagens T1 com contraste dos outros bancos para a inferência.

Um outro aspecto desse modelo é que não foi possível realizar as predições para o

² <https://github.com/BrainLesion/AURORA/>

³ <https://github.com/raidionics/Raidionics>

banco de dados *BrainMetShare* em seu formato nativo PNG, pois a conversão do diretório com as imagens em formato PNG para um arquivo em formato NIfTI gerou um erro no *software*, que indicava que o arquivo não fora carregado corretamente. Por esse motivo a aplicação desse modelo em específico não foi realizada nesse banco de dados. Além disso, para um dos exames do banco de dados de Ocaña-Tienda *et al.* (2023), o mesmo erro se repetiu e, para fins de comparação dos resultados com os outros modelos, o dado faltante foi preenchido com a média dos dados realizado para todos os exames desse banco de dados, já que deixar o dado em branco resultaria em erros de leitura dos *dataframes* que exigiam um mesmo tamanho para o vetor de dados. Essa abordagem, no entanto, não implica em diferenças significativas para a análise geral, visto que a quantidade de exames desse banco de dados é a maior entre todos os bancos.

5.4.4 Proposição de um modelo de segmentação

Tendo em vista que o objetivo deste trabalho é realizar o delineamento automático de regiões tumorais, as imagens e respectivas máscaras de segmentação obtidas dos pacientes submetidos a exames de RM foram utilizadas para o treinamento de CNNs. Para o treinamento, foi utilizado o *BrainMetShare* de Stanford, separado por tipo de Sequência de Pulso (SP), constituído por 18.189 imagens de cada tipo de ponderação, das quais 80% (17.546) foram utilizadas para treino e 20% (643) para validação. Foi realizada a variação dos hiperparâmetros de cada CNN, totalizando seis modelos treinados para cada SP. Assim, foi possível realizar a avaliação de qual SP apresenta uma melhor eficiência na detecção de metástases cerebrais.

A partir desse resultado, as melhores SP foram colocadas nos canais RGB da imagem de *input* da CNN, resultando em uma única imagem com informações sobre as três diferentes SP de melhor relevância para a classificação de metástases cerebrais.

Dessa maneira, com a troca dos hiperparâmetros e das arquiteturas de rede, obteve-se métricas que sejam próximas ou melhores das encontradas em literatura, de modo a classificar a melhor maneira de realizar a segmentação de metástases cerebrais através dos coeficientes Dice e Distância de Hausdorff. Um fluxograma com as etapas de aquisição, treino e análise de resultados está descrita no fluxograma da Figura 19.

A CNN utilizada foi uma U-Net 2D para imagens de tamanho de 256×256 *pixels* em três canais RGB, em que o *input* da rede contou com a atribuição do valor médio entre os canais de cada *pixel* para cada canal do *input*. Dessa forma, foi possível colocar imagens de três sequências de pulso em uma mesma imagem, pois cada canal da imagem do *input* representava uma modalidade de imagem. A rede U-Net possui um formato *encoder-decoder* semelhante à letra U, com a etapa de *encoder* composta por três blocos de duas convoluções seguidas por uma camada de *downsampling*, e na etapa de *decoder* possui quatro blocos de duas convoluções seguidas por uma camada de *upsampling*. Cada

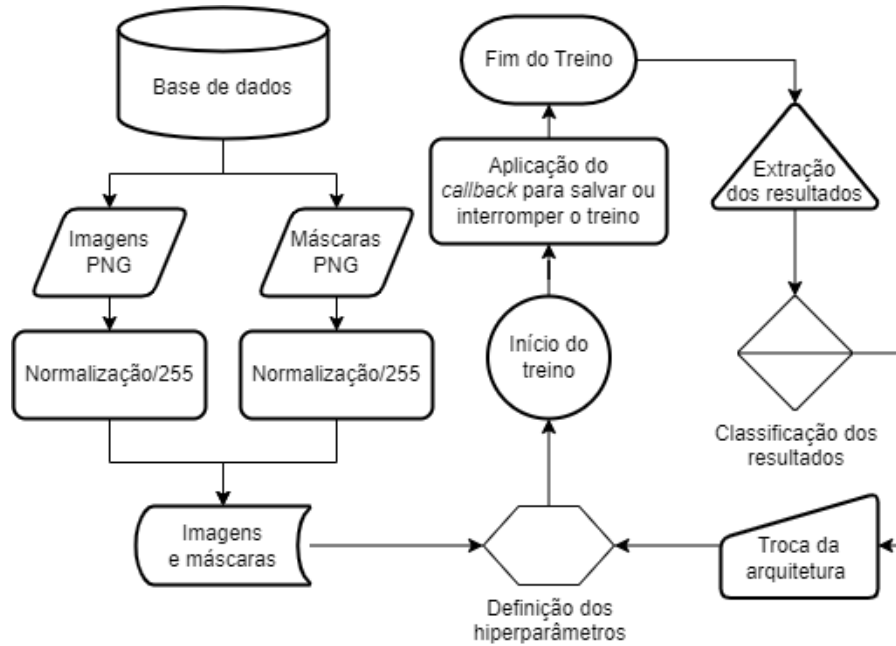


Figura 19 – Fluxograma dos passos metodológicos.

Fonte: Autoria própria.

convolução é iniciada por uma camada de ativação com função de ativação *Rectified linear units* e finalizada com uma camada de *Batch Normalization* e, após as duas convoluções, são aplicadas uma camada de *Max Pooling* seguida de uma camada residual que será alimentada nos blocos do *decoder*, conforme ilustra a Figura 20.

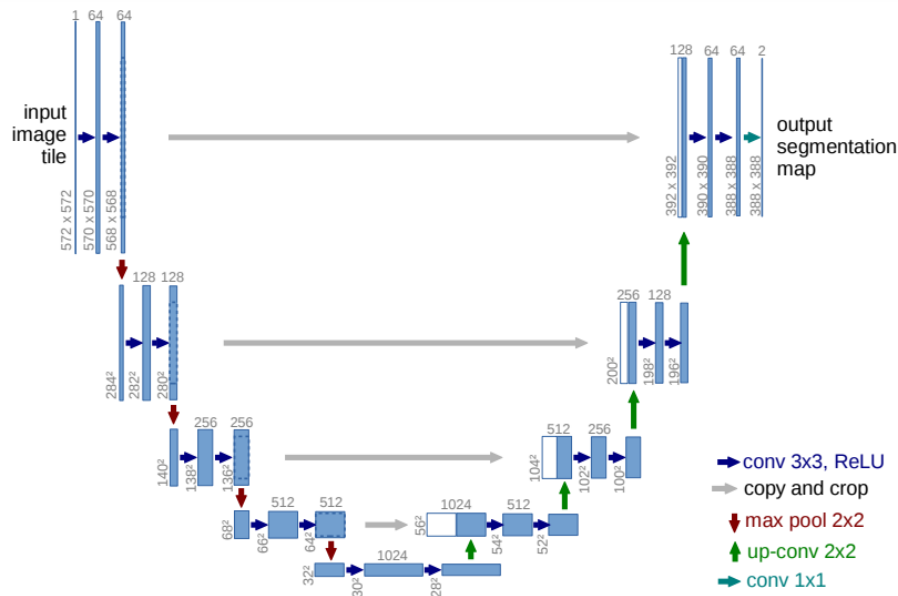


Figura 20 – Diagrama da rede U-Net

Fonte: (RONNEBERGER et al., 2015)

De acordo com (KARIMI et al., 2021), o treino do modelo *Fully Convolutional Network* no formato *encoder-decoder*, que é o caso da U-Net, possui resultados similares quando os pesos do *encoder* são congelados em valores iniciais aleatórios ou quando são

pré-treinados, e apenas os pesos do *decoder* são necessários passar pelo treino na tarefa de segmentação para imagens médicas.

Dessa forma, utilizando um modelo *dummy* formado por pacientes do banco de dados *BrainMetShare* sem a aplicação de técnicas de *Data Augmentation* ou *Transfer Learning* avaliamos a influência da função perda nos resultados do modelo. Esse modelo *dummy* possui um *batch size* de 32 amostras, foi treinado por número máximo de épocas igual a 200, a qual nunca foi alcançada devido ao uso da *callback earlystopping*, que pausa o treino quando não há melhorias na métrica de controle por 10 épocas consecutivas, outra *callback* utilizada foi a *reduceLRonplateau* de paciência de 3 épocas e fator de redução de 0,2, que diminui em 20% o valor da *learning rate* a cada 3 épocas caso não haja melhorias na métrica de controle. O formato da rede utilizada foi uma U-Net (RONNEBERGER et al., 2015), com 2 blocos de convolução com ativação ReLU e *Batch Normalization*, seguidos por uma camada de *MaxPooling/UpSampling*.

Após isso, o modelo foi aprimorado para o uso de *Data Augmentation*, e o uso de *transfer learning* não foi realizado devido à diferença na arquitetura de rede. As transformações utilizadas contaram com as características de rotação aleatória de no máximo 90 graus para a direita ou para a esquerda, deslocamento de até 10% para as quatro direções no eixo x-y, e inversão horizontal, todas as transformações possuíam 50% de probabilidade de ocorrer, um exemplo dessas transformações está apresentado na Figura 21. Essas transformações representam as diferentes formas de variações que ocorrem na prática clínica, e a transformação de inversão horizontal leva em consideração a simetria bilateral do cérebro.

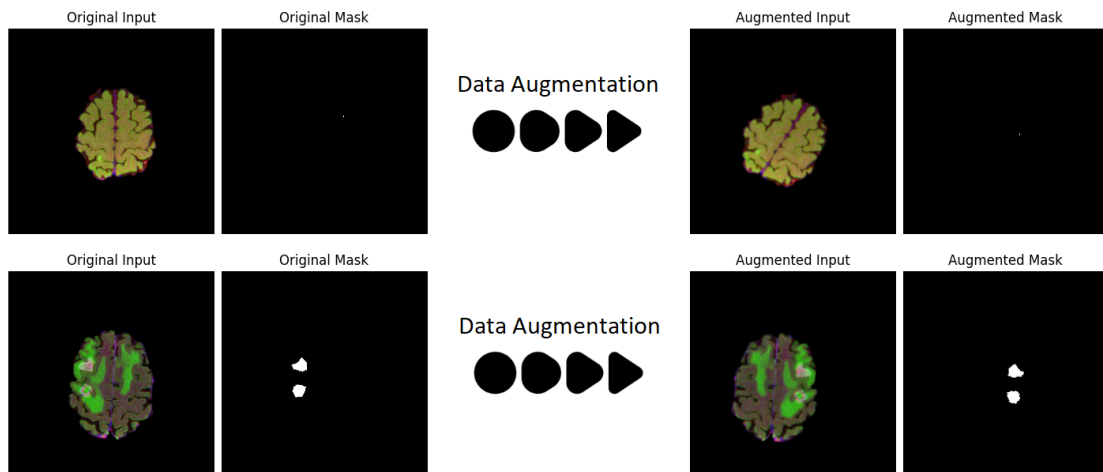


Figura 21 – Exemplo das transformações de Data Augmentation aplicadas em duas imagens de treino da rede neural.

Fonte: Autoria própria com imagens do banco BMS (GRØVIK et al., 2020)

5.5 Análise e Validação dos Modelos

A validação do modelo é um passo importante para garantir a generabilidade do modelo. Os modelos treinados serão validados utilizando-se tanto subconjuntos de teste do banco de imagens do treinamento quanto de outros bancos de imagens, para que seja realizada a validação externa. Dessa maneira, foram comparadas as métricas de avaliação de performance para cada um dos modelos nos diferentes conjuntos de validação.

Para o banco de dados utilizado no treino, *BrainMetShare* (GRØVIK et al., 2020), foi utilizado o método *k-folds* de validação cruzada, que consiste na separação em um número k de partes, ou *folds*, do conjunto de dados, reservando uma dessas partes sem a qual o treino é realizado. Assim, é possível utilizar a parte reservada para a validação, e o treino é realizado um número k de vezes, reservando uma vez cada um dos *folds* e comparando as métricas obtidas ao final de cada treino (CHOLLET, 2017), como está apresentado na Figura 22. Apesar de ter um maior custo computacional devido à maior quantidade de treinos, essa técnica é amplamente utilizada em literatura (LIU et al., 2017; DIKICI et al., 2020; XUE et al., 2020; PENNIG et al., 2021; PARK et al., 2021b).

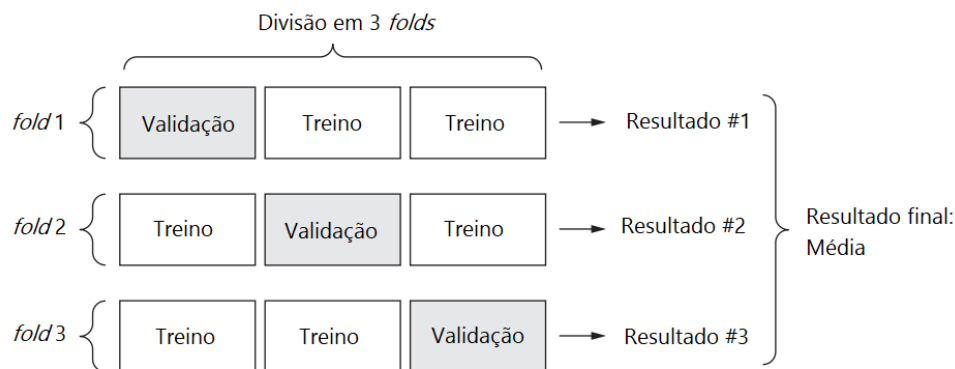


Figura 22 – Validação cruzada em *k-folds* para $k = 3$.

Fonte: Adaptado de (CHOLLET, 2017)

Para a validação da generabilidade e robustez do modelo, serão realizados treinos com os conjuntos de dados distintos (GRØVIK et al., 2020; OCAÑA-TIENDA et al., 2023) de forma independente, a fim de avaliar a aplicação de um modelo treinado com os dados de um *dataset* em outro e vice-versa.

Outra forma de avaliar a abrangência do modelo é através da subamostragem em grupos com características diferentes, como por exemplo, a instituição e aparelhos em que foram obtidas as imagens ou a quantidade, tamanho e volume das metástases cerebrais analisadas. Essa análise deve ser realizada para avaliar os efeitos de localidade do *dataset*, que pode trazer resultados bons para um conjunto de dados do mesmo local mas ruins para um conjunto de dados de outro local. Também é possível avaliar os efeitos do tamanho da lesão cerebral, pois as metástases cerebrais de menor volume são mais difíceis de serem encontradas, e o DSC para esse subgrupo é menor do que para subgrupos de metástases

maiores (BOUSABARAH et al., 2020). Assim, será possível ter um modelo treinado para cada subgrupo de metástases, que poderão passar por uma rede de identificação antes de serem colocadas na rede de segmentação, cuja performance será medida através dos coeficientes Dice, FPP e IoU para comparação com a literatura.

6 Resultados e Discussão

A apresentação dos resultados se iniciará com o a apresentação das contagens de volumes tumorais para cada banco de dados através do algoritmo de contagem de metástases, seguida pelas análises de performance dos três modelos e também a análise dos *checklists* CLAMP e CLAIM para as publicações envolvendo esses modelos. Após essa análise, será apresentado o desenvolvimento do modelo de segmentação proposto.

6.1 Algoritmo de contagem de metástases

A fim de verificar a quantidade de falsos positivos e de falsos negativos para a detecção de volumes tumorais, foi realizada uma contagem da quantidade de volumes tumorais por paciente, isto é, a quantidade de *voxels* presentes tanto nas máscaras de segmentação desenhadas por profissionais da área (GT, do inglês *Ground Truth*) quanto nas máscaras das predições dos modelos. A informação do número total de volumes tumorais está presente nos artigos de Ottesen *et al.* (2023), Ocaña-Tienda *et al.* (2023) e Ramakrishnan *et al.* (2024). No entanto, essa quantidade descrita em cada estudo é divergente da quantidade de volumes tumorais que estão disponibilizadas nos bancos de dados, o que dificulta a replicação dos resultados de cada estudo.

Como as informações sobre o número de metástases cerebrais eram limitadas, sendo o banco de Ocaña-Tienda *et al.* o único que apresentou a quantidade de volumes tumorais por exame, foi necessário implementar um algoritmo para a contagem desses volumes tumorais para os outros bancos de dados. Para isso, foi utilizado o código em linguagem Python para a contagem de *voxels* desconexos, utilizando a função *label* da biblioteca *scipy.ndimage* e o código está disponibilizado no Apêndice A. Esse algoritmo necessita de um valor de tamanho mínimo de um *cluster* para contabilizá-lo, e esse valor de tamanho de *cluster* mínimo foi considerado como aquele formado por um conjunto de 6 *voxels* por *default*.

Esse tamanho mínimo foi definido uma vez que o banco de dados de Ocaña-Tienda *et al.* (2023) possuía três etiquetas na máscara de segmentação para cada volume tumoral segmentado, correspondendo às regiões de fundo (*label 0*), região com contraste (*label N1*) e região de necrose tumoral (*label N2*), em que a letra N representa o índice da metástase. Essas duas etiquetas correspondentes à região tumoral foram unificadas, e em seguida foram contadas a quantidade de etiquetas N presentes na imagem 3D. Dessa maneira, foi possível estabelecer a quantidade exata de volumes tumorais desse banco de dados, e o algoritmo para a contagem de *voxels* desconexos foi calibrado para desconsiderar a contagem de ruído, que obteve o melhor resultado ao ignorar *clusters* de *voxels* consti-

Tabela 4 – Contagem da quantidade (Qtd.) de *clusters* para o banco de dados de Ocaña-Tienda *et al.* (2023) bem como a quantidade de acertos do número de *clusters* por paciente (TP); Os valores 591 e 373 representam os valores *Ground Truth* (GT), e vx indica o tamanho mínimo do *voxel* para ser incluído na contagem

GT	$\geq 1vx$	$\geq 2vx$	$\geq 3vx$	$\geq 4vx$	$\geq 5vx$	$\geq 6vx$	$\geq 7vx$	$\geq 8vx$
Qtd. / 591	662	607	598	595	592	590	587	584
TP / 373	328	358	366	368	369	370	368	367

Fonte: Autoria Própria.

túidos por menos do que 6 *voxels*. Essa calibração foi realizada para que fosse definido a quantidade mínima de *voxels* para contagem de *clusters* pelo algoritmo e, na Tabela 4, encontram-se os valores da quantidade total de volumes tumorais identificados pelo algoritmo para cada uma das considerações de tamanho mínimo de *voxels* por *cluster*, junto com a quantidade de acertos em cada uma delas. Essa metodologia de contagem também foi empregada no banco de dados do TCIA, que possui 975 volumes tumorais segmentados (RAMAKRISHNAN *et al.*, 2024), e resultou em um total de 993 *clusters* de tamanhos maiores ou igual a 5 *voxels*, 981 *clusters* de tamanhos maiores ou iguais a 6 *voxels* e 966 *clusters* de tamanhos maiores ou iguais a 7 *voxels*, o que concordou com a contagem para o banco de Ocaña-Tienda *et al.* (2023).

Além disso, o gráfico na Figura 23 indica a soma acumulada da quantidade de volumes tumorais encontrados utilizando o algoritmo para diferentes tamanhos mínimos de *clusters*, em comparação com a anotação realizada por Ocaña-Tienda *et al.* (2023).

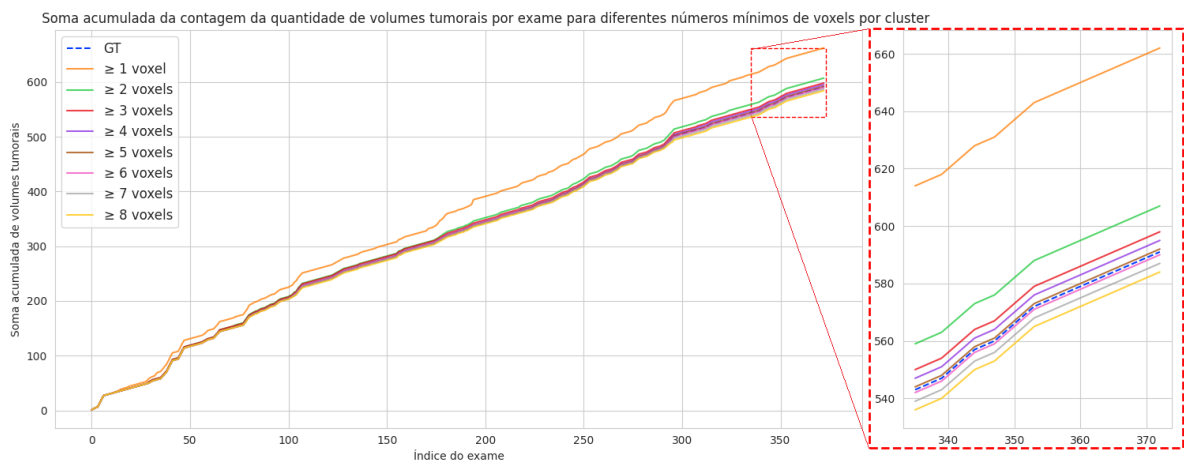


Figura 23 – Soma acumulada do número de volumes tumorais por exame, identificados pelo algoritmo de contagem de *clusters* para o banco de dados de Ocaña-Tienda *et al.* (2023).

Fonte: Autoria própria.

Dessa forma, foram considerados todos os *clusters* formados por mais do que 6 *voxels*, incluindo conjuntos formados por 6 *voxels*. A escolha para esse tamanho mínimo foi feita por ser uma medida conservadora, já que é mais seguro a presença de falsos positivos

do que falsos negativos para o número de *clusters* identificados por paciente. Somado à isso, vale salientar que a contagem do número de *clusters* realizada para o banco de dados do TCIA apresentou uma diferença significativa para os tamanhos mínimos de 5 e 6 *voxels*, com erros percentuais de 0,6% e 1,8% em relação à quantidade de volumes tumorais descrita por Ramakrishnan *et al.* (2024) e, por esse motivo, escolheu-se a contagem com o número mínimo de *voxels* sendo igual a 6.

Os resultados da contagem para as máscaras GT para todos os bancos de dados estão apresentados na Tabela 5. Vale a observação de que não houve alteração dos resultados obtidos variando-se os estruturantes volumétricos da função *label* que desconsideram conexões diagonais (em cruz) e que consideram conexões diagonais (em cubos). Nessa tabela, o número mínimo de 6 *voxels* foi considerado para contagem dos *clusters* ou volumes, e a quantidade de volumes segmentados de tamanhos menores do que 6 *voxels* estão representados entre parênteses. Note que esses volumes foram considerados como ruído e não foram adicionados à contagem de *clusters*. A partir dos dados apresentados nessa tabela, é possível interpretar que em relação ao número de volumes tumorais, o modelo Met-Seg foi o que mais se aproximou do esperado para o banco de dados do *BrainMetShare*, o que é de se esperar para o modelo treinado com esse banco de dados. Já para os outros bancos, o modelo Met-Seg apresentou uma contagem superestimada dos *clusteres*, o que significaria na prática da radioterapia um aumento desnecessário da dose de radiação total, e essa dose extra seria aplicada em tecidos saudáveis. Já para os modelos AURORA e Raidionics, ambos apresentaram uma contagem sobrestimada para o banco de dados *BrainMetShare*, o que pode ser explicado pela grande quantidade de metástases de menor volume individual e a dificuldade desses modelos em realizar a segmentação desses volumes tumorais pequenos, resultado esse que também pode ser observado quando esses modelos são aplicados no banco de dados do TCIA. Já para o banco de dados de Ocaña-Tienda *et al.* (2023), o modelo AURORA teve uma performance parecida com a do modelo Met-Seg, superestimando o valor da contagem, e com uma quantidade excessiva de ruído de 1560 *clusters*, o que pode ser muito perigoso para a prática da radioterapia. Por sua vez, o modelo Raidionics teve a melhor performance nesse mesmo banco de dados, com um resultado para a contagem cerca de três vezes menor do que para os outros dois modelos, apesar de realizar uma contagem de cerca de 43% maior em relação ao esperado.

A partir do uso desse algoritmo, foi possível extrair algumas das informações complementares às imagens que também podem ser importantes para a tarefa de segmentação, conforme foram apresentadas por Niranjana *et al.* (2019) e Cho *et al.* (2020), sendo elas a contagem do número de nódulos metastáticos, tamanho médio dos nódulos metastáticos e volume estimado do tumor cerebral para os três bancos de dados.

Além dessa contagem, foi possível realizar a contagem dos valores de tamanho de metástase para cada tumor isolado, bem como o volume tumoral total em cada um dos

Tabela 5 – Número de volumes considerados para cada um dos arquivos de segmentação, os números entre parênteses são as quantidades de volumes formados por menos do que 6 *voxels*, que foram considerados como ruído e não foram incluídos na contagem. *Ground Truth* é a anotação feita por radiologistas, Met-Seg (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) são os modelos de redes neurais. *Não foi possível carregar um paciente no modelo Raidionics, com um volume tumoral, do banco de Ocaña-Tienda et al. (2023).

	BMS	Ocaña-Tienda	TCIA
Ground Truth	1354 (354)	590 (72)	981 (190)
Predição Met-Seg	1067 (118)	2234 (148)	1928 (292)
Predição AURORA	290 (56)	2381 (1560)	510 (150)
Predição Raidionics	312 (40)	841 (87)*	469 (81)

Fonte: Autoria Própria.

Tabela 6 – Valores médios para a contagem de volumes tumorais, o volume tumoral total (V_{total}) e a mediana do volume tumoral individual ($V_{mediano}$) para os bancos de dados BMS, TCIA e Ocaña-Tienda et al. (2023).

Banco	Contagem	V_{total} (mL)	$V_{mediano}$ (mL)
BMS PNG	14 ± 27	$5,1 \pm 11,3$	$1,05 \pm 5,63$
BMS NIfTI	14 ± 26	$3,5 \pm 6,8$	$0,63 \pm 2,56$
TCIA	5 ± 7	$10,3 \pm 13,8$	$4,36 \pm 9,71$
Ocaña-Tienda et al. (2023)	2 ± 1	$4,9 \pm 6,6$	$3,40 \pm 4,65$

Fonte: Autoria Própria.

exames dos bancos de dados. Sendo assim, na Tabela 6 encontram-se os valores médios com seus desvios padrões para a quantidade de tumores por exame, o volume tumoral total por exame e a mediana do tamanho individual de cada volume tumoral por exame, onde é possível observar diferenças significativas entre os bancos de dados, como a presença de maiores quantidades de volumes tumorais por exame para o banco de dados BMS, um maior volume tumoral total para o banco de dados do TCIA, e um volume mediano menor do que 1 mL para o banco de dados BMS, em seu formato NIfTI. Dessa forma, podemos resumir o banco de dados BMS como sendo um banco formado por poucos exames, que contém um grande número de pequenas lesões, enquanto que o banco de dados de Ocaña-Tienda et al. (2023) possui uma grande quantidade de exames, sendo que cada exame contém uma ou duas lesões de tamanhos medianos. Ambos esses bancos de dados possuem um volume tumoral total de cerca de 5 mLs. Por fim, o banco de dados do TCIA possui uma quantidade intermediária de exames, que contém uma quantidade intermediária de volumes individuais, que são de tamanhos parecidos com os tamanhos dos volumes de Ocaña-Tienda et al. (2023), porém com um volume tumoral total equivalente ao dobro (10 mLs) dos volumes totais dos outros dois bancos de dados.

6.2 Análise dos modelos da literatura

6.2.1 Modelo MetSeg, desenvolvido por Ottesen *et al.* (2023)

O modelo previamente treinado que foi disponibilizado por Ottesen *et al.* (2023) e chamado de Met-Seg foi testado com os arquivos do banco de dados BMS, TCIA e Ocaña-Tienda *et al.* (2023). O banco de dados BMS também foi utilizado pelos autores para o treinamento do modelo. Nesse estudo, foi feita a remoção de 51 dos 156 pacientes do banco de dados, pois as segmentações desses 51 pacientes não estavam disponíveis publicamente, sendo esse o principal fator que impede a replicação exata da metodologia de Ottesen *et al.* (2023). Na Tabela 7 estão apresentados os valores do coeficiente de similaridade Dice apresentado pelos autores (OTTESEN *et al.*, 2023), e também quando aplicado aos 105 pacientes do banco de dados. Em seu trabalho, Ottesen *et al.* (2023) calculam os valores de DSC médio por *slice* da imagem 3D (swDSC, do inglês *slice-wise Dice Coefficient Score*), e nos casos em que a máscara de segmentação do *slice* é composta unicamente de *pixels* pretos, isto é, não possui metástase no *slice*, o valor para o swDSC é calculado como igual a 1 caso a predição do modelo esteja totalmente correta, isto é, a predição é um *slice* formado exclusivamente de *pixels* pretos, que são os verdadeiros negativos.

Tabela 7 – Métricas de performance do coeficiente Dice médio por slice (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), sensibilidade geral (Sens.), sensibilidade por paciente (Sens./pat.) e falsos positivos por paciente (FP/pat.) para as predições com todas as modalidades presentes do banco de dados *BrainMetShare*.

Estudo	swDSC _{BG}	swDSC _{noBG}	Sens.	Sens./pat.	FP/pat.
Ottesen <i>et al.</i> (2023)	0,84 ± 0,13	-	0,71	0,84 ± 0,18	3,2 ± 6,5
Calculado	0,81 ± 0,13	0,40 ± 0,16	0,61	0,74 ± 0,31	3,3 ± 3,4

Fonte: Autoria própria.

Na Tabela 7, também se encontra um valor de coeficiente Dice desconsiderando-se os verdadeiros negativos (noBG, do inglês, *no background*), esse coeficiente Dice corrigido representa uma métrica mais representativa da eficiência do modelo. Isso porque o valor do coeficiente Dice dado pela Equação 14 não leva em consideração os verdadeiros negativos em seu cálculo e adicionar os acertos de predições de *background* tendem a aumentar o valor de swDSC, mesmo em situações nas quais o modelo não é eficiente. Por exemplo, para o banco dados *BrainMetShare* o número de *slices* para todos os pacientes é igual a 18189, sendo que desses 4632 possuem metástases e 13557 não possuem metástases. Dessa forma, em uma situação hipotética em que um algoritmo nunca treinado, cuja única função seja realizar a predição de *slices* pretos, acertaria 13557 das 18189 predições, e seu swDSC teria o valor de $\frac{13557}{18189} = 0,74534059046$, o que é um valor com cerca de 10% de variação dos resultados apresentados pelos autores (OTTESEN *et al.*, 2023). Por essa razão, foi

optado o uso do cálculo do swDSC desconsiderando-se as predições sem metástases e que são compostas unicamente de *pixels* verdadeiros negativos.

Em seguida, foram realizadas predições para o modelo com dados faltantes, isto é, com a remoção de algumas das sequências de pulso no *input* do modelo. Dessa maneira, foram removidas uma de cada vez as sequências FLAIR, BRAVO, T1 pré contraste, T1 pós contraste e ambas as sequências T1. Essa análise foi realizada com o intuito de saber se existem sequências com maior importância para a predição do modelo e, se existem, quais são essas sequências. Os resultados obtidos com a remoção dessas sequências estão apresentados na Tabela 8.

Tabela 8 – Métricas de performance do coeficiente Dice médio por *slice* (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com algumas das modalidades faltantes, denotadas por mFLAIR, mBRAVO, mT1pos, mT1pre e mT1 para, respectivamente, os casos com a remoção das sequências FLAIR, BRAVO, T1 pós contraste, T1 pré contraste e ambas as sequências T1 do modelo de (OTTESEN et al., 2023) para o banco de dados *BrainMetShare*.

swDSC _{mFLAIR}		swDSC _{mBRAVO}		swDSC _{mT1pos}		swDSC _{mT1pre}		swDSC _{mT1}	
BG	noBG	BG	noBG	BG	noBG	BG	noBG	BG	noBG
0,81	0,39	0,62	0,10	0,81	0,40	0,80	0,37	0,81	0,40

Fonte: Autoria própria.

Nota-se que entre todas as sequências presentes desse banco de dados, a remoção da sequência BRAVO possui um menor resultado para o coeficiente Dice médio por *slice* tanto considerando-se os verdadeiros negativos (swDSC = 0,62 ± 0,18) quanto os desconsiderando (swDSC = 0,10 ± 0,09). Para essa avaliação, os menores valores encontrados para o valor de swDSC informa a relevância da sequência para realizar a predição, e esse resultado indica que a sequência BRAVO possui informações que, caso sejam removidas, resultam em uma menor eficiência do modelo, destacando sua importância para realizar boas predições. Em seguida, foram realizadas predições com apenas uma das sequências, removendo-se as outras três, e os resultados para os valores do coeficiente Dice para essas predições podem ser encontrados na Tabela 9.

Tabela 9 – Métricas de performance do coeficiente Dice médio por *slice* (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com apenas uma das sequências FLAIR, BRAVO, T1 pós contraste (T1pos), T1 pré contraste (T1pre) e ambas as sequências T1 do modelo de (OTTESEN et al., 2023) para o banco de dados *BrainMetShare*.

swDSC _{FLAIR}		swDSC _{BRAVO}		swDSC _{T1pos}		swDSC _{T1pre}		swDSC _{T1}	
BG	noBG	BG	noBG	BG	noBG	BG	noBG	BG	noBG
0,57	0,09	0,81	0,38	0,63	0,18	0,66	0,02	0,66	0,02

Fonte: Autoria própria.

Como pode-se observar, o uso único da sequência BRAVO ($swDSC = 0,38 \pm 0,17$) possui resultados similares ao uso de todas as sequências ($swDSC = 0,40 \pm 0,16$), o que reforça a importância da sequência BRAVO em relação às outras sequências. Um gráfico comparativo com os valores de $swDSC$ calculados em todas as combinações de sequências está apresentado da Figura 24, em que se percebe a diferença de se calcular o valor médio de DSC por *slice* assumindo os *slices* compostos exclusivamente de verdadeiros negativos como tendo $swDSC = 1$ (BG), representado pelas barras em transparência que assumem valores que chegam a quase o dobro, e desconsiderando-se esses *slices* para o cálculo do valor médio de DSC por *slice* (noBG), representado pelas barras cheias.

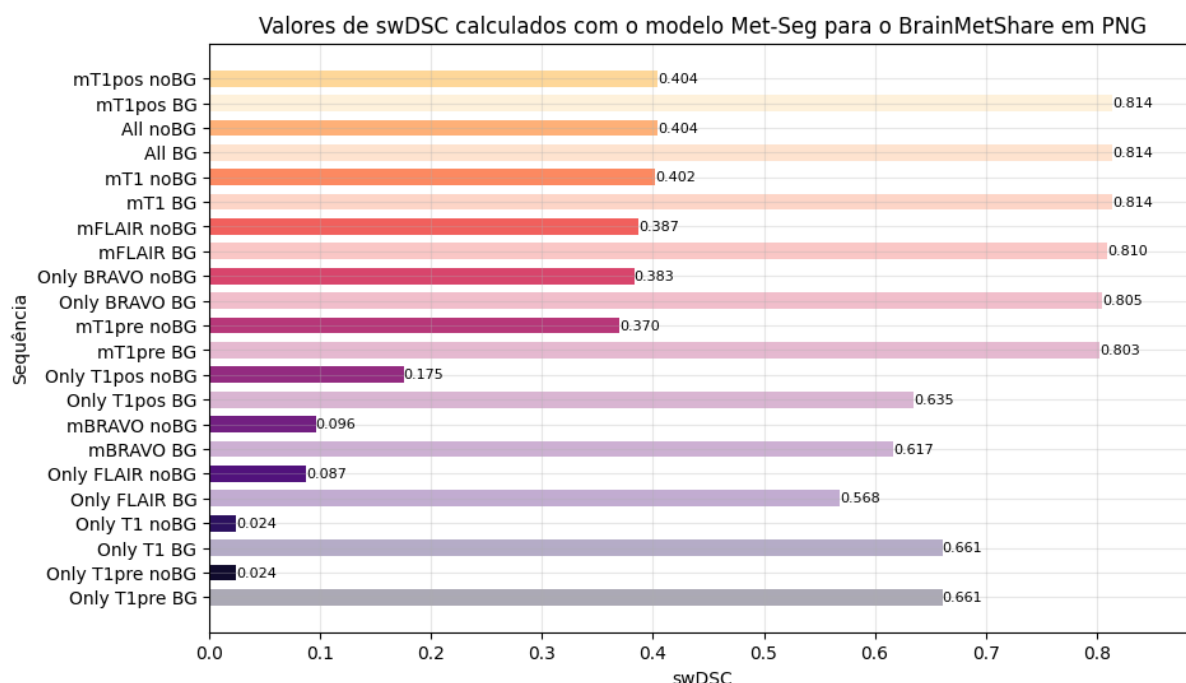


Figura 24 – Coeficiente Dice médio por *slice* $swDSC$ calculado em todas as combinações de sequências do modelo de (OTTESEN et al., 2023) para o banco de dados BMS em formato PNG considerando-se os *slices* compostos exclusivamente de verdadeiros negativos (BG) e desconsiderando-se esses *slices* (noBG) para o cálculo do valor de $swDSC$. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.

Fonte: Autoria própria.

Além disso, realizando uma análise de correlação entre os valores de $swDSC$ completo com todas as ponderações, e os valores de $swDSC$ obtidos nos demais casos apresentados nas Tabelas 8 e 9, foram encontrados os gráficos apresentados nas Figuras 25 e 26. Nesses gráficos, cada barra indica a correlação dos valores de $swDSC$ determinados para cada paciente com a remoção de alguma ponderação (em vermelho) ou apenas utilizando-se uma das ponderações (em verde) em comparação com o mesmo valor de $swDSC$ quando são utilizadas todas as ponderações para a predição. Note que as barras em cor verde indicam que a correlação entre a sequência única tem relevância para a predição do modelo como um todo, enquanto que as barras em cor vermelha indicam a

correlação com a remoção de uma sequência, isto é, que uma alta correlação significa que a sequência removida não possui relevância para a predição do modelo como um todo.

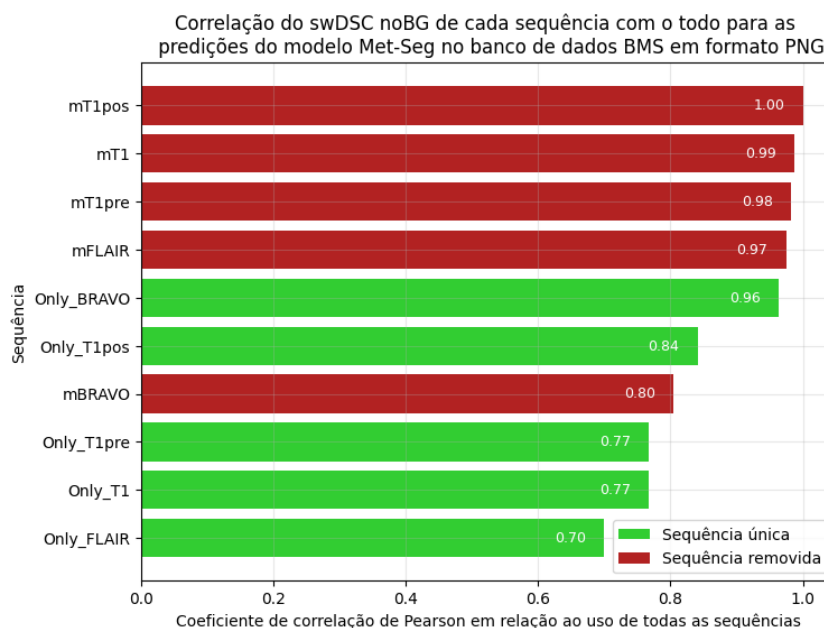


Figura 25 – Coeficientes de correlação de Pearson dos valores de swDSC do modelo de (OTTESEN et al., 2023) para o banco de dados BMS em seu formato PNG com a remoção de alguma modalidade em relação ao todo para o caso em que se considera os verdadeiros negativos dos *slices* sem metástases como tendo o valor de DSC = 1. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.

Fonte: *Autoria própria.*

A partir da análise desses gráficos de correlação, entendemos que a sequência T1 pós contraste (T1pos) não é relevante para realizar a predição, dado que a sua remoção implica a maior correlação em ambos os casos (mT1, em vermelho), indicando que realizar a predição com ou sem essa sequência pouco afeta o resultado da segmentação. No entanto, quando analisamos o uso apenas da sequência T1 pós contraste, sua correlação é maior do que utilizar apenas a sequência FLAIR, apenas a sequência T1 pré contraste, e ambas as sequências T1 pré e pós contraste. Em contrapartida, a sequência BRAVO é a que possui menor correlação quando removida (mBRAVO, em vermelho) e maior correlação quando utilizada unicamente (Only BRAVO, em verde), confirmando através de uma segunda análise sua relevância para uma predição mais eficiente.

Para o banco de dados BMS em seu formato nativo NifTI, foi realizada essa mesma análise de correlação para as modalidades faltantes, o que resultou no gráfico da Figura 27.

A partir desse gráfico, fica evidente a sequência BRAVO possui forte correlação com o resultado do modelo de modo geral, e que a remoção das outras modalidades que não são a BRAVO pouco influenciam no resultado da segmentação do modelo. Com isso,

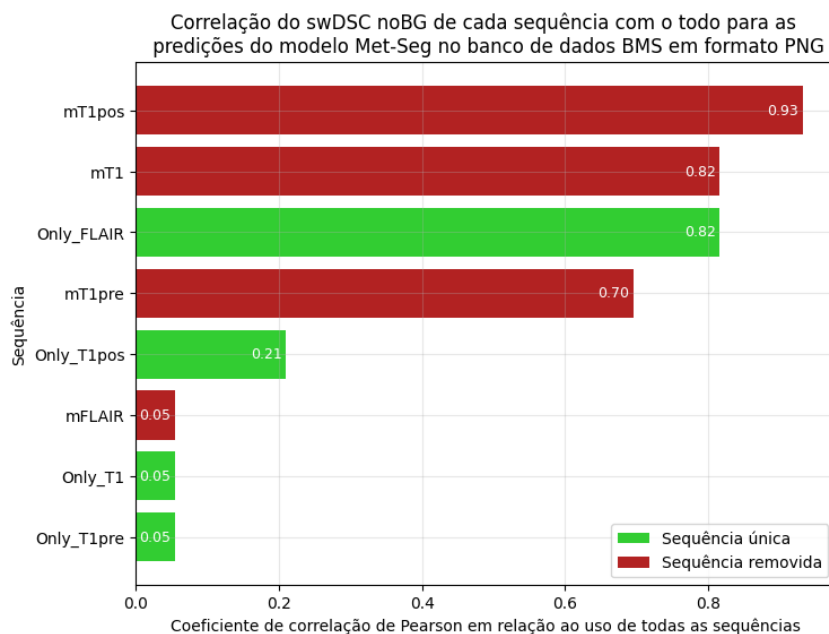


Figura 26 – Coeficientes de correlação de Pearson dos valores de DSC do modelo de (OT-TESEN et al., 2023) para o banco de dados do BMS em seu formato PNG com a remoção de alguma modalidade em relação ao todo para o caso em que não se considera os verdadeiros negativos dos *slices* sem metástases. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.

Fonte: *Autoria própria.*

é possível afirmar que o modelo Met-Seg é fortemente dependente da sequência BRAVO para realizar uma boa predição.

Sendo assim, essa mesma análise foi realizada para o banco de dados disponibilizado no *The Cancer Imaging Archive* (TCIA) por (RAMAKRISHNAN et al., 2024), que possui imagens de 200 pacientes obtidas nas sequências FLAIR, T1 pré contraste e T1 pós contraste. Como a sequência BRAVO não estava disponível nesse *dataset*, a comparação poderá ser realizada apenas para as sequências FLAIR e T1.

De forma similar ao que foi realizado para o banco de dados *BrainMetShare*, na Tabela 10 encontram-se os valores do coeficiente Dice para as diferentes combinações de sequências considerando-se os *pixels* pretos que são os verdadeiros negativos (BG) e desconsiderando-se os *pixels* verdadeiros negativos (noBG) e, na Tabela 11, encontram-se a comparação entre os bancos TCIA e *BrainMetShare* para os valores encontrados para swDSC de todos os pacientes.

Os resultados da Tabela 10 apontam para a importância da sequência FLAIR no banco de dados do TCIA, que reduz os valores de DSC caso seja removida dos arquivos utilizados para a predição, quando comparada às sequências T1 pré (T1pre) e pós (T1pos) aplicação de contraste, com uma pequena maior relevância para a sequência T1 pré contraste. Além disso, nota-se que os valores para o DSC desconsiderando-se os *slices*

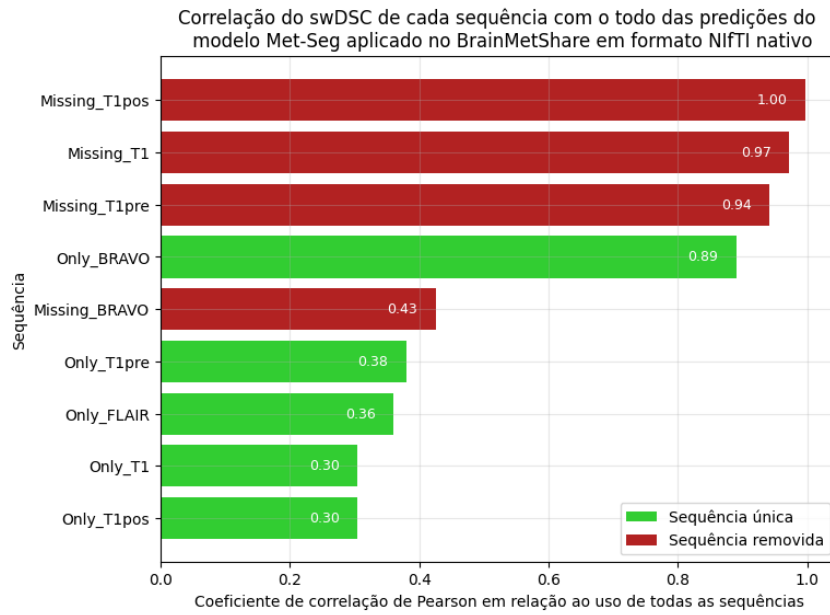


Figura 27 – Coeficientes de correlação de Pearson dos valores de swDSC do modelo Met-Seg (OTTESEN et al., 2023) para o banco de dados do BMS em formato NIfTI nativo com a remoção de alguma modalidade em relação ao todo para o caso em que não se considera os verdadeiros negativos dos *slices* sem metástases. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.

Fonte: Autoria própria.

Tabela 10 – Métricas de performance do coeficiente Dice médio por *slice* (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com algumas das modalidades faltantes, denotadas por mFLAIR, mT1pos e mT1pre para, respectivamente, os casos com a remoção das sequências FLAIR, T1 pós contraste e T1 pré contraste do modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA.

$swDSC_{mFLAIR}$		$swDSC_{mT1pos}$		$swDSC_{mT1pre}$	
BG	noBG	BG	noBG	BG	noBG
$0,74 \pm 0,08$	$0,02 \pm 0,05$	$0,77 \pm 0,08$	$0,07 \pm 0,09$	$0,78 \pm 0,09$	$0,08 \pm 0,11$

Fonte: Autoria própria.

compostos exclusivamente por *pixels* verdadeiros negativos (noBG) estão bastante abaixo dos valores considerando-se esses *slices* (BG). Para esse banco de dados, que possui um total de 31000 *slices* sendo 10600 com metástases e 20400 sem metástases, utilizando-se o mesmo algoritmo nunca treinado, cuja única função seja realizar a predição de *slices* pretos, acertaria 20400 das 31000 predições, e seu coeficiente Dice médio por *slice* teria o valor de $\frac{20400}{31000} = 0,65806451612$, o que é um valor com cerca de 12% de variação dos resultados do modelo encontrados considerando-se os *slices* totalmente verdadeiro positivos (BG). No entanto, os resultados para a média de DSC desconsiderando-se esses *slices* apresentam-se abaixo de 0,12.

Tabela 11 – Métricas de performance do coeficiente Dice (DSC) médio por slice para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com a presença de todas as sequências utilizando o modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA e o *BrainMetShare*.

$\text{swDSC}_{\text{TCIA}}$		$\text{swDSC}_{\text{BrainMetShare}}$		$\text{swDSC}_{\text{mBRAVO(BMS)}}$	
BG	noBG	BG	noBG	BG	noBG
$0,77 \pm 0,08$	$0,07 \pm 0,09$	$0,81 \pm 0,13$	$0,40 \pm 0,16$	$0,62 \pm 0,18$	$0,10 \pm 0,09$

Fonte: Autoria própria.

Já os resultados da Tabela 11 confirmam que a sequência BRAVO é crucial para o modelo Met-Seg realizar a segmentação de metástases, e que a sua ausência no banco de dados do TCIA é comparável com a remoção dessa sequência no banco de dados utilizado no treino do modelo, o *BrainMetShare*.

Quando considerados os usos únicos de cada uma das sequências, os valores de DSC encontrados estão apresentados na Tabela 12.

Tabela 12 – Métricas de performance do coeficiente Dice médio por *slice* (swDSC) para o caso considerando-se o fundo (BG) e desconsiderando-se o fundo (noBG), para as predições com apenas uma das sequências FLAIR, T1 pós contraste (T1pos) e T1 pré contraste (T1pre) do modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA.

$\text{swDSC}_{\text{FLAIR}}$		$\text{swDSC}_{\text{T1pos}}$		$\text{swDSC}_{\text{T1pre}}$	
BG	noBG	BG	noBG	BG	noBG
$0,75 \pm 0,08$	$0,07 \pm 0,10$	$0,82 \pm 0,09$	$0,20 \pm 0,17$	$0,74 \pm 0,08$	$0,02 \pm 0,05$

Fonte: Autoria própria.

Esses resultados apontam uma relevância levemente maior para a sequência FLAIR, seguida pela sequência T1 pós contraste e com a sequência T1 pré contraste sendo a menos relevante. Essa pequena variação na ordem de relevância das sequências quando comparada com a análise feita com a remoção das sequências pode ser explicada pelo fato de não existir uma sequência que seja mais importante do que outra nessa análise. Afinal, a sequência que se mostrou mais relevante no estudo realizado para o banco de dados *BrainMetShare* foi a sequência BRAVO, que não está presente no banco de dados do TCIA, e as sequências FLAIR e T1 não se mostraram significativamente importantes para a predição.

Uma análise de correlação entre os valores de swDSC de todos os pacientes do banco de dados do TCIA também foi realizado para cada uma das ponderações, e também com suas remoções, com a exceção da sequência BRAVO que não pertencem a este banco de dados. O gráfico de coeficientes de correlação do swDSC em relação ao uso de todas as sequências para o caso em que não se considera o fundo da imagem está apresentado na Figura 28, onde é possível observar a influência da sequência FLAIR na ausência da

sequência BRAVO.

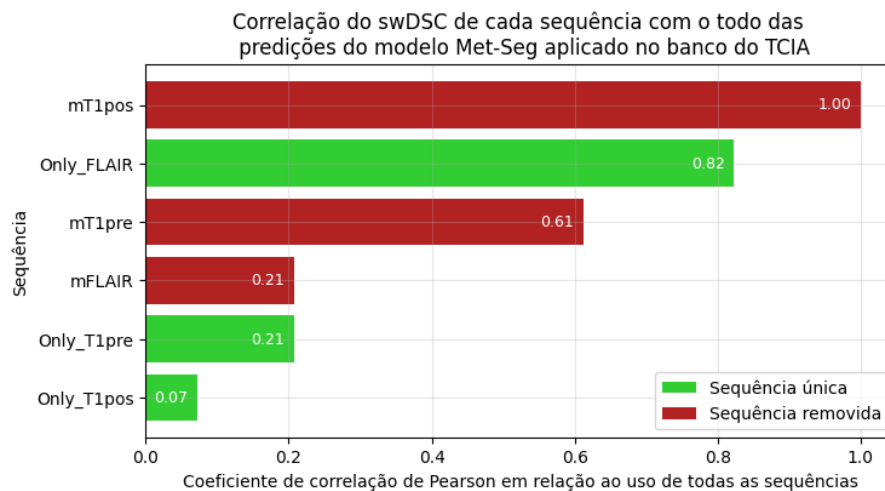


Figura 28 – Valores dos coeficientes de correlação de Pearson para os valores dos coeficientes Dice médio por *slice* do modelo Met-Seg para o banco de dados do TCIA entre os pares com e sem a remoção de alguma(s) modalidade(s) de imagem, para o caso em que não se considera os verdadeiros negativos dos *slices* sem metástases. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.

Fonte: *Autoria própria.*

Os valores de swDSC calculados em todas as combinações de sequências também está apresentado da Figura 29, em que percebe-se a significativa diferença de se calcular o valor médio de DSC por *slice* assumindo os *slices* compostos exclusivamente de verdadeiros negativos como tendo DSC = 1 (BG) e desconsiderando-se esses *slices* para o cálculo do valor médio de DSC por *slice* (noBG).

Para esse banco de dados, essa diferença foi mais perceptível e nota-se o valor de swDSC calculado para a sequência T1 pós contraste sendo usada unicamente (swDSC = $0,20 \pm 0,17$), que é mais de duas vezes maior do que os valores de swDSC calculados para a sequência FLAIR (swDSC = $0,07 \pm 0,10$), dez vezes maior do que para a sequência T1 pré contraste (swDSC = $0,02 \pm 0,05$) e inclusive para a que usa todas as sequências juntas (swDSC = $0,07 \pm 0,09$). O mesmo ocorre no caso do *dataset BrainMetShare* desconsiderando a sequência BRAVO, em que o swDSC da sequência T1 pós contraste (swDSC = $0,18 \pm 0,13$) é também duas vezes maior do que o swDSC calculado utilizando somente a sequência FLAIR (swDSC = $0,09 \pm 0,08$) e nove vezes maior do que o swDSC da sequência T1 pré contraste (swDSC = $0,02 \pm 0,07$). Isso indica que a sequência T1 pós contraste apresenta informações que são importantes para a segmentação, apesar de possuir um baixo valor para o coeficiente Dice médio por *slice* (swDSC = $0,20 \pm 0,17$), o que reforça a afirmação de Liu *et al.* (2017) sobre o uso único da sequência T1 com contraste para a identificação e consequente segmentação da região metastática. Além disso, ressalta-se que o valor de swDSC calculado utilizando-se apenas a sequência T1 pós contraste é li-

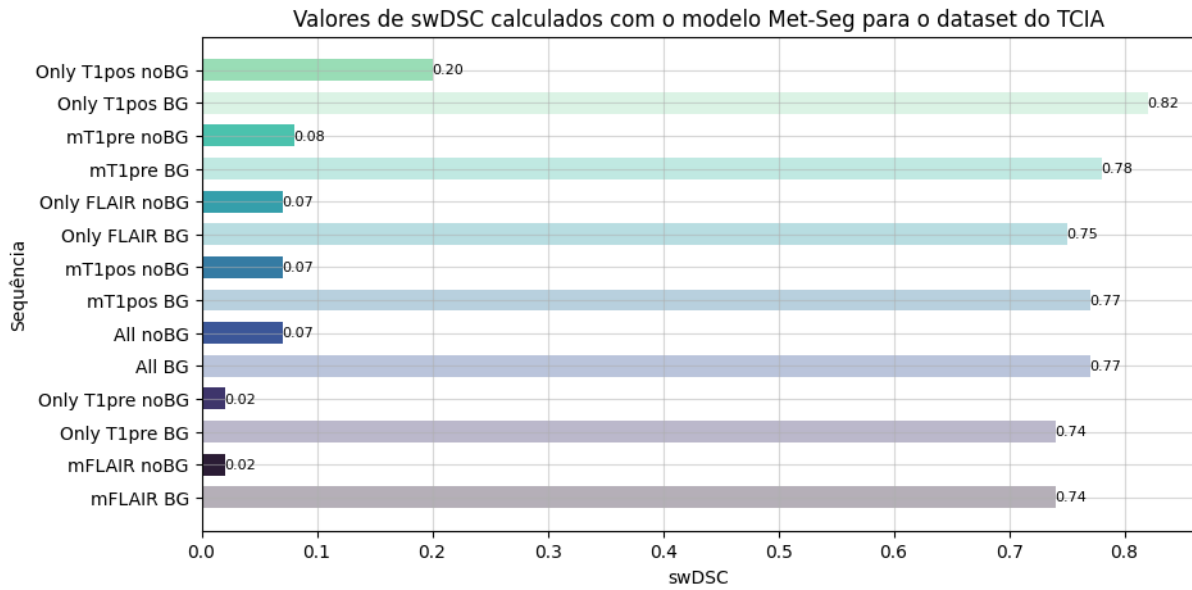


Figura 29 – Coeficiente Dice calculado em todas as combinações de sequências do modelo de (OTTESEN et al., 2023) para o banco de dados do TCIA considerando-se os *slices* compostos exclusivamente de verdadeiros negativos (BG) e desconsiderando-se esses *slices* (noBG) para o cálculo do valor médio de DSC por *slice*. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.

Fonte: Autoria própria.

geiramente maior para a aplicação no *dataset* do TCIA ($\text{swDSC} = 0,20 \pm 0,17$) do que para o *BrainMetShare* ($\text{swDSC} = 0,18 \pm 0,13$). Essa diferença absoluta dos valores do coeficiente Dice para os dois bancos de dados é explicada pelo fato do modelo de Ottesen et al. (2023) ter sido treinado com as imagens do banco de dados *BrainMetShare* e, por isso, apresenta melhores resultados já que a análise realizada neste trabalho é praticamente um resultado de validação. Enquanto isso, os resultados obtidos com o banco de dados do TCIA para o modelo de Ottesen et al. (2023) são resultados de teste, visto que as imagens apresentadas ao modelo não foram utilizadas em seu treinamento, e também possuem características de aquisição distintas, conforme foi apresentado na Tabela 3 do Capítulo 5. Por esse motivo, os resultados do modelo Met-Seg (OTTESEN et al., 2023) não foram muito satisfatórios para o banco de dados do TCIA e, apesar de acertar uma quantidade razoável de predições, em alguns outros casos o modelo realiza um número maior de predições, para regiões em que não há presença de metástase, conforme é possível observar em um caso exemplo na Figura 30.

6.2.2 Modelo AURORA, desenvolvido por Buchner et al. (2023)

Para o modelo AURORA (BUCHNER et al., 2023), foi realizada a análise através do coeficiente Dice volumétrico (3D-DSC), que foi a métrica apresentada pelos autores. Os valores obtidos de 3D-DSC para os diferentes bancos de dados estão apresentados na

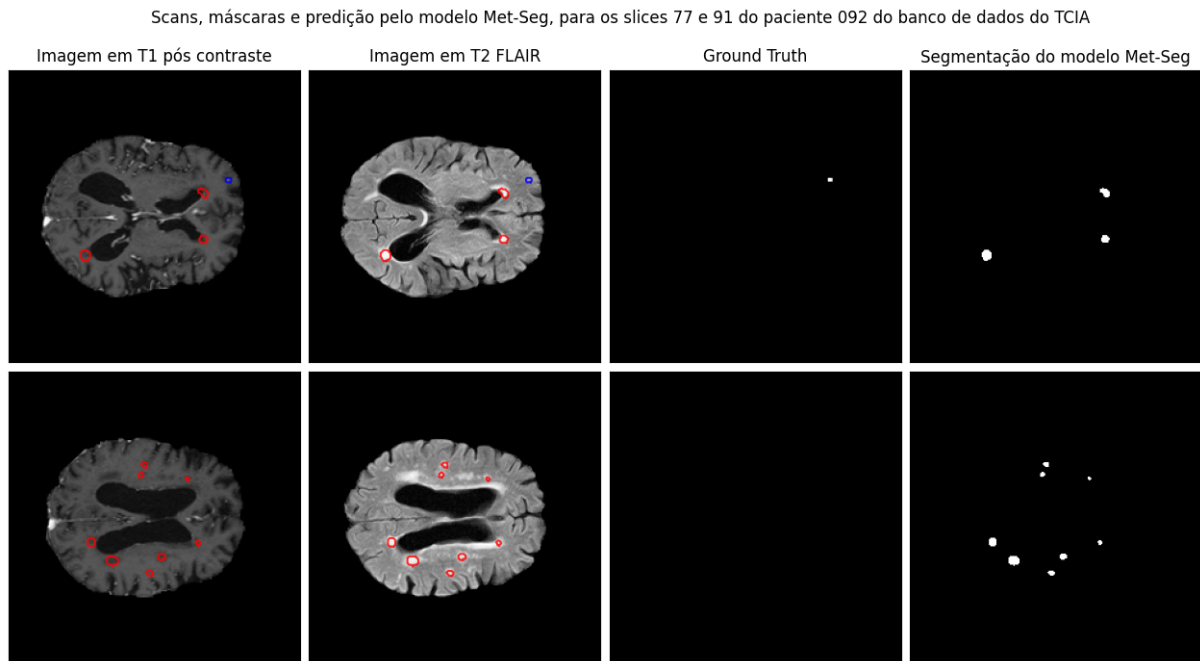


Figura 30 – Exemplo para dois *slices* em um caso do paciente de índice 92 do banco de dados do TCIA, em que o modelo Met-Seg realiza erroneamente (em vermelho) as segmentações de volumes tumorais. Em azul, nas duas primeiras imagens do topo, está destacada a região com presença metastática delimitada por radiologistas (*ground truth*).

Fonte: Autoria própria.

Tabela 13.

Tabela 13 – Comparação dos resultados do coeficiente Dice volumétrico (3D-DSC) obtidos para as predições realizadas pelo modelo AURORA (BUCHNER *et al.*, 2023) utilizando todas as modalidades de imagens presentes nos bancos de dados dos autores, BMS, TCIA e Ocaña-Tienda *et al.* (2023).

Banco de dados	3D-DSC
Buchner <i>et al.</i> (2023)	$0,92 \pm 0,07$
BMS PNG	$0,20 \pm 0,26$
BMS NIFTI	$0,41 \pm 0,31$
TCIA	$0,53 \pm 0,36$
Ocaña-Tienda <i>et al.</i> (2023)	$0,56 \pm 0,36$

Fonte: Autoria própria.

Nota-se que o valor obtido para o coeficiente Dice volumétrico no banco de dados próprio dos autores foi superior aos obtidos nos bancos de dados apresentados neste estudo, indicando que o modelo precisa passar por uma melhoria em sua robustez para que possua uma boa performance em novas imagens. Além disso, é possível que a conversão das imagens em PNG para NIFTI resultem em perda de informação, visto que o modelo teve os piores resultados para o banco de dados BMS em seu formato original em PNG.

Como o modelo AURORA permite a combinação de diferentes modalidades de

imagem no *input* do modelo, foram calculados os valores do coeficiente Dice volumétrico dessas diferentes combinações para os bancos BMS e TCIA. Sendo assim, foram obtidos os resultados apresentados na Tabela 14

Tabela 14 – Média dos coeficientes Dice volumétricos (3D-DSC) das predições do modelo AURORA (BUCHNER et al., 2023) para o banco de dados BMS em seu formato nativo NIfTI com a remoção da sequência T1 pré-contraste (mT1pre), utilizando apenas a sequência FLAIR, apenas as sequências T1 pré e pós contraste (T1) e apenas a sequência T1 pré (T1pre) e pós (T1pos) contraste. Em todos os casos a sequência BRAVO não foi utilizada, pois o modelo não foi treinado para esse tipo de imagem.

$3D-DSC_{T1pre}$	$3D-DSC_{mT1pre}$	$3D-DSC_{FLAIR}$	$3D-DSC_{T1}$	$3D-DSC_{T1pos}$
$0,15 \pm 0,24$	$0,39 \pm 0,30$	$0,19 \pm 0,26$	$0,41 \pm 0,33$	$0,40 \pm 0,33$

Fonte: Autoria própria.

Da mesma maneira que foi feita para o modelo Met-Seg, uma análise de correlação de Pearson foi realizada exame a exame para a remoção de algumas sequências, porém com a informação do 3D-DSC em vez do swDSC, pois os autores do modelo AURORA utilizaram a métrica do coeficiente Dice volumétrico em sua publicação. Essa análise pode ser encontrada na Figura 31, em que é possível observar uma correlação muito forte ($p = 0,92$) entre o uso das sequências T1 com e sem contraste e o uso de todas as sequências, indicando que essas duas sequências possuem maior influência na hora da inferência realizada pelo modelo. Além disso, nota-se que o uso apenas da sequência T1 pós contraste é similar ao uso das sequências FLAIR e T1 pós contraste combinadas (mT1pre), mantendo uma correlação forte ($p = 0,82$ e $p = 0,81$), e que a adição da sequência FLAIR não gera uma melhoria no resultado do modelo, o que pode ser observado na correlação moderada ($p = 0,60$) entre o uso único da sequência FLAIR e o uso de todas as sequências. Sendo assim, a combinação da sequência T1 pós contraste com as sequências T1 pré contraste ou T2 FLAIR gerou melhores resultados do que o uso individual da sequência T1 pós contraste.

O uso da métrica 3D-DSC em vez da swDSC não apresenta uma variação significativa da análise de dados, pois as duas métricas também estão fortemente correlacionadas em um valor de coeficiente de Pearson médio de $0,931 \pm 0,016$ quando comparados as correlações de swDSC e 3D-DSC para todos os modelos em todos os bancos de dados. Sendo assim, as análises realizadas para as métricas 3D-DSC e swDSC se tornam análogas. Como algumas publicações utilizam a métrica IoU e sua equação é similar à equação de DSC, também foi realizada uma análise adicional para a correlação entre IoU e swDSC, o índice de correlação para essas métricas resultou em um valor de $0,941 \pm 0,018$, indicando correlação muito forte. Dessa maneira, concluí-se que a apresentação de apenas uma dessas três métricas é suficiente para a análise da performance da segmentação.

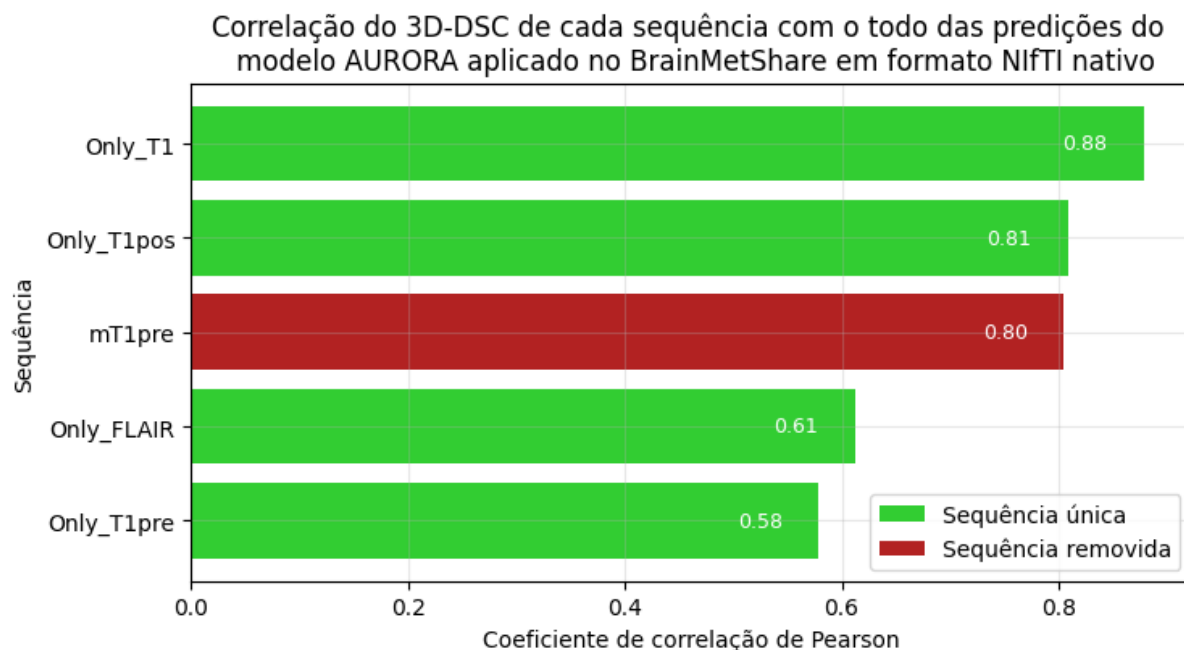


Figura 31 – Valores dos coeficientes de correlação de Pearson para os valores dos coeficientes Dice volumétricos (3D-DSC) do modelo AURORA para o banco de dados BMS em seu formato NIfTI nativo entre os pares com e sem a remoção de alguma(s) modalidade(s) de imagem. Only indica que apenas aquela sequência foi utilizada, e a letra m indica que a sequência seguida da letra m foi removida.

Fonte: Autoria própria.

6.2.3 Modelo Raidionics, desenvolvido por Bouget *et al.* (2023)

As inferências realizadas pelo modelo Raidionics (BOUGET *et al.*, 2023) só foram possíveis de serem realizadas para a sequência T1 pós contraste e os resultados para o coeficiente Dice estão apresentados na Tabela 15, que compara os valores encontrados para os bancos de dados deste estudo com o banco de dados dos autores do modelo, que separam os resultados de performance entre geral, para metástases menores do que 2 mL e metástases maiores ou de tamanho igual à 2 mL. Nesta Tabela, pode-se observar que a performance geral do modelo apresentada pelos autores sofre um decréscimo de 29% para os bancos de dados TCIA e Ocaña-Tienda *et al.* (2023), sem muita variação de performance nesses dois bancos que possuem, respectivamente, um valor médio de 4,36 mL e 3,40 mL para a mediana do tamanho tumoral por paciente, que são de 4 a 5 vezes menores do que o tamanho das lesões na qual o modelo foi inicialmente testado. Esse resultado aponta para uma robustez do modelo para diferentes bancos de dados, uma vez que possui uma performance similar para bancos diferentes, e é condizente com o esperado pelo que foi apontado pelos autores, que informam um decréscimo na performance com a diminuição do tamanho tumoral. Já para o banco BMS, que possui um valor médio para a mediana do volume tumoral por paciente de 0,63 mL, esse resultado é reduzido em 57% para o caso geral e fica 47% abaixo do valor obtido pelos autores para o caso

de metástases menores de 2 mL, o que confirma a dificuldade do modelo em realizar a segmentação de metástases pequenas.

Tabela 15 – Comparação dos resultados do coeficiente Dice volumétrico (3D-DSC) obtidos para as predições realizadas pelo modelo Raidionics (BOUGET et al., 2023) utilizando apenas a modalidade T1 pós contraste dos bancos de dados dos autores, BMS, TCIA e Ocaña-Tienda *et al.* (2023).

Banco de dados	Volume do tumor (V)	3D-DSC
	$\bar{V} = 17$ mL	$0,866 \pm 0,191$
Bouget et al. (2023)	$V < 2$ mL	$0,701 \pm 0,309$
	$V \geq 2$ mL	$0,895 \pm 0,132$
BMS NIFTI	$\bar{V} = 0,63$ mL	$0,374 \pm 0,310$
TCIA	$\bar{V} = 4,36$ mL	$0,611 \pm 0,346$
Ocaña-Tienda <i>et al.</i> (2023)	$\bar{V} = 3,40$ mL	$0,610 \pm 0,354$

Fonte: Autoria própria.

6.2.4 Análise dos *checklists* CLAMP e CLAIM para as publicações envolvendo os modelos da literatura

Foram verificados os itens dos *checklists* nas publicações dos três modelos: Met-Seg (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2023) e Raidionics (BOUGET et al., 2023). Os itens marcados com o *checkmark* verde (✓) indicam que o artigo apresentou o item conforme descrito pelos autores dos *checklists*, enquanto que os itens marcados com a cruz vermelha (✗) indicam que o artigo não apresentou ou que apresentou parcialmente o item conforme descrito pelos autores dos *checklists*. No final, há soma total de itens que estavam de acordo com cada *checklist*.

Sendo assim, pode-se dizer que a publicação envolvendo o modelo AURORA (BUCHNER et al., 2023) possui uma maior qualidade de apresentação dos resultados, uma vez que consta com 64% dos itens de ambos os *checklists* de acordo com sua publicação. A publicação envolvendo o modelo Met-Seg (OTTESEN et al., 2023) consta com 51% de itens de acordo com ambos os *checklists* e a publicação envolvendo o modelo Raidionics (BOUGET et al., 2023) consta com 26% de itens de acordo com os *checklists*, sendo essa a que menos se encontra adequada pela análise através dos *checklists*. Apesar de nenhuma publicação ultrapassar a margem de dois terços de adequação dos *checklists*, as revistas nas quais foram publicadas possuem um bom fator de impacto, de valores 4,9 para a publicação do modelo AURORA, 2,5 para a publicação do modelo Met-Seg e 3,8 para a publicação do modelo Raidionics.

Quadro 1 – Checklists CLAIM e CLAMP para as publicações dos três modelos.

Checklist	CLAIM			CLAMP		
	Met-Seg	AURORA	Raidionics	Met-Seg	AURORA	Raidionics
Item 1	✓	✓	×	×	✓	✓
Item 2	×	×	×	×	✓	×
Item 3	✓	✓	×	×	×	×
Item 4	×	✓	×	✓	✓	×
Item 5	✓	✓	×	✓	✓	✓
Item 6	×	✓	✓	✓	✓	✓
Item 7	✓	✓	×	✓	✓	×
Item 8	✓	✓	×	×	×	×
Item 9	✓	✓	✓	✓	×	×
Item 10	×	✓	✓	✓	✓	×
Item 11	×	×	×	N/A	N/A	N/A
Item 12	✓	✓	×	✓	✓	✓
Item 13	×	✓	×	×	×	×
Item 14	×	×	×	✓	✓	×
Item 15	×	×	×	✓	N/A	N/A
Item 16	×	×	×	×	✓	✓
Item 17	×	✓	×	✓	✓	×
Item 18	×	×	×	✓	✓	✓
Item 19	✓	×	×	✓	✓	×
Item 20	✓	✓	✓	×	×	×
Item 21	×	×	×	×	×	×
Item 22	×	✓	✓	✓	✓	×
Item 23	×	✓	✓	✓	✓	×
Item 24	×	×	×	✓	✓	✓
Item 25	×	✓	×	✓	✓	×
Item 26	×	×	×	×	✓	✓
Item 27	N/A	N/A	N/A	-	-	-
Item 28	×	✓	×	-	-	-
Item 29	×	×	×	-	-	-
Item 30	✓	✓	×	-	-	-
Item 31	✓	✓	×	-	-	-
Item 32	✓	✓	×	-	-	-
Item 33	✓	✓	×	-	-	-
Item 34	✓	×	×	-	-	-
Item 35	×	×	×	-	-	-
Item 36	✓	×	×	-	-	-
Item 37	×	×	×	-	-	-
Item 38	✓	✓	✓	-	-	-
Item 39	✓	✓	×	-	-	-
Item 40	✓	✓	×	-	-	-
Item 41	×	✓	✓	-	-	-
Item 42	✓	✓	×	-	-	-
Item 43	✓	✓	✓	-	-	-
Item 44	×	×	✓	-	-	-
✓ Total	20	27	10	16	18	8

Fonte: Autoria própria.

6.2.5 Comparação dos resultados obtidos nos três modelos da literatura

Os resultados para os coeficientes Dice volumétricos (3D-DSC) dos três modelos para os três bancos de dados utilizando todas as sequências possíveis estão apresentados na Tabela 16 e, na Tabela 17, encontra-se os mesmos resultados para o uso apenas da sequência T1 pós contraste, que é a única sequência disponível no banco de dados de Ocaña-Tienda *et al.* (2023).

Tabela 16 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelos modelos Met-Seg, AURORA e Raidionics nos bancos de dados *BrainMetShare* em formato original PNG, *BrainMetShare* em formato original NIfTI, TCIA e Ocaña-Tienda *et al.* (2023).

Banco de dados	Met-Seg	AURORA	Raidionics
BMS PNG	$0,58 \pm 0,21$	$0,20 \pm 0,26$	-
BMS NIfTI	$0,75 \pm 0,13$	$0,41 \pm 0,31$	$0,37 \pm 0,31$
TCIA	$0,12 \pm 0,17$	$0,53 \pm 0,36$	$0,61 \pm 0,35$
Ocaña-Tienda <i>et al.</i> (2023)	$0,50 \pm 0,30$	$0,56 \pm 0,36$	$0,61 \pm 0,35$

Fonte: Autoria própria.

Tabela 17 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelos modelos Met-Seg, AURORA e Raidionics nas imagens em ponderação T1 com contraste dos bancos de dados *BrainMetShare* em formato original NIfTI, TCIA e Ocaña-Tienda *et al.* (2023).

Banco de dados	Met-Seg	AURORA	Raidionics
BMS NIfTI	$0,48 \pm 0,21$	$0,40 \pm 0,33$	$0,37 \pm 0,31$
TCIA	$0,34 \pm 0,27$	$0,68 \pm 0,33$	$0,61 \pm 0,35$
Ocaña-Tienda <i>et al.</i> (2023)	$0,50 \pm 0,30$	$0,56 \pm 0,36$	$0,61 \pm 0,35$

Fonte: Autoria própria.

A partir da Tabela 17, pode-se ver que quando as mesmas condições são oferecidas para todos os modelos, isto é, no caso em que apenas a sequência T1 pós contraste é utilizada, o modelo AURORA é o que possui a melhor performance de segmentação. Apesar disso, o modelo Met-Seg é o que apresentou melhor performance para metástases de menor volume, enquanto que o modelo Raidionics apresentou uma performance mais estável entre os bancos de dados que não são compostos por volumes tumorais de menor tamanho.

Com o intuito de verificar alguma relação entre o coeficiente Dice e algumas características da imagem, foram realizadas quatro análises envolvendo a quantidade de volumes tumorais, o volume tumoral total, a mediana dos volumes tumorais e o tamanho da imagem utilizada como *input*. Essas variáveis foram comparadas com os valores do swDSC médio por exame, e os gráficos com as distribuições de todos os modelos e banco de dados estarão dispostos a seguir, onde, na parte inferior, também é possível ter uma visão geral dessas distribuições para cada modelo e cada banco de dados isoladamente.

Na Figura 32, está apresentado o coeficiente Dice médio por *slice* em função do número de tumores (contabilizados até 100) para as predições dos três modelos nas imagens dos três bancos de dados, em que é possível observar o decréscimo do coeficiente Dice com o número de volumes tumorais. Para pequenas quantidades de volumes tumorais, não há uma relação explícita para a melhora da eficiência do modelo em realizar a segmentação.

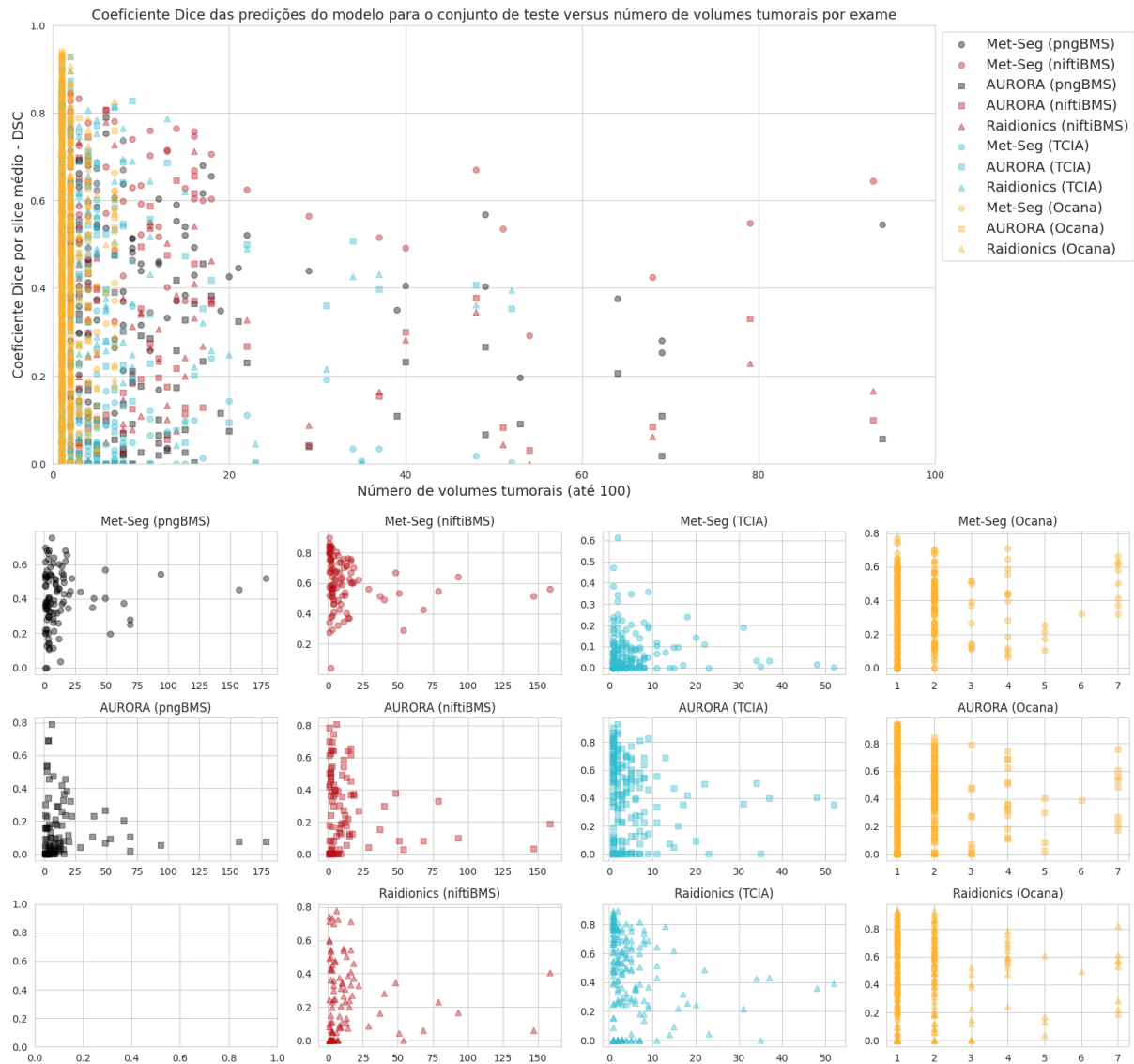


Figura 32 – Coeficiente Dice médio por *slice* versus a quantidade de volumes tumorais por paciente para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: *BrainMetShare* (em vermelho), TCIA (em azul) e Ocanã-Tienda (em amarelo).

Fonte: Autoria própria.

Na Figura 33, está apresentado o coeficiente Dice médio por *slice* em função do volume tumoral total para as predições dos três modelos nas imagens dos três bancos de dados, em que é possível observar que maiores volumes tumorais são mais facilmente

detectados pelos modelos AURORA (quadrados azuis) e Raidionics (triângulos azuis), já para o modelo Met-Seg (círculos azuis), os maiores volumes tumorais totais foram mais difíceis de serem detectados. Ressalta-se que o modelo Raidionics foi melhor do que o modelo Met-Seg para o banco de dados do BMS (em vermelho), que é o banco em que o modelo Met-Seg foi treinado. Esse padrão se repetiu para o gráfico do coeficiente Dice pela mediana dos volumes tumorais por exame, o que reforça que não apenas grandes volumes tumorais totais são mais dificilmente detectados pelo modelo Met-Seg, mas que metástases de tamanhos maiores também se mostram desafiadoras.

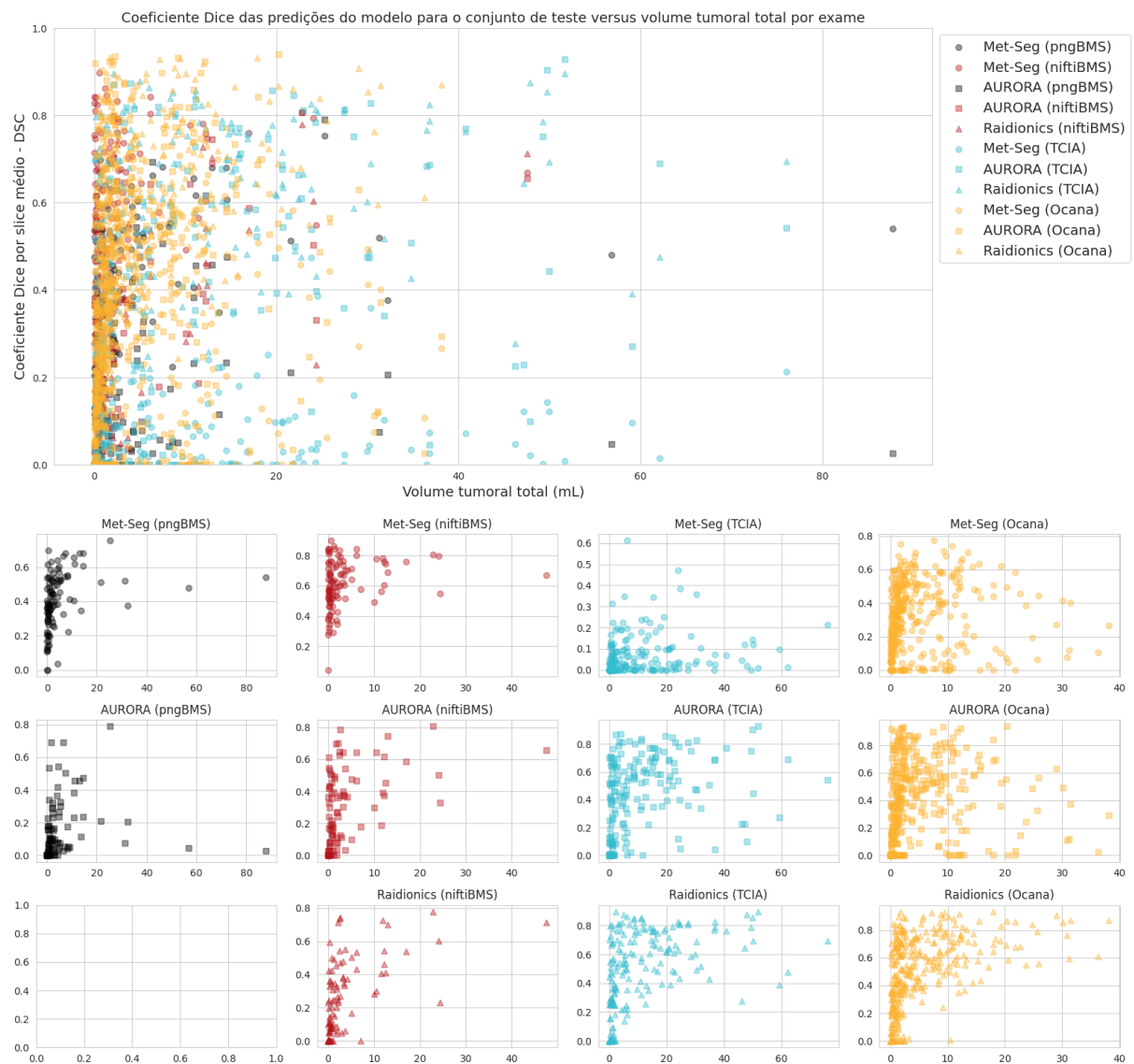


Figura 33 – Coeficiente Dice médio por *slice* versus a o volume total tumoral por paciente para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: *BrainMetShare* (em vermelho), TCIA (em azul) e Ocanã-Tienda (em amarelo).

Fonte: Autoria própria.

A respeito da influência do tamanho das imagens, o plot do coeficiente Dice médio

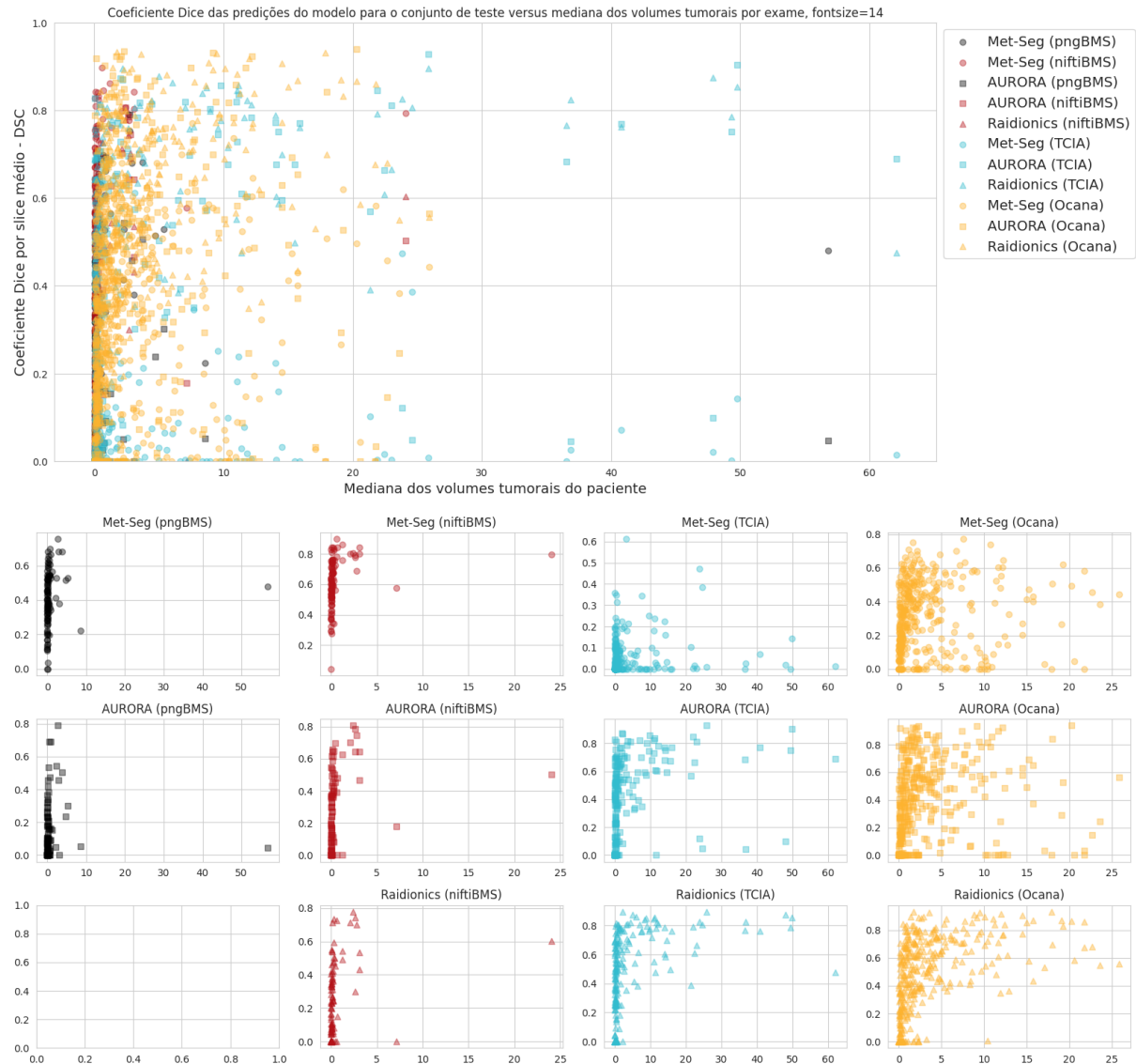


Figura 34 – Coeficiente Dice médio por *slice* versus a mediana de volumes tumorais por paciente para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: *BrainMetShare* (em vermelho), TCIA (em azul) e Ocaña-Tienda (em amarelo).

Fonte: Autoria própria.

por *slice* em função do tamanho em voxels está apresentado na Figura 35. A partir desse gráfico pode-se concluir que as imagens de maior tamanho possuem melhores resultados para o modelo Raidionics, o que indica que para esse modelo em específico, a resolução da imagem e a capacidade do modelo em realizar a segmentação podem estar associadas. No entanto, essa afirmativa não é verdadeira para os modelos AURORA e Met-Seg.

Além disso, como o banco de dados de Ocaña-Tienda *et al.* (2023) possui a informação da origem do tumor primário, foi possível realizar a distinção dos coeficientes Dice encontrados para cada uma das origens, o que encontra-se apresentado na Figura 36, em que é possível observar um maior coeficiente Dice para as metástases que possuem

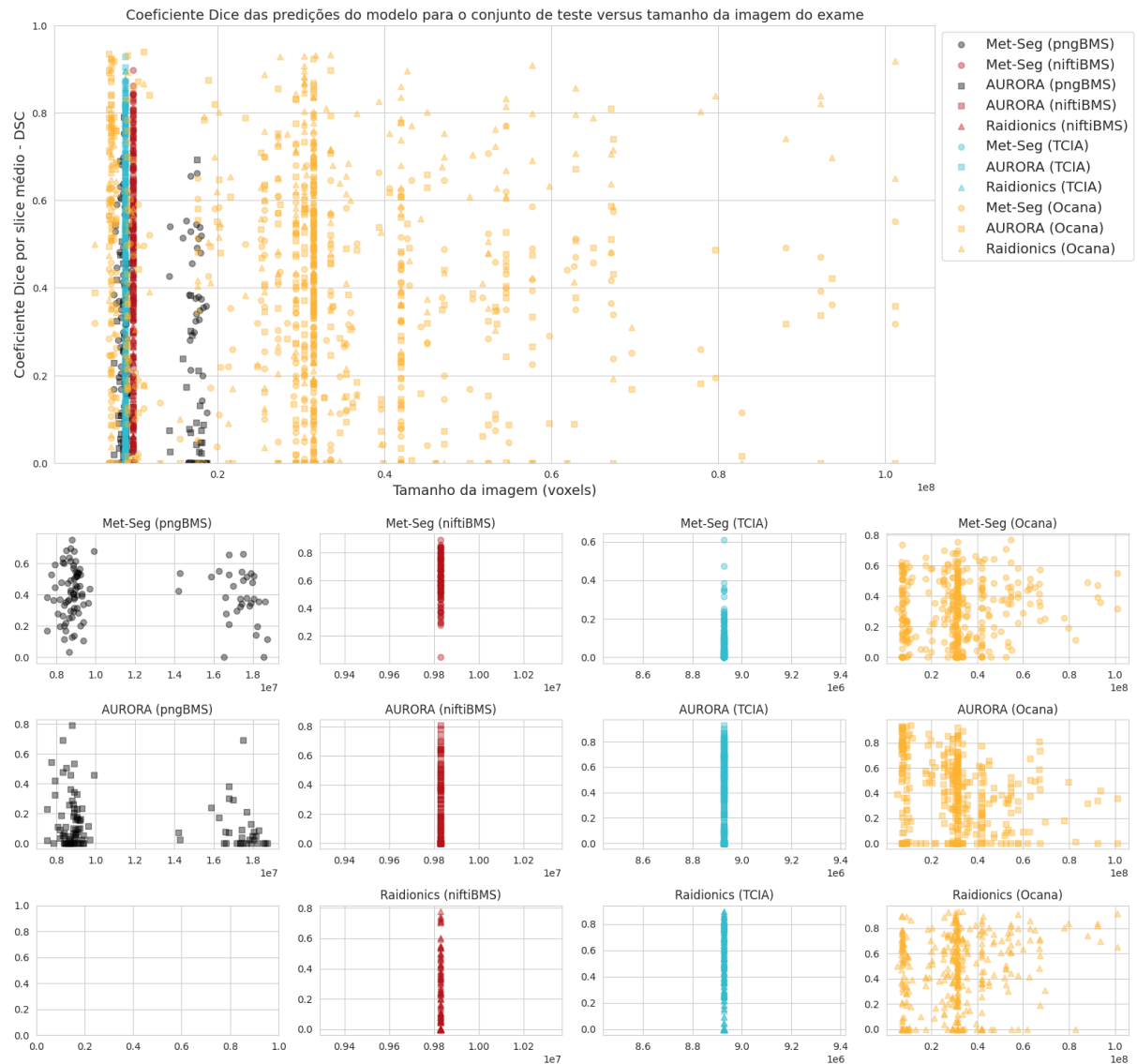


Figura 35 – Coeficiente Dice versus o tamanho da imagem do exame para os modelos Met-Seg (círculo) (OTTESEN et al., 2023), AURORA (quadrado) (BUCHNER et al., 2022) e Raidionics (triângulo) (BOUGET et al., 2023), para os três bancos de dados: *BrainMetShare* (em vermelho), TCIA (em azul) e Ocaña-Tienda (em amarelo).

Fonte: Autoria própria.

como tumor primário o Câncer de pulmão de pequenas células (SCLC, do inglês *Small-Cell Lung Cancer*), seguidos pelo Câncer de pulmão de células não pequenas (NSCLC, do inglês *Non-Small Cell Lung Cancer*), câncer de mama e outros tipos de câncer. Assim como informado por Buchner *et al.* (2022) em um estudo posterior à publicação do modelo AURORA, a performance do modelo foi superior para metástases oriundas de SCLC (BUCHNER et al., 2022). Além disso, fica evidente que o modelo Met-Seg possui resultados piores quando comparados com os resultados dos modelos AURORA e Raidionics, sendo esse último o melhor modelo para todos os tipos de tumor primário, no dataset de Ocaña-Tienda *et al.* (2023).

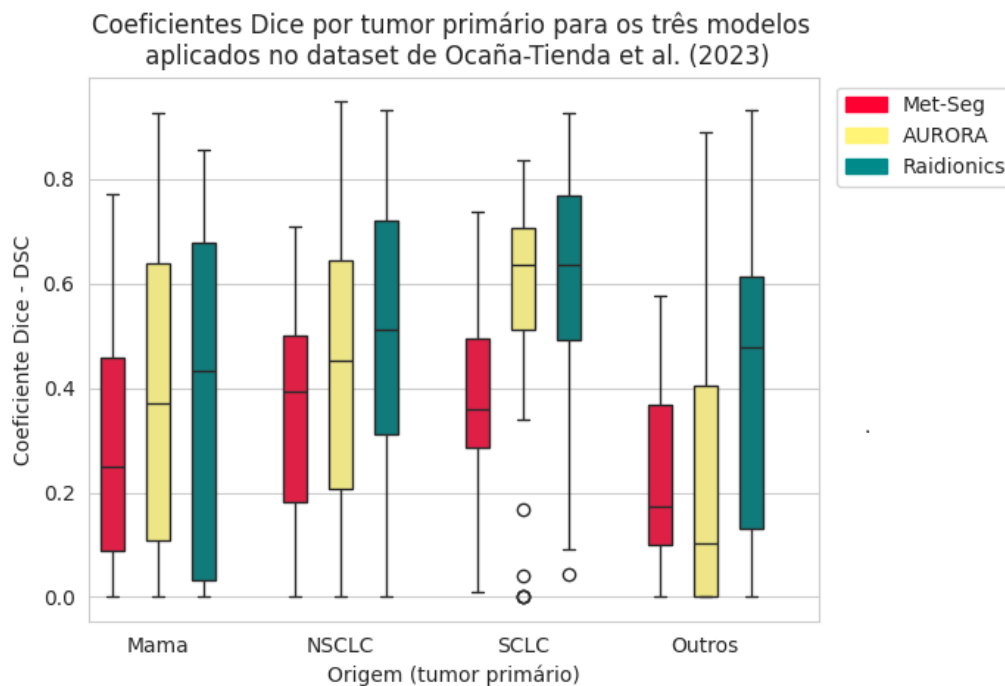


Figura 36 – Coeficientes Dice (DSC) médio por *slice* para as diferentes origens de tumor primário, calculados para os três modelos: Met-Set (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) no banco de dados Ocaña-Tienda (OCAÑA-TIENDA et al., 2023). SCLC - *Small Cell Lung Cancer*, NSCLC - *Non Small Cell Lung Cancer*.

Fonte: Autoria própria.

Ao analisar as Distâncias de Hausdorff das segmentações encontradas pelos três modelos para os três bancos de dados, apresentado na Figura 37, observa-se que o modelo Met-Seg apresentou um melhor resultado para os bancos de dados do BMS, tanto para a versão em PNG quanto para a versão em NIfTI, e para o TCIA, porém foi o pior modelo para o banco de dados de Ocaña-Tienda *et al.* (2023). Isso pode se dar pelo fato de que o modelo Met-Seg foi treinado com o banco de dados *BrainMetShare* que, por sua vez, é constituinte do banco de dados do TCIA. Além disso, o modelo Met-Seg apresentou menos *outliers* do que os outros dois modelos para todos os bancos de dados, e possui um menor número de falsos negativos na identificação de pacientes com metástases cerebrais. Vale observar que o banco de dados BMS em formato PNG apresentou uma distância de Hausdorff maior do que para o mesmo banco de dados em formato NIfTI, o que indica que o formato NIfTI é mais adequado para uso na tarefa de segmentação tumoral. As segmentações realizadas pelos modelos AURORA e Raidionics para o banco de dados do TCIA tiveram resultados relativamente melhores quando comparadas às do banco BMS, enquanto que para o modelo Met-Seg o resultado foi pior, o que, de novo, deve ser devido ao fato desse modelo ter sido treinado com o banco BMS. Em relação ao banco de dados de Ocaña-Tienda *et al.* (2023), as HD das predições foram as maiores para os três modelos, sendo Met-Seg o pior e Raidionics o melhor. Como esse banco de

dados é composto majoritariamente por metástase cerebral individual de grande volume, é possível que os modelos treinados com bancos de dados com outras características tenham uma performance pior, o que está de acordo com o apontado por Bousabarah *et al.* (2020) e Zhou *et al.* (2020).

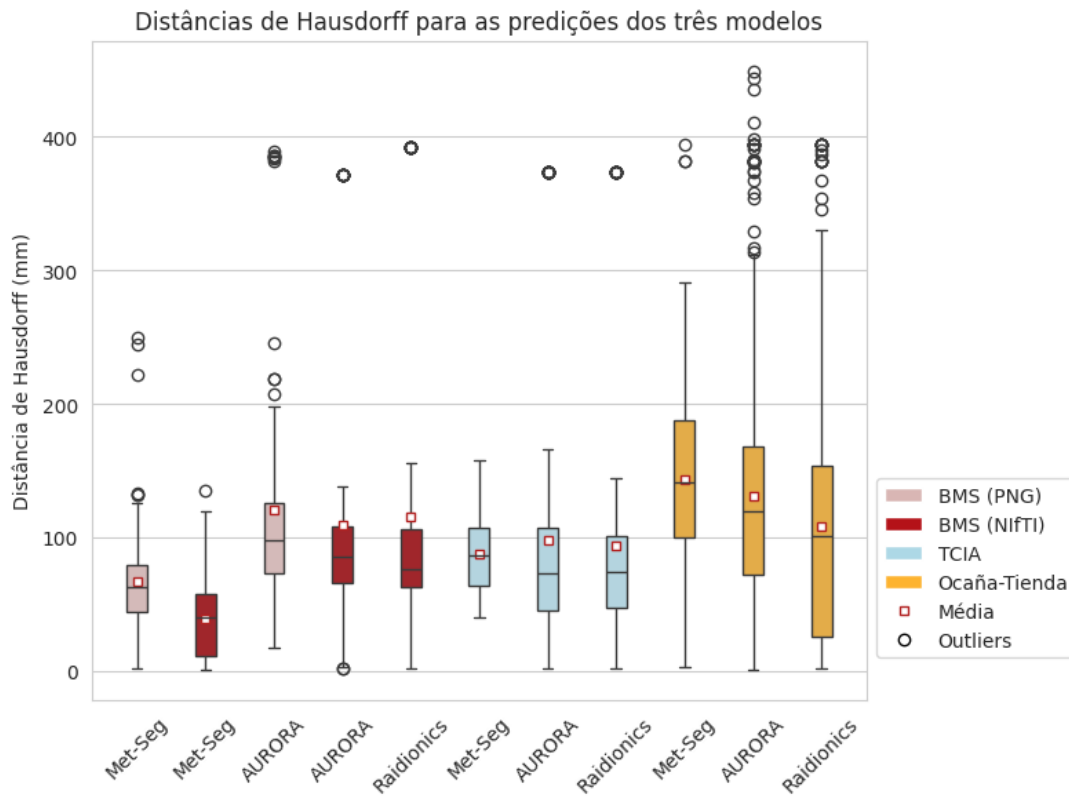


Figura 37 – Distâncias de Hausdorff (HD) calculadas para os três modelos: Met-Set (OT-TESEN *et al.*, 2023), AURORA (BUCHNER *et al.*, 2022) e Raidionics (BOUGET *et al.*, 2023) em cada um dos três bancos de dados: *BrainMetShare* (BMS) (GRØVIK *et al.*, 2020), TCIA (RAMAKRISHNAN *et al.*, 2024) e Ocaña-Tienda (OCAÑA-TIENDA *et al.*, 2023). Note que o banco BMS possui duas versões, a primeira em PNG e a segunda em NIFTI, e o modelo ideal apresenta valores de distância de Hausdorff próximas de 0 milímetros. Além disso, os *outliers* muito acima do restante do conjunto de previsões representam erros em que o modelo não conseguiu identificar a presença da metástase.

Fonte: Autoria própria.

Em relação aos coeficiente Dice, apresentados por modelo e por banco de dados na Figura 38, o modelo Met-Seg aplicado no banco de dados *BrainMetShare* foi o que apresentou o melhor resultado. No entanto, esse apresenta um viés de *dataset* em que foi treinado, e os resultados para os outros bancos de dados foram os piores quando comparado com os outros dois modelos, apresentando o pior resultado para o banco de dados do TCIA. Já o modelo Raidionics aplicado nos bancos de Ocaña-Tienda *et al.* (2023) e TCIA teve resultados consideráveis, e o modelo AURORA teve resultados um pouco menores do que o modelo Raidionics para esses bancos, porém ambos os modelos tiveram um resultado insatisfatório para o banco BMS, indicando que esses modelos não podem

ser utilizados para a segmentação de metástases pequenas. Além disso, assim como para o cálculo da métrica da distância de Hausdorff, tanto o modelo Met-Seg quanto o AURORA tiveram uma pior performance na métrica de coeficiente Dice ao serem aplicados no banco BMS composto por imagens PNG, o que indica a importância da aquisição dos dados no padrão de imagem em NIFTI.

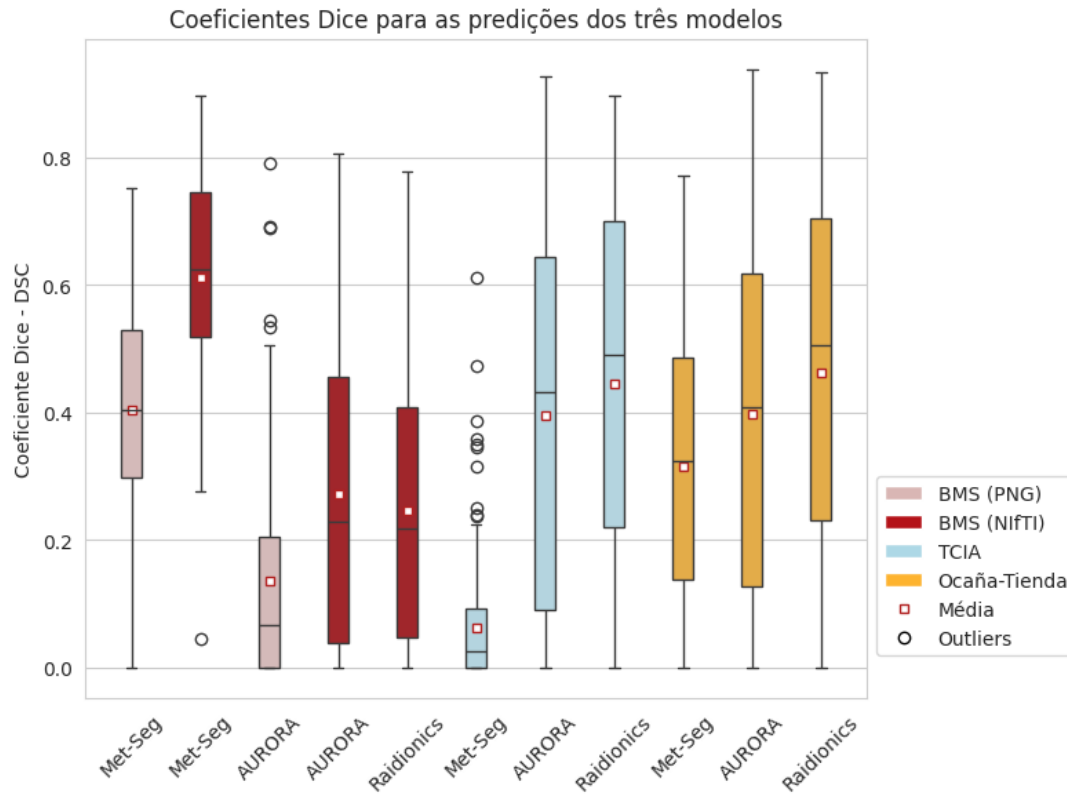


Figura 38 – Coeficientes Dice (DSC) calculados para os três modelos: Met-Set (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) em cada um dos três bancos de dados: BrainMetShare (GRØVIK et al., 2020), TCIA (RAMAKRISHNAN et al., 2024) e Ocaña-Tienda (OCAÑA-TIENDA et al., 2023). Note que o banco BMS possui duas versões, a primeira em PNG e a segunda em NIFTI, e o modelo ideal apresenta valores de Coeficientes Dice próximos a 1.

Fonte: Autoria própria.

Por fim, na Figura 39 está a comparação dos resultados obtidos para os valores de coeficiente Dice por *slice* médio, o coeficiente Dice para todo o volume e o valor encontrado para a métrica de interseção sobre a união (IoU), em que é possível perceber que não há muita diferença da informação que cada uma das métricas trás pois possuem um alto coeficiente de correlação de Pearson ($p > 0,9$) que indica uma correlação muito forte. Sendo assim, a análise feita para apenas uma delas é suficiente para averiguar a eficiência do modelo.

Em suma, os resultados obtidos para a predição dos três modelos de rede neural indicam que o banco de dados utilizado no treino é essencial para a boa performance do

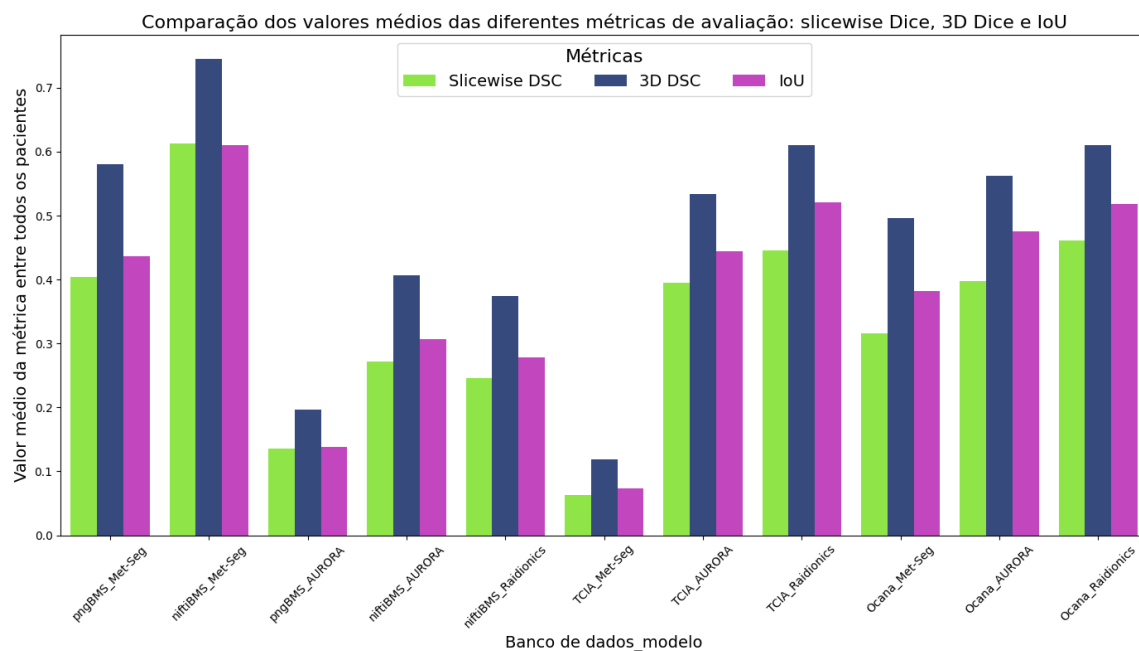


Figura 39 – Comparação das métricas de coeficientes Dice por *slice* médio (slicewise DSC), coeficiente Dice para todo o volume (3D DSC), e o valor da interseção sobre a união (IoU) calculados para os três modelos: Met-Set (OTTESEN et al., 2023), AURORA (BUCHNER et al., 2022) e Raidionics (BOUGET et al., 2023) em cada um dos três bancos de dados: *BrainMetShare* em PNG (old_BMS) e em NIfTI (BMS) (GRØVIK et al., 2020), TCIA (RAMAKRISHNAN et al., 2024) e Ocaña-Tienda (Ocana) (OCAÑA-TIENDA et al., 2023). Note que o banco BMS possui duas versões, a primeira em PNG e a segunda em NIfTI, e o modelo ideal apresenta valores de DSC e IoU próximos a 1.

Fonte: Autoria própria.

modelo, e que bancos de dados mais heterogêneos podem ser desafiadores para os modelos de redes neurais. Além disso, destaca-se a importância na padronização de banco de imagens obtidas no formato em NIfTI para melhores resultados. Não foram testadas imagens em formato DICOM e estudos de resultados para imagens nesse formato são necessários para análise comparativa. O modelo Raidionics foi o que apresentou os melhores resultados de forma geral, porém com um valor de DSC abaixo de 0,50, o que torna sua aplicação inviável para utilização na prática clínica da segmentação de metástases cerebrais. Os valores medianos para as distâncias de Hausdorff no entorno de 100 milímetros para todos os bancos de dados representa erros clínicos inaceitáveis, indicando que esses modelos precisam passar por refinamentos, para os quais se é feita necessária a criação de mais bancos de dados específicos para os casos de metástases cerebrais.

6.3 Proposição de um modelo para segmentação de metástases cerebrais

Nessa seção, estará descrito o desenvolvimento do modelo de segmentação de metástases cerebrais, que começa pela identificação de sequências de RM ideais para a identificação de *slices* com presença de metástases e segue com a determinação de melhores configurações de rede para a tarefa de segmentação.

6.3.1 Modelo para avaliação das melhores sequências de pulso para identificação de Metástases Cerebrais

Em uma primeira parte do estudo, foi feita uma comparação da performance de detecção de metástases cerebrais por meio de redes neurais artificiais em três ponderações de sequências de pulso em imagens de ressonância magnética de crânio: BRAVO, 3D CUBE e FLAIR. Para a sequência 3D CUBE, foram realizadas duas aquisições, uma sem aplicação de contraste e outra com aplicação de contraste

Sendo assim, foram realizados treinos de CNNs para identificação das melhores combinações de sequências de pulso e ponderações em RM através das métricas de acurácia, precisão, sensibilidade, especificidade, AuROCC e F1-Score. E as redes neurais utilizadas consistiam de uma rede neural convolucional simples, com arquitetura de três camadas de convolução com 8, 16 e 32 *kernels* de dimensão 3x3, seguidas por uma camada de *max pooling* com *kernels* 2x2 e uma camada de .5, .3 e .2 de *Dropout*, respectivamente. Depois dessas camadas, havia uma camada de *Flatten* para uma camada Densa de 64 neurônios com ativação 'relu' e uma camada de 0.4 de *Dropout*. Ao final, uma camada Densa de 1 neurônio com ativação sigmoide para classificação binária. Na Figura 40 encontra-se o diagrama com a arquitetura da rede utilizada.

Dessa maneira, foi possível realizar a variação dos hiperparâmetros a fim de se determinar a melhor ponderação para identificação das metástases. Em um primeiro momento, foi feita uma investigação com variação multiparamétrica para que algumas variações de hiperparâmetros fossem analisadas e ponderadas para a criação de um protocolo de treino. A partir dessa primeira investigação, a variação dos hiperparâmetros foram feitas realizando-se 6 passos para cada uma das sequências de pulso. Esses passos estão listados a seguir:

1. Iniciar o treino com *batch size* de 128, *learning rate* de 5E-4, durante 100 épocas, com *Dropouts* de, respectivamente, 0.5, 0.3, 0.2 e 0.4 e 64 neurônios na última camada;
2. Aumentar o *batch size* de 128 para 256;
3. Diminuir a *learning rate* de 5E-04 para 1E-04;

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 8)	80
max_pooling2d (MaxPooling2D)	(None, 128, 128, 8)	0
dropout (Dropout)	(None, 128, 128, 8)	0
conv2d_1 (Conv2D)	(None, 128, 128, 16)	1168
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 16)	0
dropout_1 (Dropout)	(None, 64, 64, 16)	0
conv2d_2 (Conv2D)	(None, 64, 64, 32)	4640
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 32)	0
dropout_2 (Dropout)	(None, 32, 32, 32)	0
flatten (Flatten)	(None, 32768)	0
dense (Dense)	(None, 64)	2097216
dropout_3 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65

Figura 40 – Arquitetura de rede utilizada para identificação
 Fonte: Autoria própria.

4. Diminuir todos os *Dropouts* pela metade;
5. Aumentar o tempo de treino para 300 épocas e o número de neurônios na última camada para 128;
6. Aumentar o *batch size* para 1024;

As matrizes de confusão obtidas para o conjunto de teste para cada um dos treinos do protocolo estão descritas na Figura 41:

1 - T1GRE	tp 264 130 fn 175 74	fn tn	1 - T1SE	tp 164 230 fn 128 121	fn tn	1 - T1SECE	tp 191 203 fn 117 132	fn tn	1 - T2FLAIR	tp 157 237 fn 170 79	fn tn
2 - T1GRE	tp 116 278 fn 81 168	fn tn	2 - T1SE	tp 172 222 fn 118 131	fn tn	2 - T1SECE	tp 146 248 fn 118 131	fn tn	2 - T2FLAIR	tp 173 221 fn 146 103	fn tn
3 - T1GRE	tp 89 305 fn 51 198	fn tn	3 - T1SE	tp 117 277 fn 140 109	fn tn	3 - T1SECE	tp 105 289 fn 102 147	fn tn	3 - T2FLAIR	tp 316 78 fn 238 11	fn tn
4 - T1GRE	tp 65 329 fn 37 212	fn tn	4 - T1SE	tp 136 258 fn 88 161	fn tn	4 - T1SECE	tp 100 294 fn 52 197	fn tn	4 - T2FLAIR	tp 251 143 fn 196 53	fn tn
5 - T1GRE	tp 288 106 fn 211 38	fn tn	5 - T1SE	tp 124 270 fn 79 170	fn tn	5 - T1SECE	tp 116 278 fn 68 181	fn tn	5 - T2FLAIR	tp 144 250 fn 59 190	fn tn
6 - T1GRE	tp 284 110 fn 211 38	fn tn	6 - T1SE	tp 191 203 fn 114 135	fn tn	6 - T1SECE	tp 128 266 fn 75 174	fn tn	6 - T2FLAIR	tp 301 93 fn 230 19	fn tn

Figura 41 – Matrizes de confusão para diferentes sequências de pulso onde: *tp* - *true positive*, *fp* - *false positive*, *tn* - *true negative*, *fn* - *false negative*
 Fonte: Autoria própria.

A partir desses resultados, foi possível categorizar as melhores sequências de pulso quanto à capacidade de identificação de metástases cerebrais, na seguinte ordem, da melhor para a pior: FLAIR, BRAVO e 3D CUBE com contraste.

Os resultados para cada um dos modelos treinados M em relação à precisão P, acurácia A, sensibilidade S, especificidade E e F1-Score F1 estão dispostos na Figura 18 a seguir:

Tabela 18 – Resultados dos treinos para determinação das melhores sequências de pulso no diagnóstico de metástases cerebrais pela rede neural, com medições de M - modelo, P - precisão, A - acurácia, S - sensibilidade, E - especificidade e F1 - *F1-Score*. Os melhores valores estão destacados em negrito.

Modalidade	M	P	A	S	E	F1
BRAVO	1	0,601	0,526	0,670	0,297	0,398
BRAVO	2	0,568	0,487	0,675	0,189	0,283
BRAVO	3	0,636	0,446	0,226	0,795	0,707
BRAVO	4	0,637	0,431	0,165	0,851	0,729
BRAVO	5	0,577	0,507	0,731	0,153	0,241
BRAVO	6	0,574	0,501	0,721	0,153	0,241
3D CUBE	1	0,562	0,443	0,416	0,486	0,521
3D CUBE	2	0,593	0,471	0,437	0,526	0,558
3D CUBE	3	0,455	0,351	0,297	0,438	0,446
3D CUBE	4	0,607	0,462	0,345	0,647	0,626
3D CUBE	5	0,611	0,457	0,315	0,683	0,645
3D CUBE	6	0,626	0,507	0,485	0,542	0,581
3D CUBE CC	1	0,620	0,502	0,485	0,530	0,572
3D CUBE CC	2	0,553	0,431	0,371	0,526	0,539
3D CUBE CC	3	0,507	0,392	0,266	0,590	0,546
3D CUBE CC	4	0,658	0,462	0,254	0,791	0,718
3D CUBE CC	5	0,630	0,462	0,294	0,727	0,675
3D CUBE CC	6	0,593	0,510	0,637	0,309	0,407
FLAIR	1	0,524	0,411	0,416	0,402	0,455
FLAIR	2	0,542	0,429	0,439	0,414	0,469
FLAIR	3	0,570	0,509	0,802	0,044	0,082
FLAIR	4	0,562	0,473	0,637	0,213	0,309
FLAIR	5	0,709	0,519	0,365	0,763	0,735
FLAIR	6	0,567	0,498	0,764	0,076	0,135

Fonte: Autoria própria.

Após o treino de 24 CNNs, foram obtidas todas as matrizes de confusão e mediu-se os valores da métrica F1-Score: 0,728 para BRAVO, 0,645 para 3D CUBE sem contraste, 0,718 para 3D CUBE e 0,735 para FLAIR. Quanto a análise isolada das métricas de precisão, acurácia, sensibilidade e especificidade, as melhores foram, respectivamente, 70,9% para FLAIR, 52,6% para BRAVO, 80,2% para FLAIR e 85,1% para BRAVO.

A partir desses resultados, foi possível identificar as sequências de pulso com mai-

ores contribuições para a identificação de metástases cerebrais através do treino de redes neurais: FLAIR e BRAVO. Os resultados para a sequência FLAIR indicaram precisão de 70,9%, acurácia de 51,9%, sensibilidade de 36,5% e especificidade de 76,3%. Além disso, não foi possível identificar uma diferença significativa na capacidade de detecção de metástases através das sequências 3D CUBE com e sem contraste.

Vale observar que os resultados encontrados nessa etapa da pesquisa reforçam os encontrados por (CHARRON et al., 2018), uma vez que as sequências com maior valor para a métrica DSC para a segmentação em ponderações 2D foram para FLAIR, seguida da T1. Além disso, no estudo de (CHARRON et al., 2018) a utilização de uma sequência 3DT1 se mostrou superior às duas sequências 2D.

A partir das imagens nas SP FLAIR, BRAVO e 3D CUBE CC, foi possível gerar uma imagem para o *input* de uma rede neural de arquitetura U-Net 2D, com atribuições para os canais RGB dos valores de intensidade de pixel para as três sequências com maior eficácia, FLAIR para o canal vermelho, BRAVO para o canal verde e 3D CUBE com contraste para o canal azul. Isso foi possível pois três os valores RGB em cada canal são os mesmos quando em escala de cinza, que é a escala da imagem original, como pode ser observado na Figura 42.

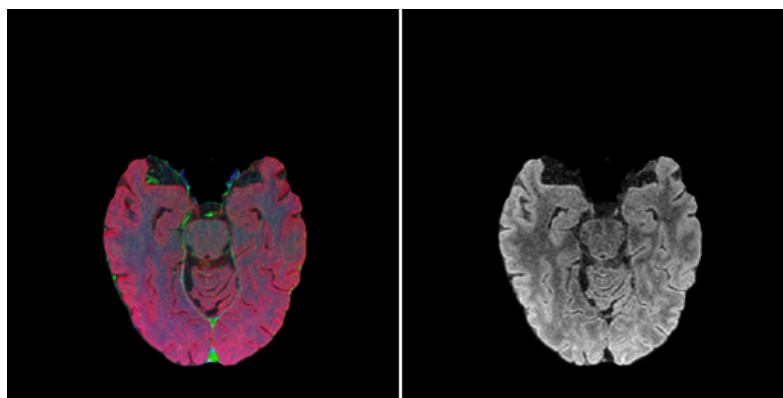
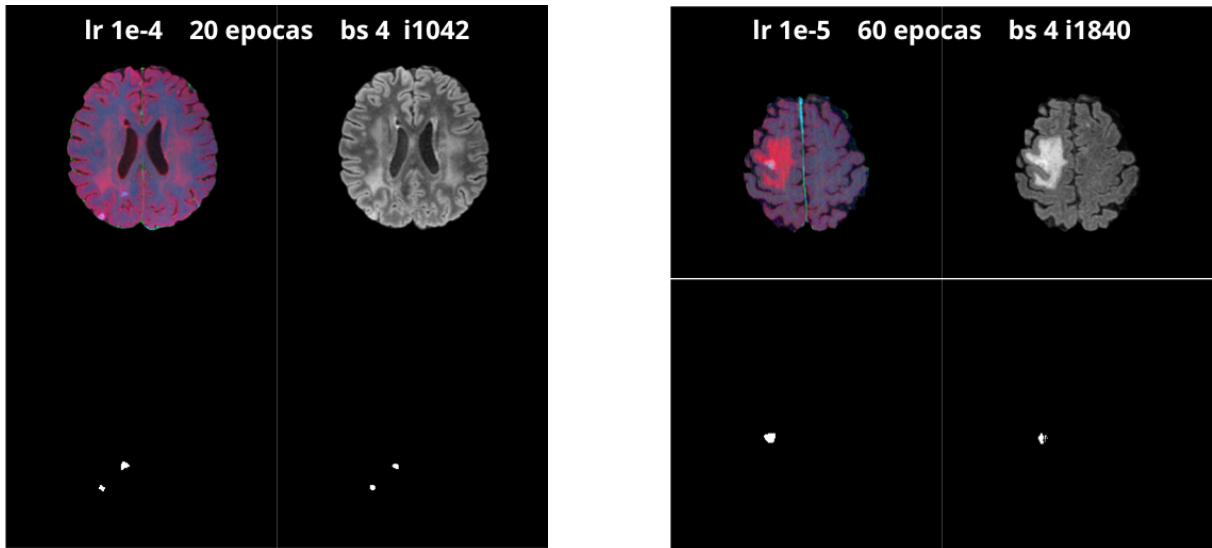


Figura 42 – Imagem com uma sequência em cada canal RGB (direita) e Imagem original FLAIR (esquerda)

Fonte: *BrainMetShare*

Dessa maneira, foi possível realizar o treino em uma rede neural do tipo 2D U-Net (RONNEBERGER et al., 2015) que mostrou bastante eficácia em tarefas de segmentação em visão computacional e, após o treino utilizando essa rede U-Net, foi possível realizar a segmentação de algumas metástases, como pode ser visto nas Figuras 43a e 43b. Nota-se que essa segmentação mostrou-se eficaz para metástases de tamanhos pequenos, podendo auxiliar a identificação realizada pelos radiologistas.



(a) Treino em 20 épocas com *learning rate* de $1e-4$ e *batch size* igual a 4

(b) Treino em 60 épocas com *learning rate* de $1e-5$ e *batch size* igual a 4

Figura 43 – Imagens mostrando a eficiência da U-Net em realizar segmentações. Imagem com os três canais de cores para cada modalidade de aquisição (acima à esquerda), imagem da aquisição FLAIR (acima à direita), máscara de segmentação feita pelos neurorradiologistas (abaixo à esquerda) e máscara de segmentação realizada pela U-Net (abaixo à direita) para duas imagens (a) com 2 metástases e em (b) com 1 metástase pequena.

Fonte: Autoria própria.

6.3.2 Avaliação das funções de perda do modelo de segmentação

Dessa forma, foi utilizado um modelo *dummy* formado por pacientes do banco de dados *BrainMetShare* sem a aplicação de técnicas de *Data Augmentation* e *Transfer Learning* com intuito de avaliar a influência de diferentes funções perdas variando-se seus parâmetros. Para isso, foi considerado que, caso o modelo *dummy* tenha como resultado final um coeficiente Dice para o conjunto de teste desprezível, isto é, abaixo de 0,01, esse resultado não seria melhorado suficientemente para aplicação com o uso de *Data Augmentation*. Essa estratégia permite avaliar os casos em que o coeficiente Dice para o modelo *dummy* sejam maiores, o que poupa tempo de máquina e permite avaliações com um maior número de variações de hiperparâmetros envolvidos no treino do modelo. Sendo assim, o detalhamento de cada função perda, bem como o resultados da variação dos parâmetros de cada uma delas estão apresentados a seguir.

6.3.2.1 Binary Cross Entropy

A função perda BCE foi aplicada em um treino do modelo *dummy* para avaliar sua influência nos resultados para o coeficiente Dice. No entanto, o valor de DSC obtido para o conjunto de teste em um treino com *learning rate* de 0,001 foi de 0,0005634. Outros treinos

realizados com variação de hiperparâmetros não resultaram em valores significativos, o que indicou para um descarte do uso dessa função perda ou para a necessidade da combinação dessa função com outras funções perda.

6.3.2.2 Binary Balanced Focal Cross Entropy

A função perda BBFCE, que é uma versão da BBCE com pesos focais, também foi testada no modelo *dummy*, e envolveu a variação dos parâmetros α de balanceamento e γ de peso focal. Dessa forma, um total de 34 combinações desses fatores, com $0, 1 \leq \alpha < 1$ e $2 \leq \gamma \leq 8$, o valor para o coeficiente Dice obtido não foi superior a 0,075. Sendo assim, o uso dessa função perda foi descartado.

6.3.2.3 Tversky

A função Tversky foi utilizada individualmente para o treino do modelo *dummy*, e inicialmente os parâmetros $\frac{\eta}{2}$ e $\frac{\psi}{2}$ da função foram variados para números inteiros indo de 1 a 10. Note que os valores estão divididos por dois, já que a implementação em relação à Equação 18 se dá com uma diferença de um fator de dois. Para comparação, essa variação foi realizada com os valores de *learning rate* iguais a 0,01 e 0,001. O mapa de calor para essa análise está apresentado na Figura 44, em que percebe-se que a *learning rate* de 0,001 apresentou melhores resultados do que a *learning rate* de 0,01. Além disso, foram realizados os treinos para *learning rate* de 0,0001 e 0,005, em que os resultados obtidos para o DSC ficaram abaixo de 2 ordens de grandeza dos resultados obtidos para a *learning rate* de 0,001.

Dessa forma, foi avaliado uma segunda faixa de valores para os coeficientes $\frac{\eta}{2}$ e $\frac{\psi}{2}$ com valores entre 0 e 4, que está ilustrado na Figura 45. Nesse mapa de calor, os quadrados tracejados em vermelho e amarelo representam as funções perda Dice e IoU, sendo possível observar que não há ganhos significativos em utilizá-las em comparação com outras variações da função de Tversky. No entanto, vale ressaltar que os resultados obtidos apenas com o uso da função de Tversky foram todos abaixo do ideal, com valores de DSC abaixo de 0,13 e, ademais, para valores acima de 1,5 para o coeficiente $\frac{\eta}{2}$, a variação dos hiperparâmetros não resultou em uma variação contínua de resultados, o que pode indicar que tais resultados são específicos do treino em questão, são suscetíveis a pequenas variações na rede, como inicializadores aleatórios, e não são generalizáveis.

6.3.2.4 BCE + Dice + Boundary

O uso da função perda apresentada por Kervadec *et al.* (2021) para a diminuição das distâncias entre as superfícies das segmentações foi utilizada em combinação com as funções BCE e Dice, o que melhorou os resultados em cerca de 5% quando comparado ao uso dessas duas apresentadas na subseção anterior. Além disso, esses resultados foram

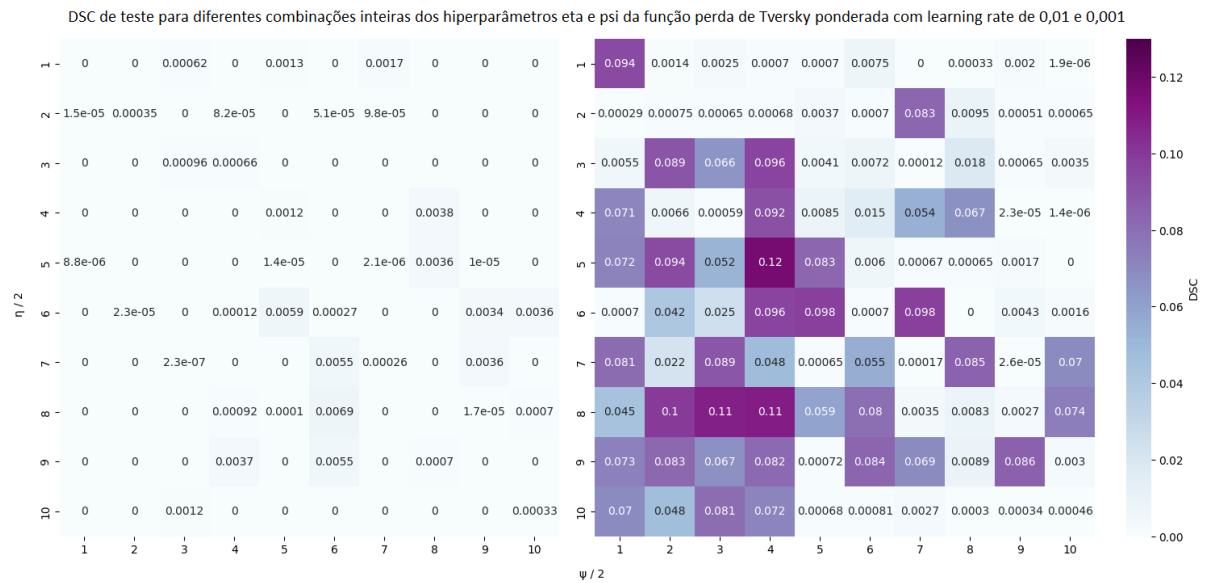


Figura 44 – Mapas de calor dos coeficientes Dice para a variação dos coeficientes $\frac{\eta}{2}$ e $\frac{\psi}{2}$ em números inteiros de 1 a 10 da função perda Tversky. O mapa da esquerda foi feito para a *learning rate* de 0,01 e o mapa da direita para a *learning rate* de 0,001.

Fonte: Autoria própria.

comparados com o uso da função de Tversky utilizando os parâmetros que a deixam equivalente à função Dice, e um gráfico com os valores para diferentes pesos de balanceamento estão apresentados na Figura 46, em que é possível observar que pesos pequenos para o balanceamento tendem a trazer piores resultados para os modelos treinados sem o uso da função *Boundary* e, com o seu uso, o efeito desse parâmetro de peso de balanceamento parece não afetar os resultados, que se mantêm entre 0,33 e 0,42. Sendo assim, o acréscimo da função *Boundary Loss* proposto por Kervadec *et al.* (2021) deve ser levado em consideração para a construção de um modelo de segmentação.

6.3.2.5 BCE + Dice + HD

A combinação das funções perda BCE, Dice e *Hausdorff Distance* foi testada, porém em nenhum dos treinos para o modelo *dummy* o valor de DSC do conjunto de teste foi superior a 0,001. Por esse motivo, o seu uso não foi continuado para os casos com *Data Augmentation*.

6.3.2.6 BCE + Tversky ponderada

Para avaliar a combinação da função BCE com a Tversky ponderada, foi utilizado o modelo *dummy* para uma combinação de hiperparâmetros η e ψ indo de 1 a 5 para diferentes ponderações para o balanceamento de classes.

Na Figura 47, encontra-se o mapa de calor para os coeficientes Dice do conjunto

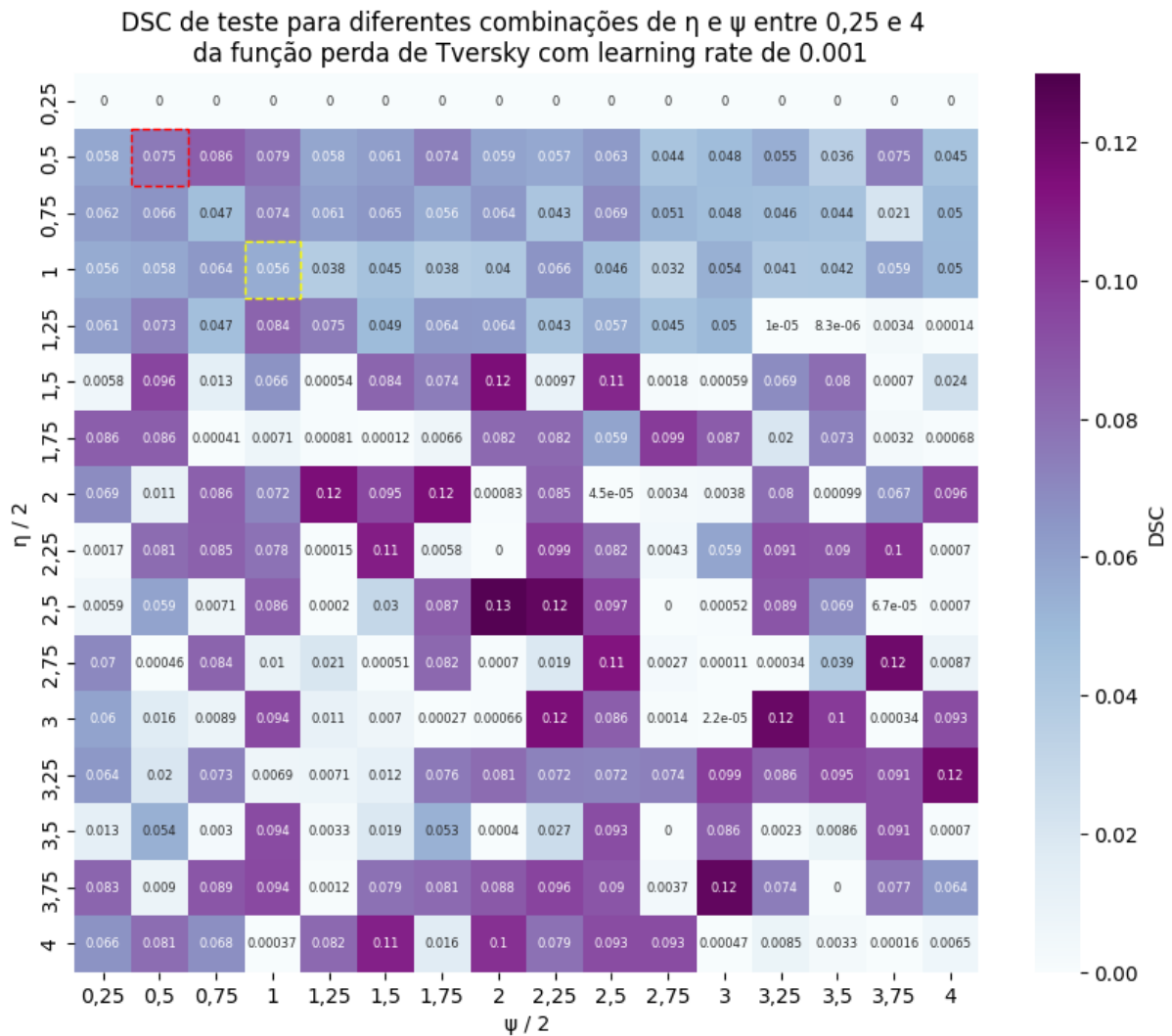


Figura 45 – Mapas de calor dos coeficientes Dice para a variação dos coeficientes η e ψ entre 0,25 e 4 da função perda Tversky ponderada para a *learning rate* de 0,001. Os quadrados destacados em vermelho e amarelo representam, respectivamente, as funções perda Dice e IoU

Fonte: Autoria própria.

de teste do modelo *dummy*. Nota-se que a ponderação para o balanço de classes tem um impacto positivo para valores menores do que 0,01 e alcança um platô a partir de 10^{-6} , o que pode estar em desacordo com o informado por Salehi *et al.* sobre a desnecessidade de realizar o balanceamento de classe quando utilizada a função Tversky (SALEHI *et al.*, 2017), os autores, no entanto, não informam quais métodos de balanceamento estavam se referindo. Além disso, e de acordo com os resultados obtidos por Salehi *et al.* (2017), fica evidente que o valor de $\psi = 1$ otimiza os resultados para o coeficiente Dice, o que é perceptível nas colunas mais claras que precedem o bloco de colunas para um mesmo valor de η . Isso indica que o coeficiente ψ que multiplica os falsos negativos não precisa ser alterado na equação de Tversky original, em que $\psi = 1$.

Um outro aspecto que pode ser observado são que os blocos de 5 colunas de um

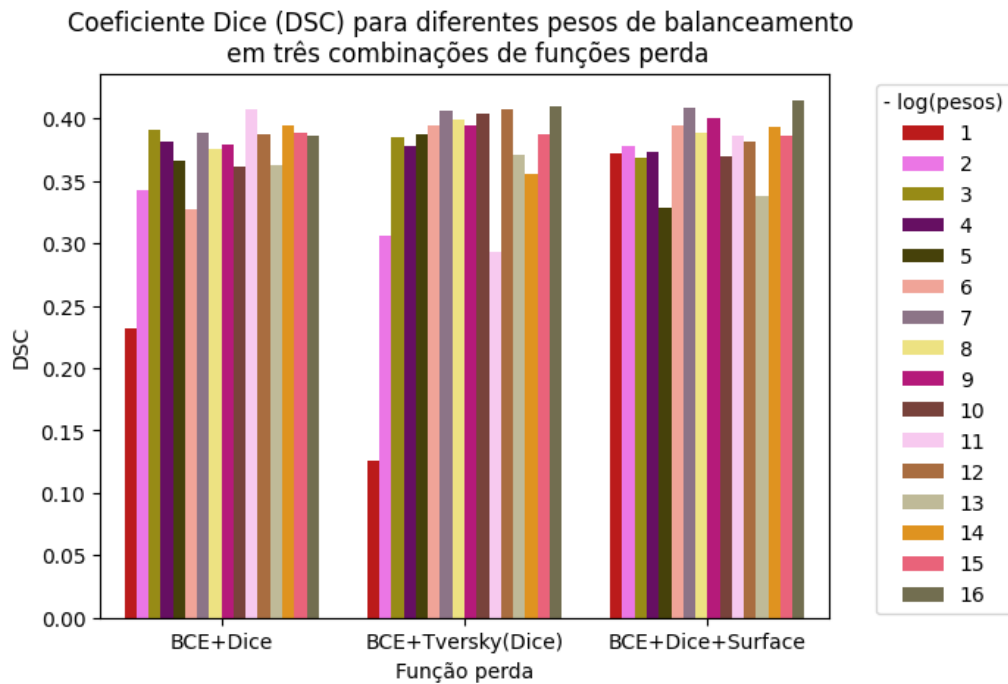


Figura 46 – Comparação dos valores do coeficiente Dice para o conjunto de teste com o uso de três funções perdas e com variação dos pesos para o balanceamento das classes, cujos valores vão de de 0,1 até 10^{-16} e estão apresentados na legenda pelo negativo do logaritmo.

Fonte: Autoria própria.

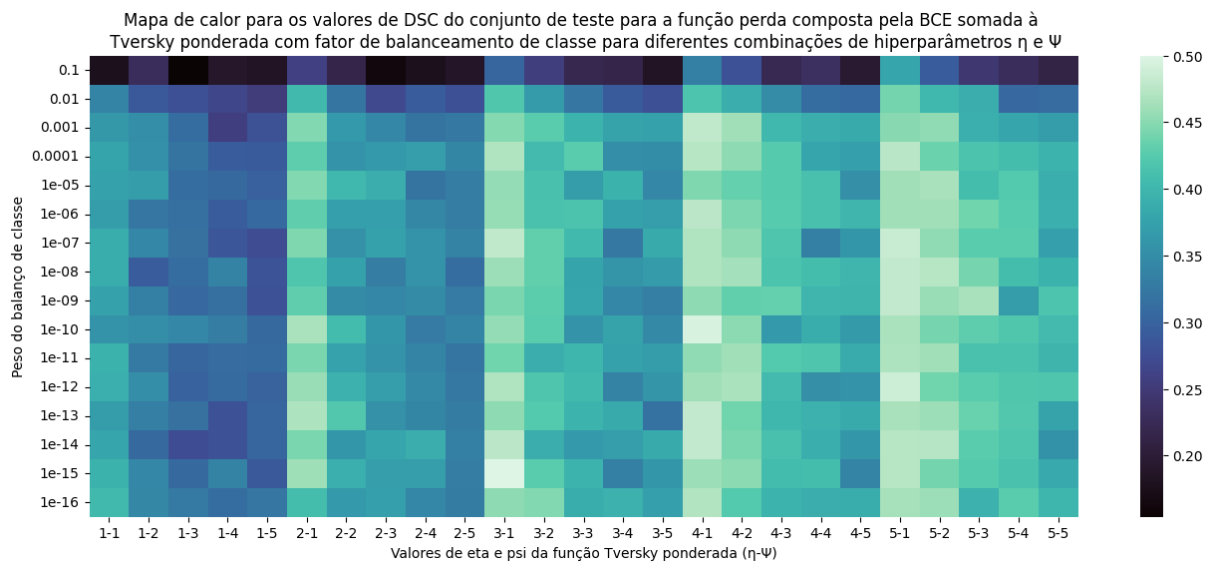


Figura 47 – Mapa de calor dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.

Fonte: Autoria própria.

mesmo valor η se tornam mais claros conforme o valor desse coeficiente aumenta, indicando que os resultados tendem a melhorar para maiores valores de η . Por esses motivos, a fim de verificar a tendência e investigar um valor ótimo para os dois coeficientes, foram realizadas

variações de $\eta = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ e de $\psi = 0.01, 0.25, 0.5, 0.75, 1, 2$, para um número reduzido de pesos para o balanço de classes, uma vez que após atingir o platô não houveram muitas variações nos treinos anteriores. Os 5 valores para os pesos do balanceamento das classes foram 10^{-1} , 10^{-6} , 10^{-9} , 10^{-12} e 10^{-16} e o mapa de calor obtido para esse conjunto de hiperparâmetros está disponível na Figura 48.

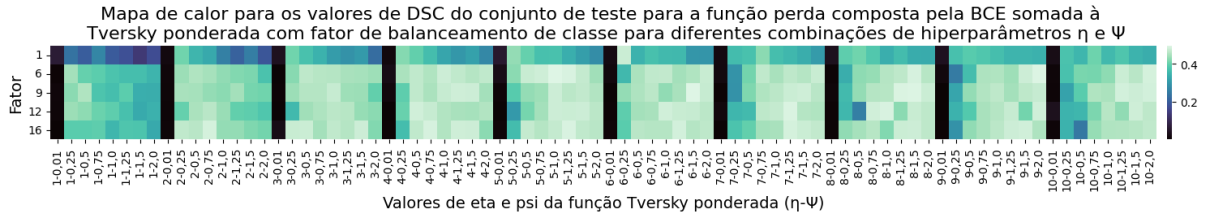


Figura 48 – Mapa de calor dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.

Fonte: Autoria própria.

Através desse gráfico, foi possível verificar que um valor de $\psi = 0,01$ resultou em perdas significativas para a performance do modelo (colunas escuras), porém não foi possível encontrar alguma tendência de melhoria do resultado da segmentação para um grupo de valores de η com essa faixa de valores ψ . Tampouco foi possível observar alguma variação significativa para os diferentes pesos do balanceamento de classes, o que fica ainda mais evidente no gráfico da Figura 49, onde é possível observar a estabilidade dos resultados para diferentes combinações dos hiperparâmetros η e ψ a partir do valor de $\psi = 1$ nos casos em que $\eta \geq 4$.

Gráfico de colunas empilhadas para os valores de DSC do conjunto de teste para a função perda composta pela BCE somada à Tversky ponderada com fator de balanceamento de classe para diferentes combinações de hiperparâmetros η e ψ

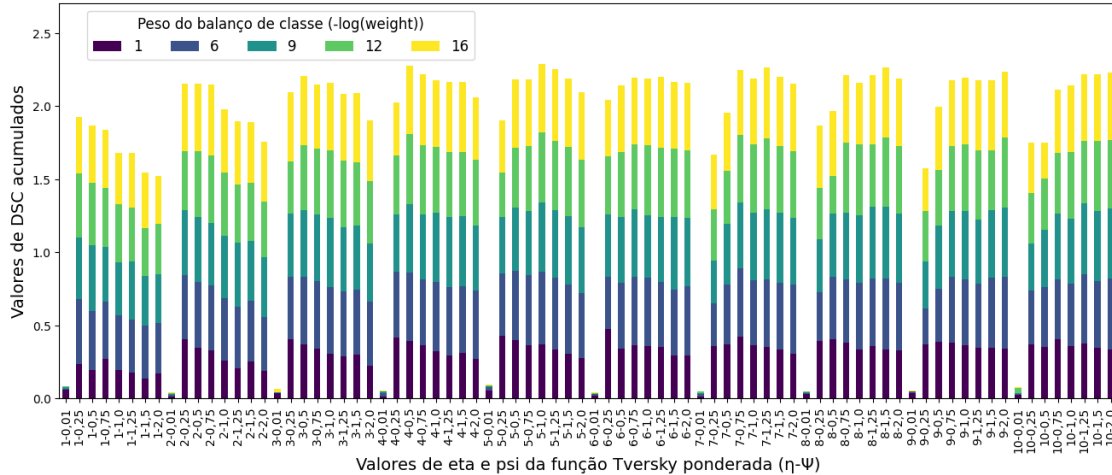


Figura 49 – Gráfico de barras empilhadas dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.

Fonte: Autoria própria.

6.3.2.7 BCE + Tversky ponderada + Boundary

O uso das funções perda BCE somada à função *Boundary Loss* (KERVADEC et al., 2021) e Tversky ponderada foi a combinação de funções que se apresentou mais adequada para a tarefa de segmentação de metástases cerebrais. Sendo assim, foi realizado o treino através do uso de *Data Augmentation* e os valores obtidos de DSC para o conjunto de teste podem ser encontrados nas Figuras 50 e 51, onde é possível identificar o melhor modelo como sendo aquele composto por um fator de peso igual a 0,1 com fatores $\eta = 9$ e $\psi = 2$.

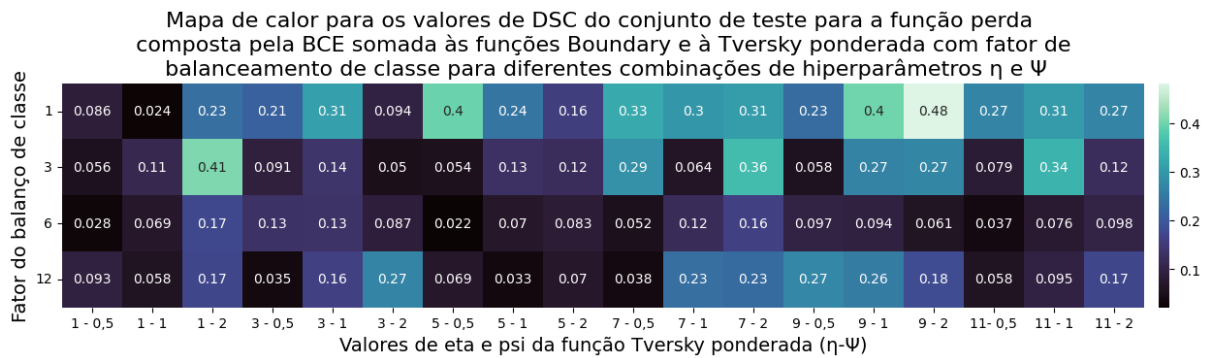


Figura 50 – Gráfico de barras empilhadas dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE e pela Tversky ponderada com diferentes pesos para o balanceamento de classe.

Fonte: Autoria própria.

Gráfico de colunas empilhadas para os valores de DSC do conjunto de teste para a função perda composta pela BCE somada à Tversky ponderada com fator de balanceamento de classe para diferentes combinações de hiperparâmetros η e Ψ

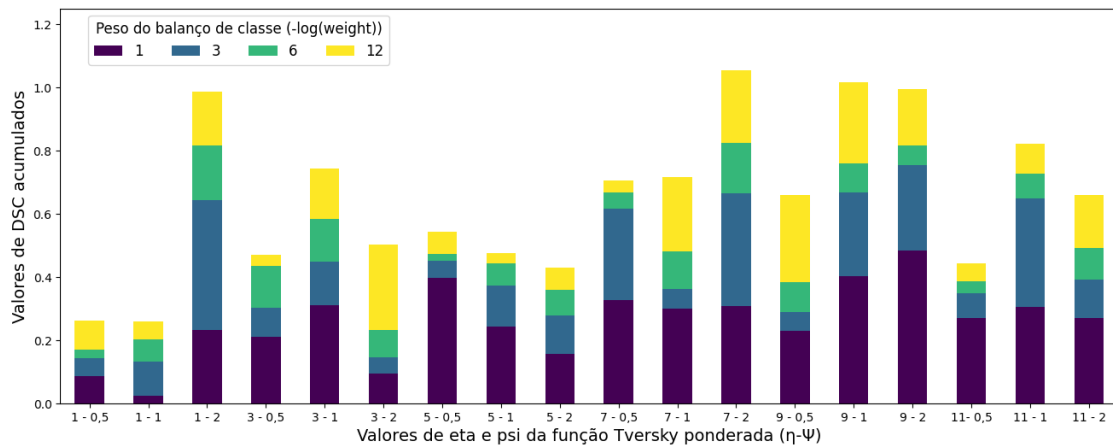


Figura 51 – Gráfico de barras empilhadas dos coeficientes Dice para a variação dos coeficientes η e ψ da função perda composta pela BCE, Tversky e Boundary ponderada com diferentes pesos para o balanceamento de classe.

Fonte: Autoria própria.

Além disso, nota-se que os resultados para o valor de $\psi = 2$ foram superiores aos resultados para valores de $\psi < 2$, o que está em desacordo com os testes no modelo *dummy*. Sendo assim, análise quantitativa do uso dessas funções de perda combinadas permitiu

afirmar que são adequadas para o uso em um treino de redes neurais convolucionais. No entanto, ainda não é muito evidente a relação entre os fatores de balanceamento da equação e o resultado final da segmentação. Por esse motivo, uma análise qualitativa para avaliar os resultados da segmentação foi realizada utilizando um *slice* que não serviu de *input* para o modelo e que está apresentado junto de sua máscara de segmentação na Figura 52. A seguir, serão apresentados e comentados os resultados do modelo para diferentes combinações de parâmetros da função de perda BCE com Tversky e *Boundary*.

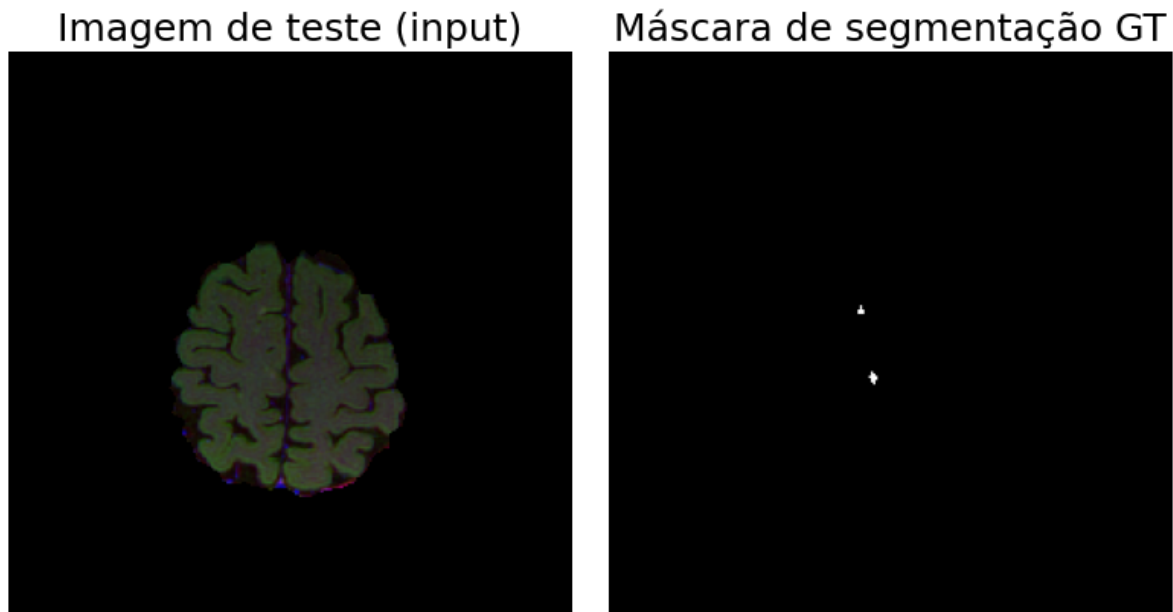


Figura 52 – Imagem de teste e sua máscara de segmentação. Esse exemplo foi escolhido por representar uma metástase múltipla de difícil identificação devido ao tamanho.

Fonte: Autoria própria.

Na Figura 53, temos o uso do parâmetro de balanceamento de classe 0,1 e valores de η indo de 1 até 15, com passo de 2, e os parâmetros ψ assumindo valores 0,5, 1 e 2, uma vez que sua contribuição não resultou em melhorias no modelo *dummy* para valores muito acima de 1. O melhor (swDSC = 0,483) e o pior (swDSC = 0,024) modelo estão destacados com uma seta dentro de um círculo verde e vermelho, respectivamente. A escolha do melhor e pior modelo foi realizada a partir dos valores de swDSC obtidas para todo o conjunto de teste, e não apenas para o *slice* em questão. Sendo assim, o aumento do valor de η aparenta estar de acordo com os testes feitos no modelo *dummy* e resultam em melhorias na segmentação. Por outro lado, o fator ψ apresentou melhores resultados para seu valor sendo igual a 2, que não está de acordo com os testes do modelo *dummy*, cujo resultado não variava muito a partir de ψ sendo próximo a 1. Além disso, é possível verificar que a identificação dos volumes tumorais, isto é, a diminuição do número de falsos negativos por *slice*, é influenciado pelo valor de η maior que 5, e também que há uma diminuição do ruído (falsos positivos) com o aumento de ψ , o que é esperado a partir

da Equação 19 de Tversky.

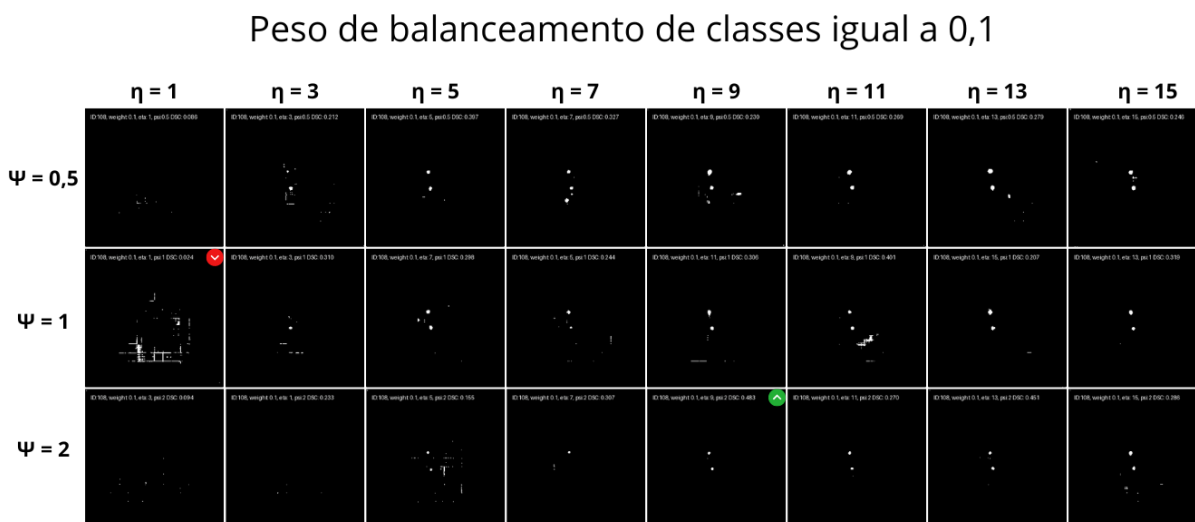


Figura 53 – Resultados da segmentação do modelo com uso de *Data Augmentation* para diferentes composições de η e ψ utilizando o peso de balanceamento de classes igual a 0,1 na função de perda BCE combinada à Tversky ponderada e à *Boundary*. O melhor e o pior modelo estão destacados no canto superior direito com uma seta dentro de um círculo verde e vermelho, respectivamente.

Fonte: Autoria própria.

Ao diminuir os valores do peso de balanceamento para 0,001, os resultados para a segmentação não apresentaram melhoria do modelo, que teve um valor máximo de swDSC de 0,411, e uma análise qualitativa pode ser feita a partir da Figura 54.

A partir dessa análise, há indícios de uma mesma tendência para a diminuição de falsos negativos a partir de η maior do que 5 e diminuição de falsos positivos com o aumento dos valores de ψ . Para esse conjunto em específico, o melhor modelo possuía $\eta = 1$, porém não foi capaz de identificar a metástase da imagem de teste, visto que o resultado da segmentação para esse *slice* é constituído de uma imagem quase que inteiramente preta. Isso ocorre pois a definição do melhor modelo é feita para todos os *slices* de todos os exames do conjunto de teste, e não apenas para a imagem de teste apresentada para análise qualitativa. Por fim, os resultados para os pesos de 10^{-6} e 10^{-12} não foram superiores aos resultados para os pesos 0,1 e 0,001, e a análise qualitativa para esses dois pesos não foi realizada.

Dessa maneira, o modelo final foi escolhido com o conjunto de hiperparâmetros listados a seguir e os resultados obtidos para os três bancos de dados estão apresentados na próxima seção.

- **Learning Rate:** 0,001 com decrescimento de fator multiplicativo igual 0,2 a cada 6 épocas sem diminuição da função de perda utilizando o otimizador Adam.

Peso de balanceamento de classes igual a 0,001

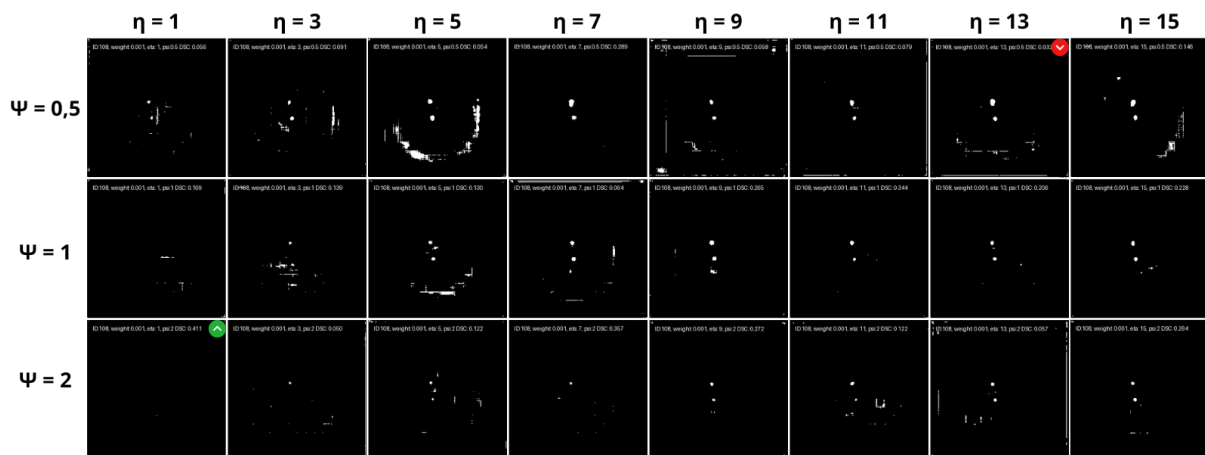


Figura 54 – Resultados da segmentação do modelo com uso de *Data Augmentation* para diferentes composições de η e ψ utilizando o peso de balanceamento de classes igual a 0,001 na função de perda BCE combinada à Tversky ponderada e à *Boundary*. O melhor e o pior modelo estão destacados no canto superior direito com uma seta dentro de um círculo verde e vermelho, respectivamente.

Fonte: Autoria própria.

- **Número de épocas:** Máximo de 200 ou até o modelo ficar 20 épocas consecutivas sem diminuição da função de perda.
- **Função de perda:** Soma das funções *Binary Cross Entropy*, Tversky ponderada ($\eta = 9$ e $\psi = 2$) e *Boundary*.
- **Peso do balanceamento de classes:** 0,2
- **Batch Size:** 16
- **Data Augmentation:** 50% de chance de aplicar pelo menos uma das seguintes transformações (128 épocas com 465 *steps* por época, com *batch size* de 16 totalizando uma média de 476160 imagens por treino, representando um aumento de cerca de 51 vezes da quantidade original de imagens $4632 \text{ slices} \times 2 \text{ classes} = 9264$):
 1. Rotação de até 90° com preenchimento de bordas em preto.
 2. Deslocamento de até 10% nos eixos horizontal e vertical com preenchimento de bordas em preto.
 3. Inversão Horizontal.

As curvas de aprendizado para o modelo com melhor performance para o teste externo estão apresentadas na Figura 55. Neste treino, as imagens utilizadas no *input* foram apenas imagens T1 com contraste do banco de dados BMS em formato PNG.

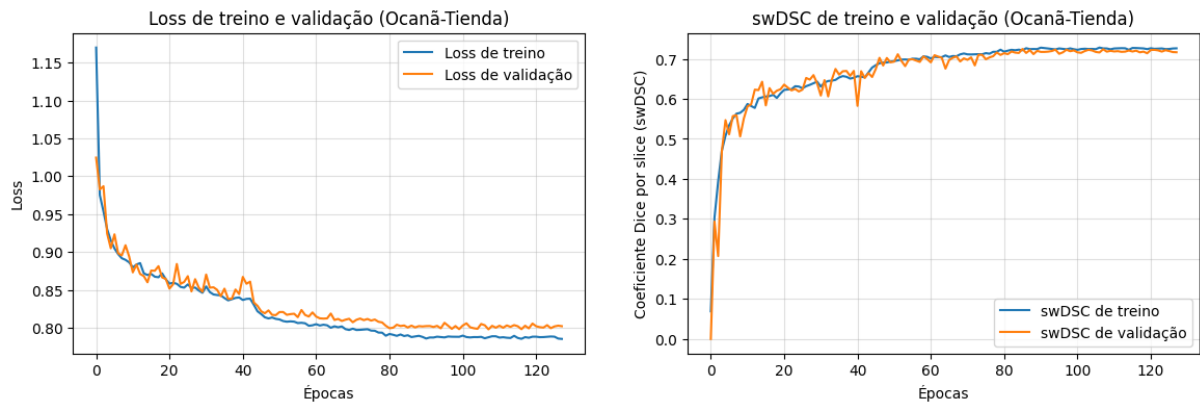


Figura 55 – Curvas de aprendizado para o modelo final em treino realizado utilizando imagens apenas em T1 com contraste do banco de dados *BrainMetShare* em formato PNG.

Fonte: Autoria própria.

6.3.3 Comparação dos resultados do modelo desenvolvido com os modelos de literatura

Nas Tabelas 19 e 20 encontram-se os resultados do coeficiente Dice volumétrico do nosso modelo comparados com os modelos da literatura para todos os conjuntos de imagens (Tabela 19) e para apenas as imagens ponderadas em T1 com contraste (Tabela 20). Nota-se que apesar do nosso modelo não apresentar resultados médios superiores quando comparado aos outros modelos, ainda se encontra dentro do desvio padrão dos outros resultados. Além disso, nosso modelo possui limitações relativas ao uso de apenas imagens do banco de dados BMS em formato PNG, sendo possível aprimorar os resultados com o uso das imagens em formato NIfTI e realizando o *shuffling* dos dados, que pode ser desenvolvido em trabalhos futuros.

Tabela 19 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelo nosso modelo e comparado aos modelos Met-Seg, AURORA e Raidionics nos bancos de dados *BrainMetShare* em formato original NIfTI, TCIA e Ocaña-Tienda *et al.* (2023). Os melhores resultados por banco de dados estão destacados em negrito.

Banco de dados	Nosso Modelo	Met-Seg	AURORA	Raidionics
BMS NIfTI	0,25 ± 0,21	0,75 ± 0,13	0,41 ± 0,31	0,37 ± 0,31
TCIA	0,19 ± 0,22	0,12 ± 0,17	0,53 ± 0,36	0,61 ± 0,35
Ocaña-Tienda	0,31 ± 0,25	0,50 ± 0,30	0,56 ± 0,36	0,61 ± 0,35

Fonte: Autoria própria.

Tabela 20 – Média e desvio padrão dos valores de coeficiente Dice volumétrico encontrados pelo modelo desenvolvido e os modelos da literatura Met-Seg, AURORA e Raidionics nas imagens em ponderação T1 com contraste dos bancos de dados *BrainMetShare* em formato original NIfTI, TCIA e Ocaña-Tienda *et al.* (2023). Os melhores resultados por banco de dados estão destacados em negrito.

Banco de dados	Nosso Modelo	Met-Seg	AURORA	Raidionics
BMS NIfTI	0,45 ± 0,24	0,48 ± 0,21	0,40 ± 0,33	0,37 ± 0,31
TCIA	0,20 ± 0,19	0,34 ± 0,27	0,68 ± 0,33	0,61 ± 0,35
Ocaña-Tienda	0,44 ± 0,29	0,50 ± 0,30	0,56 ± 0,36	0,61 ± 0,35

Fonte: Autoria própria.

7 Conclusão

Um modelo baseado em redes neurais para a segmentação de metástases cerebrais possui alguns desafios associados à quantidade de bancos de dados disponíveis publicamente, que é escassa, e ao tamanho dos volumes tumorais, que são relativamente pequenos, na ordem de 1 a 5 mLs. Sendo assim, a partir da avaliação dos modelos da literatura que realizam essa tarefa, foi possível observar que os modelos treinados com bancos de dados que possuem imagens de metástases de menor volume, terá um resultado melhor para a tarefa de segmentação de metástases menores, porém esse resultado do modelo tende a piorar quando ele é aplicado em imagens contendo metástases de maior volume. A situação inversa também é válida, e os modelos treinados com imagens de metástases maiores não possuem uma boa performance quando aplicados em imagens de metástases de menor tamanho. Um outro fator que se mostrou relevante foi a quantidade de volumes tumorais por exame, já que uma grande quantidade de volumes tumorais dificultou a tarefa de segmentação para dois dos modelos que não foram treinados em bancos de dados com um grande número de tumores por exame. Além disso, o tipo do tumor primário da metástase pode influenciar a performance de segmentação do modelo, que possui resultados diferentes para os casos de câncer de mama e câncer de pulmão de células pequenas e não pequenas. No entanto, essa análise foi realizada apenas com um banco de dados e mais pesquisas são necessárias para reforçar essa afirmação.

Logo, é necessária a diversidade dos bancos de dados utilizados para o treinamento do modelo para se obter resultados mais generalizáveis. Atualmente, os modelos Met-Seg, AURORA e Raidionics apresentam resultados consideráveis para os valores de swDSC, sendo o modelo Raidionics o mais estável entre os bancos de dados do TCIA e Ocaña-Tienda *et al.* (2023) com o valor para o 3D-DSC médio de $0,53 \pm 0,29$ nos três bancos. O modelo AURORA é o que apresenta o melhor resultado médio de segmentação 3D-DSC = $0,55 \pm 0,29$. Apesar desses resultados, nenhum deles é suficiente para ser aplicado na prática clínica, sendo necessário o desenvolvimento de modelos mais robustos que possuem um valor para o coeficiente Dice mais próximos de 1.

Um outro ponto que merece destaque é o fato de que os modelos Met-Seg, AURORA e o desenvolvido por nós apresentaram melhoria ao se utilizar apenas a sequência ponderada em T1 com adição de contraste em comparação ao uso de todas as sequências disponíveis para cada paciente, indicando que um protocolo de aquisição de imagem apenas dessa sequência pode ser suficiente para a detecção de metástases cerebrais.

Ao se desenvolver um modelo de redes neurais para a tarefa de segmentação, é necessária atenção na definição da função de perda utilizada, e diferentes combinações dessas funções podem melhorar ou piorar o resultado do modelo. Neste trabalho foi pos-

sível concluir que o uso da função de perda *Boundary Loss* somada às funções BCE e Tversky resultou em uma maior homogeneidade dos resultados de swDSC para diferentes pesos de balanceamento de classes, podendo ser considerada sua adição à tradicional função perda constituída pela soma da BCE e Tversky.

Em relação à variação dos fatores η e ψ da função de perda de Tversky, concluí-se que a tendência de platô para os resultados obtidos sem o *Data Augmentation* é perdido, e que o teste realizado com o modelo *dummy* não é suficiente para replicar os mesmos resultados em casos de *Data Augmentation*. Além disso, é notável a diminuição da quantidade de falsos negativos e falsos positivos pela variação dos parâmetros η e ψ , respectivamente, da função perda de Tversky. Cabe ressaltar que esses testes foram realizados utilizando-se apenas as imagens do banco de dados BMS em formato PNG, que é uma das limitações do estudo. Outras limitações da pesquisa envolvem o uso de uma rede U-Net bidimensional, quando o estado da arte é a rede 3D, que demanda maior poder computacional. Sendo assim, o uso de outros bancos de dados em formato NIfTI farão parte dos próximos passos da pesquisa.

Por fim, a análise de diferentes combinações desses parâmetros da função de Tversky deve ser realizada para se desenvolverem modelos mais otimizados para a tarefa de segmentação de metástases cerebrais para a prática clínica, uma vez que os modelos disponíveis não se mostraram robustos o suficiente para manterem, nos testes externos, os resultados obtidos em testes internos, principalmente para os casos em que os bancos de dados de treino e de teste possuem características divergentes. Com isso em vista, estratégias de *shuffling* de diferentes bancos de dados podem ser utilizadas para tornar o modelo treinado mais robusto a novos casos obtidos prospectivamente.

Referências

- ACHROL, A. S.; RENNERT, R. C.; ANDERS, C.; SOFFIETTI, R.; AHLUWALIA, M. S.; NAYAK, L.; PETERS, S.; ARVOLD, N. D.; HARSH, G. R.; STEEG, P. S.; CHANG, S. D. Brain metastases. **Nature reviews. Disease primers**, v. 5, n. 1, p. 5, January 2019. ISSN 2056-676X. Disponível em: <<https://doi.org/10.1038/s41572-018-0055-y>>.
- AKERET, K.; NIFTRIK, C. H. B. van; SEBÖK, M.; MUSCAS, G.; VISSER, T.; STA-ARTJES, V. E.; MARINONI, F.; SERRA, C.; REGLI, L.; KRAYENBÜHL, N.; PICCIRELLI, M.; FIERSTRA, J. Topographic volume-standardization atlas of the human brain. **Brain Struct Funct**, Germany, v. 226, n. 6, p. 1699–1711, maio 2021.
- AMSBAUGH, M. J.; KIM, C. S. Brain metastasis. In: **StatPearls**. Treasure Island (FL): StatPearls Publishing, 2023.
- ANDERSEN, E. Imagedata: A python library to handle medical image data in numpy array subclass series. **Journal of Open Source Software**, The Open Journal, v. 7, n. 73, p. 4133, 2022. Disponível em: <<https://doi.org/10.21105/joss.04133>>.
- ARORA, R. D.; AGARWAL, M. S.; MAANI, E. V.; CASCELLA, M. Palliative radiation therapy for brain metastases. In: **StatPearls**. Treasure Island (FL): StatPearls Publishing, 2023.
- BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 12, p. 2481–2495, 2017.
- BALLANGRUD, Å.; KUO, L. C.; HAPPERSETT, L.; LIM, S. B.; BEAL, K.; YAMADA, Y.; HUNT, M.; MECHALAKOS, J. Institutional experience with SRS VMAT planning for multiple cranial metastases. **J Appl Clin Med Phys**, United States, v. 19, n. 2, p. 176–183, fev. 2018.
- BANGERTER, N. K.; HARGREAVES, B. A.; GOLD, G. E.; STUCKER, D. T.; NISHIMURA, D. G. Fluid-attenuated inversion-recovery SSFP imaging. **J Magn Reson Imaging**, United States, v. 24, n. 6, p. 1426–1431, dez. 2006.
- BIBAULT, J.-E.; BURGUN, A.; FOURNIER, L.; DEKKER, A.; LAMBIN, P. Chapter 18 - artificial intelligence in oncology. In: XING, L.; GIGER, M. L.; MIN, J. K. (Ed.). **Artificial Intelligence in Medicine**. Academic Press, 2021. p. 361–381. ISBN 978-0-12-821259-2. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780128212592000181>>.
- BIDGOOD JR, W. D.; HORII, S. C.; PRIOR, F. W.; SYCKLE, D. E. V. Understanding and using DICOM, the data interchange standard for biomedical imaging. **J Am Med Inform Assoc**, England, v. 4, n. 3, p. 199–212, maio 1997.
- BOER, P.-T. de; KROESE, D. P.; MANNOR, S.; RUBINSTEIN, R. Y. A tutorial on the Cross-Entropy method. **Annals of Operations Research**, v. 134, n. 1, p. 19–67, fev. 2005.

BOUGET, D.; ALSINAN, D.; GAITAN, V.; HELLAND, R. H.; PEDERSEN, A.; SOLHEIM, O.; REINERTSEN, I. Raidionics: an open software for pre-and postoperative central nervous system tumor segmentation and standardized reporting. **Scientific Reports**, v. 13, 09 2023.

BOUGET, D.; PEDERSEN, A.; JAKOLA, A. S.; KAVOURIDIS, V.; EMBLEM, K. E.; EIJGELAAR, R. S.; KOMMERS, I.; ARDON, H.; BARKHOF, F.; BELLO, L.; BERGER, M. S.; NIBALI, M. C.; FURTNER, J.; HERVEY-JUMPER, S.; IDEMA, A. J. S.; KIESEL, B.; KLOET, A.; MANDONNET, E.; MÜLLER, D. M. J.; ROBE, P. A.; ROSSI, M.; SCIORTINO, T.; BRINK, W. A. Van den; WAGEMAKERS, M.; WIDHALM, G.; WITTE, M. G.; ZWINDERMAN, A. H.; HAMER, P. C. D. W.; SOLHEIM, O.; REINERTSEN, I. Preoperative brain tumor imaging: Models and software for segmentation and standardized reporting. **Front Neurol**, Switzerland, v. 13, p. 932219, jul. 2022.

BOUSABARAH, K.; RUGE, M.; BRAND, J.-S.; HOVELS, M.; RUESS, D.; BORG-GREFE, J.; HOKAMP, N. G.; VISSER-VANDEWALLE, V.; MAINTZ, D.; TREUER, H.; KOCHER, M. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. **Radiation Oncology**, v. 15, n. 1, p. 87, Apr 2020. ISSN 1748-717X. Disponível em: <<https://doi.org/10.1186/s13014-020-01514-6>>.

BUCHNER, J. A.; KOFLER, F.; ETZEL, L.; MAYINGER, M.; CHRIST, S. M.; BRUNNER, T. B.; WITTIG, A.; MENZE, B.; ZIMMER, C.; MEYER, B. et al. Development and external validation of an mri-based neural network for brain metastasis segmentation in the aurora multicenter study. **Radiotherapy and Oncology**, Elsevier, 2022.

BUCHNER, J. A.; KOFLER, F.; ETZEL, L.; MAYINGER, M.; CHRIST, S. M.; BRUNNER, T. B.; WITTIG, A.; MENZE, B.; ZIMMER, C.; MEYER, B.; GUCKENBERGER, M.; ANDRATSCHKE, N.; El Shafie, R. A.; DEBUS, J.; ROGERS, S.; RIESTERER, O.; SCHULZE, K.; FELDMANN, H. J.; BLANCK, O.; ZAMBOGLOU, C.; FERENTINOS, K.; WOLFF, R.; EITZ, K. A.; COMBS, S. E.; BERNHARDT, D.; WIESTLER, B.; PEEKEN, J. C. Development and external validation of an mri-based neural network for brain metastasis segmentation in the aurora multicenter study. **Radiotherapy and Oncology**, v. 178, p. 109425, 2023. ISSN 0167-8140. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167814022045625>>.

CASTELLANO, A.; BAILO, M.; CICONE, F.; CARIDEO, L.; QUARTUCCIO, N.; MORTINI, P.; FALINI, A.; CASCINI, G. L.; MINNITI, G. Advanced imaging techniques for radiotherapy planning of gliomas. **Cancers (Basel)**, v. 13, n. 5, mar. 2021.

CEAUȘESCU, C.-M.; ALEXE, B. Cerebral metastases segmentation using transfer gliomas learning and grabcut. In: **2022 24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)**. [S.l.: s.n.], 2022. p. 342–346.

CHARRON, O.; LALLEMENT, A.; JARNET, D.; NOBLET, V.; CLAVIER, J.-B.; MEYER, P. Automatic detection and segmentation of brain metastases on multimodal mr images with a deep convolutional neural network. **Computers in Biology and Medicine**, v. 95, p. 43–54, 2018. ISSN 0010-4825. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0010482518300325>>.

CHARTRAND, G.; EMILIANI, R. D.; PAWLOWSKI, S. A.; MARKEL, D. A.; BAHIG, H.; CENGARLE-SAMAK, A.; RAJAKESARI, S.; LAVOIE, J.; DUCHARME, S.; ROBERGE, D. Automated detection of brain metastases on T1-Weighted MRI using a convolutional neural network: Impact of volume aware loss and sampling strategy. **J Magn Reson Imaging**, United States, v. 56, n. 6, p. 1885–1898, maio 2022.

CHO, J.; KIM, Y. J.; SUNWOO, L.; LEE, G. P.; NGUYEN, T. Q.; CHO, S. J.; BAIK, S. H.; BAE, Y. J.; CHOI, B. S.; JUNG, C.; SOHN, C.-H.; HAN, J.-H.; KIM, C.-Y.; KIM, K. G.; KIM, J. H. Deep learning-based computer-aided detection system for automated treatment response assessment of brain metastases on 3d mri. **Frontiers in Oncology**, v. 11, 2021. ISSN 2234-943X. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fonc.2021.739639>>.

CHO, S. J.; SUNWOO, L.; BAIK, S. H.; BAE, Y. J.; CHOI, B. S.; KIM, J. H. Brain metastasis detection using machine learning: a systematic review and meta-analysis. **Neuro-Oncology**, v. 23, n. 2, p. 214–225, 10 2020. ISSN 1522-8517. Disponível em: <<https://doi.org/10.1093/neuonc/noaa232>>.

CHOLLET, F. **Deep Learning with Python**. [S.l.]: Manning, 2017. ISBN 9781617294433.

DEMIRAL, S.; SAGER, O.; DINCOGLAN, F.; BEYZADEOGLU, M. Assessment of Computed Tomography (CT) And Magnetic Resonance Imaging (MRI) Based Radio-surgery Treatment Planning for Pituitary Adenomas. **Cancer Therapy & Oncology International Journal**, v. 13, n. 2, p. 59–62, March 2019. Disponível em: <<https://ideas.repec.org/a/adp/jctoj/v13y2019i2p59-62.html>>.

DESAI, R.; RICH, K. M. Therapeutic role of gamma knife stereotactic radiosurgery in Neuro-Oncology. **Mo Med**, United States, v. 117, n. 1, p. 33–38, jan. 2020.

DIKICI, E.; RYU, J. L.; DEMIRER, M.; BIGELOW, M.; WHITE, R. D.; SLONE, W.; ERDAL, B. S.; PREVEDELLO, L. M. Automated brain metastases detection framework for T1-Weighted Contrast-Enhanced 3D MRI. **IEEE J Biomed Health Inform**, United States, v. 24, n. 10, p. 2883–2893, mar. 2020.

DUCHI, J. C.; HAZAN, E.; SINGER, Y. Adaptive subgradient methods for online learning and stochastic optimization. **J. Mach. Learn. Res.**, v. 12, p. 2121–2159, 2011. Disponível em: <<https://api.semanticscholar.org/CorpusID:538820>>.

FILIPEK, P. A.; RICHELME, C.; KENNEDY, D. N.; CAVINESS VERNE S., J. The Young Adult Human Brain: An MRI-based Morphometric Analysis. **Cerebral Cortex**, v. 4, n. 4, p. 344–360, 07 1994. ISSN 1047-3211. Disponível em: <<https://doi.org/10.1093/cercor/4.4.344>>.

FINK, K. R.; FINK, J. R. Imaging of brain metastases. **Surg Neurol Int**, United States, v. 4, n. Suppl 4, p. S209–19, maio 2013.

GRØVIK, E.; YI, D.; IV, M.; TONG, E.; RUBIN, D.; ZAHARCHUK, G. Deep learning enables automatic detection and segmentation of brain metastases on multisequence mri. **Journal of Magnetic Resonance Imaging**, v. 51, n. 1, p. 175–182, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.26766>>.

GRØVIK, E.; YI, D.; IV, M.; TONG, E.; NILSEN, L. B.; LATYSHEVA, A.; SAXHAUG, C.; JACOBSEN, K. D.; HELLAND, Å.; EMBLEM, K. E.; RUBIN, D. L.; ZAHARCHUK, G. Handling missing mri sequences in deep learning segmentation of brain metastases: a multicenter study. **npj Digital Medicine**, v. 4, n. 1, p. 33, Feb 2021. ISSN 2398-6352. Disponível em: <<https://doi.org/10.1038/s41746-021-00398-4>>.

HAMMER, Y.; NAJJAR, W.; KAHANOV, L.; JOSKOWICZ, L.; SHOSHAN, Y. Two is better than one: longitudinal detection and volumetric evaluation of brain metastases after stereotactic radiosurgery with a deep learning pipeline. **J. Neurooncol.**, Springer Science and Business Media LLC, v. 166, n. 3, p. 547–555, fev. 2024.

HAMMERS, A.; ALLOM, R.; KOEPP, M. J.; FREE, S. L.; MYERS, R.; LEMIEUX, L.; MITCHELL, T. N.; BROOKS, D. J.; DUNCAN, J. S. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. **Human Brain Mapping**, v. 19, n. 4, p. 224–247, 2003. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.10123>>.

HSU, D. G.; BALLANGRUD, Å.; SHAMSEDDINE, A.; DEASY, J. O.; VEERARAGHAVAN, H.; CERVINO, L.; BEAL, K.; ARISTOPHANOUS, M. Automatic segmentation of brain metastases using T1 magnetic resonance and computed tomography images. **Phys Med Biol**, England, v. 66, n. 17, ago. 2021.

HU, S.-Y.; WENG, W.-H.; LU, S.-L.; CHENG, Y.-H.; XIAO, F.; HSU, F.-M.; LU, J.-T. Multimodal volume-aware detection and segmentation for brain metastases radiosurgery. In: NGUYEN, D.; XING, L.; JIANG, S. (Ed.). **Artificial Intelligence in Radiation Therapy**. Cham: Springer International Publishing, 2019. p. 61–69. ISBN 978-3-030-32486-5.

HUANG, Y.; BERT, C.; SOMMER, P.; FREY, B.; GAJPL, U.; DISTEL, L. V.; WEISSMANN, T.; UDER, M.; SCHMIDT, M. A.; DÖRFLER, A.; MAIER, A.; FIETKAU, R.; PUTZ, F. Deep learning for brain metastasis detection and segmentation in longitudinal MRI data. **Med Phys**, United States, v. 49, n. 9, p. 5773–5786, jul. 2022.

HUTTENLOCHER, D.; KLANDERMAN, G.; RUCKLIDGE, W. Comparing images using the hausdorff distance. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 15, n. 9, p. 850–863, 1993.

IAEA. **Artificial Intelligence in Medical Physics**. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2023. (Training Course Series, 83). Disponível em: <<https://www.iaea.org/publications/15450/artificial-intelligence-in-medical-physics>>.

ICRU. **Prescribing, Recording, and Reporting Photon Beam Therapy**. 7910 Woodmont Avenue, Suite 800, Bethesda, Maryland 20814, U.S.A., 2015.

ISENSEE, F.; JAEGER, P. F.; KOHL, S. A. A.; PETERSEN, J.; MAIER-HEIN, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. **Nat. Methods**, Springer Science and Business Media LLC, v. 18, n. 2, p. 203–211, fev. 2021.

ISENSEE, F.; SCHELL, M.; PFLUEGER, I.; BRUGNARA, G.; BONEKAMP, D.; NEUBERGER, U.; WICK, A.; SCHLEMMER, H.-P.; HEILAND, S.; WICK, W.; BENDSZUS,

- M.; MAIER-HEIN, K. H.; KICKINGEREDER, P. Automated brain extraction of multi-sequence mri using artificial neural networks. **Human Brain Mapping**, v. 40, n. 17, p. 4952–4964, 2019. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24750>>.
- JALALIFAR, A.; SOLIMAN, H.; SAHGAL, A.; SADEGHI-NAINI, A. A cascaded Deep-Learning framework for segmentation of metastatic brain tumors before and after stereotactic radiation therapy(). **Annu Int Conf IEEE Eng Med Biol Soc**, United States, v. 2020, p. 1063–1066, jul. 2020.
- JALALIFAR, S. A.; SOLIMAN, H.; SAHGAL, A.; SADEGHI-NAINI, A. Impact of tumour segmentation accuracy on efficacy of quantitative mri biomarkers of radiotherapy outcome in brain metastasis. **Cancers**, v. 14, n. 20, 2022. ISSN 2072-6694. Disponível em: <<https://www.mdpi.com/2072-6694/14/20/5133>>.
- JALALIFAR, S. A.; SOLIMAN, H.; SAHGAL, A.; SADEGHI-NAINI, A. Automatic assessment of stereotactic radiation therapy outcome in brain metastasis using longitudinal segmentation on serial MRI. **IEEE J Biomed Health Inform**, United States, PP, jan. 2023.
- JAYAMANI, J.; OSMAN, N. D.; TAJUDDIN, A. A.; SALEHI, Z.; ALI, M. H.; AZIZ, M. Z. A. Determination of computed tomography number of high-density materials in 12-bit, 12-bit extended and 16-bit depth for dosimetric calculation in treatment planning system. **Journal of Radiotherapy in Practice**, v. 18, n. 3, p. 285–294, 2019.
- JÜNGER, S. T.; HOYER, U. C. I.; SCHAUFLER, D.; LAUKAMP, K. R.; GOERTZ, L.; THIELE, F.; GRUNZ, J.-P.; SCHLAMANN, M.; PERKUHN, M.; KABBASCH, C.; PER-SIGEHL, T.; GRAU, S.; BORGGREFE, J.; SCHEFFLER, M.; SHAHZAD, R.; PENNIG, L. Fully automated MR detection and segmentation of brain metastases in non-small cell lung cancer using deep learning. **J Magn Reson Imaging**, United States, v. 54, n. 5, p. 1608–1622, maio 2021.
- KAMNITSAS, K.; LEDIG, C.; NEWCOMBE, V. F.; SIMPSON, J. P.; KANE, A. D.; MENON, D. K.; RUECKERT, D.; GLOCKER, B. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. **Medical Image Analysis**, v. 36, p. 61–78, 2017. ISSN 1361-8415. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1361841516301839>>.
- KARIMI, D.; WARFIELD, S. K.; GHOLIPOUR, A. Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations. **Artif Intell Med**, Netherlands, v. 116, p. 102078, abr. 2021.
- KERVADEC, H.; BOUCHTIBA, J.; DESROSIERS, C.; GRANGER, E.; DOLZ, J.; Ben Ayed, I. Boundary loss for highly unbalanced segmentation. **Medical Image Analysis**, v. 67, p. 101851, 2021. ISSN 1361-8415. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1361841520302152>>.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.
- KINOSHITA, T.; OGAWA, T.; YOSHIDA, Y.; TAMURA, H.; KADO, H.; OKUDERA, T. Curvilinear T1 hyperintense lesions representing cortical necrosis after cerebral infarction. **Neuroradiology**, Germany, v. 47, n. 9, p. 647–651, jul. 2005.

LANCASTER, J. L.; CYKOWSKI, M. D.; MCKAY, D. R.; KOCHUNOV, P. V.; FOX, P. T.; ROGERS, W.; TOGA, A. W.; ZILLES, K.; AMUNTS, K.; MAZZIOTTA, J. Anatomical global spatial normalization. **Neuroinformatics**, v. 8, n. 3, p. 171–182, out. 2010.

LEE, N. Y. **Target volume delineation and field setup A practical guide for conformal and intensity-modulated radiation therapy**. Germany: Springer, 2013. RADIOLOGY AND NUCLEAR MEDICINE. ISBN 978-3-642-28859-3. Disponível em: <http://inis.iaea.org/search/search.aspx?orig_q=RN:44039493>.

LEE, W.-K.; YANG, H.-C.; LEE, C.-C.; LU, C.-F.; WU, C.-C.; CHUNG, W.-Y.; WU, H.-M.; GUO, W.-Y.; WU, Y.-T. Lesion delineation framework for vestibular schwannoma, meningioma and brain metastasis for gamma knife radiosurgery using stereotactic magnetic resonance images. **Comput Methods Programs Biomed**, Ireland, v. 229, p. 107311, dez. 2022.

LI, C.; LI, K.; ZHONG, S.; TANG, M.; SHI, X.; BAO, Y. Which is the best treatment for melanoma brain metastases? a bayesian network meta-analysis and systematic review. **Critical Reviews in Oncology/Hematology**, v. 194, p. 104227, 2024. ISSN 1040-8428. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1040842823003153>>.

LI, R.; GUO, Y.; ZHAO, Z.; CHEN, M.; LIU, X.; GONG, G.; WANG, L. MRI-based two-stage deep learning model for automatic detection and segmentation of brain metastases. **Eur Radiol**, Germany, v. 33, n. 5, p. 3521–3531, jan. 2023.

LIANG, Y.; LEE, K.; BOVI, J. A.; PALMER, J. D.; BROWN, P. D.; GONDI, V.; TOMÉ, W. A.; BENZINGER, T. L. S.; MEHTA, M. P.; LI, X. A. Deep Learning-Based automatic detection of brain metastases in heterogenous Multi-Institutional magnetic resonance imaging sets: An exploratory analysis of NRG-CC001. **Int J Radiat Oncol Biol Phys**, United States, v. 114, n. 3, p. 529–536, jul. 2022.

LIEW, A.; LEE, C. C.; SUBRAMANIAM, V.; LAN, B. L.; TAN, M. Gradual Self-Training via confidence and volume based domain adaptation for multi dataset deep Learning-Based brain metastases detection using nonlocal networks on MRI images. **J Magn Reson Imaging**, United States, out. 2022.

LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLÁR, P. Focal loss for dense object detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 42, n. 2, p. 318–327, 2020.

LIN, X.; DEANGELIS, L. M. Treatment of brain metastases. **J Clin Oncol**, United States, v. 33, n. 30, p. 3475–3484, ago. 2015.

LIU, Y.; STOJADINOVIC, S.; HRYCUSHKO, B.; WARDAK, Z.; LAU, S.; LU, W.; YAN, Y.; JIANG, S. B.; ZHEN, X.; TIMMERMAN, R.; NEDZI, L.; GU, X. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. **PLoS One**, United States, v. 12, n. 10, p. e0185844, out. 2017.

LU, S.-L.; XIAO, F.-R.; CHENG, J. C.-H.; YANG, W.-C.; CHENG, Y.-H.; CHANG, Y.-C.; LIN, J.-Y.; LIANG, C.-H.; LU, J.-T.; CHEN, Y.-F.; HSU, F.-M. Randomized multi-reader evaluation of automated detection and segmentation of brain tumors in stereotactic

radiosurgery with deep neural networks. **Neuro. Oncol.**, Oxford University Press (OUP), v. 23, n. 9, p. 1560–1568, set. 2021.

LYU, Q.; NAMJOSHI, S. V.; MCTYRE, E.; TOPALOGLU, U.; BARCUS, R.; CHAN, M. D.; CRAMER, C. K.; DEBINSKI, W.; GURCAN, M. N.; LESSER, G. J.; LIN, H.-K.; MUNDEN, R. F.; PASCHE, B. C.; SAI, K. K. S.; STROWD, R. E.; TATTER, S. B.; WATABE, K.; ZHANG, W.; WANG, G.; WHITLOW, C. T. A transformer-based deep-learning approach for classifying brain metastases into primary organ sites using clinical whole-brain MRI images. **Patterns (N Y)**, United States, v. 3, n. 11, p. 100613, out. 2022.

MACKIE, T. R.; GREGOIRE, V. Assessment and management of uncertainties: Overview and examples in different diseases sites. In: AAPM SUMMER SCHOOL. 2011. Disponível em: <<https://www.aapm.org/meetings/2011SS/documents/MackieUncertainty.pdf>>.

MASON, D. Su-e-t-33: Pydicom: An open source dicom library. **Medical Physics**, v. 38, p. 3493–3493, 2011.

MAZZOLA, A. A. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. **Revista Brasileira de Física Médica**, v. 3, n. 1, p. 117–129, out. 2015. Disponível em: <<https://www.rbmf.org.br/rbmf/article/view/51>>.

MEHRABIAN, H.; DETSKY, J.; SOLIMAN, H.; SAHGAL, A.; STANISZ, G. J. Advanced magnetic resonance imaging techniques in management of brain metastases. **Frontiers in Oncology**, v. 9, 2019. ISSN 2234-943X. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fonc.2019.00440>>.

MENZE, B. H.; JAKAB, A.; BAUER, S.; KALPATHY-CRAMER, J.; FARAHANI, K.; KIRBY, J.; BURREN, Y.; PORZ, N.; SLOTBOOM, J.; WIEST, R.; LANCZI, L.; GERSTNER, E.; WEBER, M.-A.; ARBEL, T.; AVANTS, B. B.; AYACHE, N.; BUENDIA, P.; COLLINS, D. L.; CORDIER, N.; CORSO, J. J.; CRIMINISI, A.; DAS, T.; DELINGETTE, H.; DEMIRALP, C.; DURST, C. R.; DOJAT, M.; DOYLE, S.; FESTA, J.; FORBES, F.; GEREMIA, E.; GLOCKER, B.; GOLLAND, P.; GUO, X.; HAMAMCI, A.; IFTEKHARUDDIN, K. M.; JENA, R.; JOHN, N. M.; KONUKOGLU, E.; LASHKARI, D.; MARIZ, J. A.; MEIER, R.; PEREIRA, S.; PRECUP, D.; PRICE, S. J.; RAVIV, T. R.; REZA, S. M. S.; RYAN, M.; SARIKAYA, D.; SCHWARTZ, L.; SHIN, H.-C.; SHOTTON, J.; SILVA, C. A.; SOUSA, N.; SUBBANNA, N. K.; SZEKELY, G.; TAYLOR, T. J.; THOMAS, O. M.; TUSTISON, N. J.; UNAL, G.; VASSEUR, F.; WINTERMARK, M.; YE, D. H.; ZHAO, L.; ZHAO, B.; ZIKIC, D.; PRASTAWA, M.; REYES, M.; LEEMPUT, K. V. The multimodal brain tumor image segmentation benchmark (brats). **IEEE Transactions on Medical Imaging**, v. 34, n. 10, p. 1993–2024, 2015.

MONGAN, J.; MOY, L.; KAHN, C. E. Checklist for artificial intelligence in medical imaging (claim): A guide for authors and reviewers. **Radiology: Artificial Intelligence**, v. 2, n. 2, p. e200029, 2020. PMID: 33937821. Disponível em: <<https://doi.org/10.1148/ryai.2020200029>>.

MORI, Y.; HAYASHI, N.; IWASE, M.; YAMADA, M.; TAKIKAWA, Y.; UCHIYAMA, Y.; ODA, K.; KAI, O. Stereotactic imaging for radiosurgery: localization accuracy of magnetic resonance imaging and positron emission tomography compared with computed tomography. **Stereotact Funct Neurosurg**, Switzerland, v. 84, n. 4, p. 142–146, ago. 2006.

NAQA, I. E.; BOONE, J. M.; BENEDICT, S. H.; GOODSITT, M. M.; CHAN, H.-P.; DRUKKER, K.; HADJIISKI, L.; RUAN, D.; SAHINER, B. Ai in medical physics: guidelines for publication. **Medical Physics**, v. 48, n. 9, p. 4711–4714, 2021. Disponível em: <<https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.15170>>.

NIRANJAN, A.; MONACO, E.; FLICKINGER, J.; LUNSFORD, L. D. Guidelines for multiple brain metastases radiosurgery. **Prog Neurol Surg**, Switzerland, v. 34, p. 100–109, maio 2019.

NOMURA, Y.; HANAOKA, S.; TAKENAGA, T.; NAKAO, T.; SHIBATA, H.; MIKI, S.; YOSHIKAWA, T.; WATADANI, T.; HAYASHI, N.; ABE, O. Preliminary study of generalized semiautomatic segmentation for 3D voxel labeling of lesions based on deep learning. **Int. J. Comput. Assist. Radiol. Surg.**, Springer Science and Business Media LLC, v. 16, n. 11, p. 1901–1913, nov. 2021.

OCAÑA-TIENDA, B.; PÉREZ-BETETA, J.; VILLANUEVA-GARCÍA, J. D.; ROMERO-ROSALES, J. A.; MOLINA-GARCÍA, D.; SUTER, Y.; ASENJO, B.; ALBILLO, D.; MENDIVIL, A. Ortiz de; PÉREZ-ROMASANTA, L. A.; PORTILLO, E. G.-D.; LLORENTE, M.; CARBALLO, N.; NAGIB-RAYA, F.; VIDAL-DENIS, M.; LUQUE, B.; REYES, M.; ARANA, E.; PÉREZ-GARCÍA, V. M. A comprehensive dataset of annotated brain metastasis mr images with clinical and radiomic data. **Scientific Data**, v. 10, n. 1, p. 208, Apr 2023. ISSN 2052-4463. Disponível em: <<https://doi.org/10.1038/s41597-023-02123-0>>.

OKUDA, T.; KOROGI, Y.; SHIGEMATSU, Y.; SUGAHARA, T.; HIRAI, T.; IKUSHIMA, I.; LIANG, L.; TAKAHASHI, M. Brain lesions: when should fluid-attenuated inversion-recovery sequences be used in MR evaluation? **Radiology**, United States, v. 212, n. 3, p. 793–798, set. 1999.

OTTESEN, J. A.; YI, D.; TONG, E.; IV, M.; LATYSHEVA, A.; SAXHAUG, C.; JACOBSEN, K. D.; HELLAND, A.; EMBLEM, K. E.; RUBIN, D. L.; BJØRNERUD, A.; ZAHARCHUK, G.; GRØVIK, E. 2.5d and 3d segmentation of brain metastases with deep learning on multinational mri data. **Frontiers in Neuroinformatics**, v. 16, 2023. ISSN 1662-5196. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fninf.2022.1056068>>.

PAGE, M. J.; MCKENZIE, J. E.; BOSSUYT, P. M.; BOUTRON, I.; HOFFMANN, T. C.; MULROW, C. D.; SHAMSEER, L.; TETZLAFF, J. M.; AKL, E. A.; BRENNAN, S. E.; CHOU, R.; GLANVILLE, J.; GRIMSHAW, J. M.; HRÓBJARTSSON, A.; LALU, M. M.; LI, T.; LODER, E. W.; MAYO-WILSON, E.; MCDONALD, S.; MCGUINNESS, L. A.; STEWART, L. A.; THOMAS, J.; TRICCO, A. C.; WELCH, V. A.; WHITING, P.; MOHER, D. The prisma 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, BMJ Publishing Group Ltd, v. 372, 2021. Disponível em: <<https://www.bmj.com/content/372/bmj.n71>>.

PARK, Y. R.; KIM, Y. J.; KIM, K. G. Deep learning-based brain metastatic detection and treatment response assessment system on 3d mri. In: **2021 International Conference on Information and Communication Technology Convergence (ICTC)**. [S.l.: s.n.], 2021. p. 1–3.

PARK, Y. W.; JUN, Y.; LEE, Y.; HAN, K.; AN, C.; AHN, S. S.; HWANG, D.; LEE, S.-K. Robust performance of deep learning for automatic detection and segmentation

of brain metastases using three-dimensional black-blood and three-dimensional gradient echo imaging. **Eur Radiol**, Germany, v. 31, n. 9, p. 6686–6695, mar. 2021.

PENNIG, L.; SHAHZAD, R.; CALDEIRA, L.; LENNARTZ, S.; THIELE, F.; GOERTZ, L.; ZOPFS, D.; MEISSNER, A.-K.; FÜRTJES, G.; PERKUHN, M.; KABBASCH, C.; GRAU, S.; BORGGREFE, J.; LAUKAMP, K. R. Automated detection and segmentation of brain metastases in malignant melanoma: Evaluation of a dedicated deep learning model. **AJNR Am J Neuroradiol**, United States, v. 42, n. 4, p. 655–662, fev. 2021.

POPE, W. B. Brain metastases: neuroimaging. **Handb Clin Neurol**, Netherlands, v. 149, p. 89–112, 2018.

PREZELSKI, K.; HSU, D. G.; BALZO, L. D.; HELLER, E.; MA, J.; PIKE, L. R. G.; BALLANGRUD, Å.; ARISTOPHANOUS, M. Artificial-intelligence-driven measurements of brain metastases' response to SRS compare favorably with current manual standards of assessment. **Neurooncol. Adv.**, Oxford University Press (OUP), v. 6, n. 1, p. vdae015, jan. 2024.

QU, J.; ZHANG, W.; SHU, X.; WANG, Y.; WANG, L.; XU, M.; YAO, L.; HU, N.; TANG, B.; ZHANG, L.; LUI, S. Construction and evaluation of a gated high-resolution neural network for automatic brain metastasis detection and segmentation. **Eur. Radiol.**, v. 33, n. 10, p. 6648–6658, out. 2023.

RAMAKRISHNAN, D.; JEKEL, L.; CHADHA, S.; JANAS, A.; MOY, H.; MALEKI, N.; SALA, M.; KAUR, M.; PETERSEN, G. C.; MERKAJ, S.; REPPERT, M. von; BAID, U.; BAKAS, S.; KIRSCH, C.; DAVIS, M.; BOUSABARAH, K.; HOLLER, W.; LIN, M.; WESTERHOFF, M.; ANEJA, S.; MEMON, F.; ABOIAN, M. S. A large open access dataset of brain metastasis 3d segmentations on mri with clinical and imaging information. **Scientific Data**, v. 11, n. 1, p. 254, Feb 2024. ISSN 2052-4463. Disponível em: <<https://doi.org/10.1038/s41597-024-03021-9>>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). **Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015**. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4.

RUDER, S. An overview of gradient descent optimization algorithms. **CoRR**, abs/1609.04747, 2016. Disponível em: <<http://arxiv.org/abs/1609.04747>>.

RUDIE, J. D.; WEISS, D. A.; COLBY, J. B.; RAUSCHECKER, A. M.; LAGUNA, B.; BRAUNSTEIN, S.; SUGRUE, L. P.; HESS, C. P.; VILLANUEVA-MEYER, J. E. Three-dimensional u-net convolutional neural network for detection and segmentation of intracranial metastases. **Radiology: Artificial Intelligence**, v. 3, n. 3, p. e200204, 2021. Disponível em: <<https://doi.org/10.1148/ryai.2021200204>>.

SALEHI, S. S. M.; ERDOGMUS, D.; GHOLIPOUR, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: WANG, Q.; SHI, Y.; SUK, H.-I.; SUZUKI, K. (Ed.). **Machine Learning in Medical Imaging**. Cham: Springer International Publishing, 2017. p. 379–387. ISBN 978-3-319-67389-9.

SARVAMANGALA, D. R.; KULKARNI, R. V. Convolutional neural networks in medical image understanding: a survey. **Evolutionary Intelligence**, v. 15, n. 1, p. 1–22, Mar 2022. ISSN 1864-5917. Disponível em: <<https://doi.org/10.1007/s12065-020-00540-3>>.

SCHNEIDER, T.; KUHNE, J. F.; BITTRICH, P.; SCHROEDER, J.; MAGNUS, T.; MOHME, M.; GROSSER, M.; SCHOEN, G.; FIEHLER, J.; SIEMONSEN, S. Edema is not a reliable diagnostic sign to exclude small brain metastases. **PLOS ONE**, Public Library of Science, v. 12, n. 5, p. e0177217, May 2017. Disponível em: <<https://doi.org/10.1371/journal.pone.0177217>>.

SHAW, E. G.; GASPAR, L. E.; GIBBS, F. A.; LEWIN, A. A.; JR, M. D. W.; LARSON, D.; BLOOMER, W. D.; BUCKLEY, J. A.; LOEFFLER, J. S.; MALCOLM, A. W.; MENDENHALL, W. M.; SCHNEIDER, J. F.; SCHUPAK, K. D.; SIMPSON, J. R.; GUTIN, P. H.; ROGERS, L.; LEIBEL, S. **Critérios de Adequação de Exames de Imagem e Radioterapia: Metástases cerebrais múltiplas**. Colégio Brasileiro de Radiologia e Diagnóstico por Imagem, 2005. ISBN 85-87950-06-1. Disponível em: <https://cbr.org.br/wp-content/uploads/2017/06/07_03.pdf>.

SHIROKIKH, B.; DALECHINA, A.; SHEVTSOV, A.; KRIVOV, E.; KOSTJUCHENKO, V.; DURGARYAN, A.; GALKIN, M.; GOLANOV, A.; BELYAEV, M. Systematic clinical evaluation of a deep learning method for medical image segmentation: Radiosurgery application. **IEEE Journal of Biomedical and Health Informatics**, v. 26, n. 7, p. 3037–3046, 2022.

SHU, X.; ZHANG, L.; QU, J.; WANG, L.; WANG, Z.; ZHANG, W.; WANG, Y.; LUI, S. Deep slice-crossed network with local weighted loss for brain metastases segmentation. **IEEE Transactions on Cognitive and Developmental Systems**, v. 15, n. 3, p. 1419–1429, 2023.

SON, S.; JOO, B.; PARK, M.; SUH, S. H.; OH, H. S.; KIM, J. W.; LEE, S.; AHN, S. J.; LEE, J.-M. Development of RLK-Unet: a clinically favorable deep learning algorithm for brain metastasis detection and treatment response assessment. **Front. Oncol.**, v. 13, p. 1273013, 2023.

STANLEY, D. W.; KOTSENAS, A. L.; KAUFMANN, T. J.; EDMONSON, H. A.; RETTMANN, D. W.; HAN, E. T. T1 cube compared to fast spin echo t1 weighted and bravo in post contrast enhanced brain mri at 3t. In: . [s.n.], 2009. Disponível em: <<https://api.semanticscholar.org/CorpusID:51778093>>.

SUDRE, C. H.; LI, W.; VERCAUTEREN, T.; OURSELIN, S.; CARDOSO, M. J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: CARDOSO, M. J.; ARBEL, T.; CARNEIRO, G.; SYEDA-MAHMOOD, T.; TAVARES, J. M. R.; MORADI, M.; BRADLEY, A.; GREENSPAN, H.; PAPA, J. P.; MADABHUSHI, A.; NASCIMENTO, J. C.; CARDOSO, J. S.; BELAGIANNIS, V.; LU, Z. (Ed.). **Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support**. Cham: Springer International Publishing, 2017. p. 240–248. ISBN 978-3-319-67558-9.

SUGINO, T.; KAWASE, T.; ONOGI, S.; KIN, T.; SAITO, N.; NAKAJIMA, Y. Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks. **Healthcare (Basel)**, Switzerland, v. 9, n. 8, jul. 2021.

SUN, K.; XIAO, B.; LIU, D.; WANG, J. Deep high-resolution representation learning for human pose estimation. In: **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2019. p. 5686–5696.

SUN, Y.-C.; HSIEH, A.-T.; FANG, S.-T.; WU, H.-M.; KAO, L.-W.; CHUNG, W.-Y.; CHEN, H.-H.; LIOU, K.-D.; LIN, Y.-S.; GUO, W.-Y.; LU, H. H.-S. Can 3D artificial intelligence models outshine 2D ones in the detection of intracranial metastatic tumors on magnetic resonance images? **J Chin Med Assoc**, Netherlands, v. 84, n. 10, p. 956–962, out. 2021.

TEJANI, A. S.; KLONTZAS, M. E.; GATTI, A. A.; MONGAN, J. T.; MOY, L.; PARK, S. H.; KAHN, C. E. a. Checklist for artificial intelligence in medical imaging (claim): 2024 update. **Radiology: Artificial Intelligence**, v. 0, n. ja, p. e240300, 2024. PMID: 38809149. Disponível em: <<https://doi.org/10.1148/ryai.240300>>.

TOZUKA, T.; MINEGISHI, Y.; YAMAGUCHI, O.; WATANABE, K.; TOI, Y.; SAITO, R.; NAGAI, Y.; TAMURA, Y.; SHOJI, T.; ODAGIRI, H.; EBI, N.; SAKAI, K.; KANAJI, N.; IZUMI, M.; SODA, S.; WATANABE, S.; MORITA, S.; KOBAYASHI, K.; SEIKE, M. Immunotherapy with radiotherapy for brain metastases in patients with NSCLC: NEJ060. **JTO Clin Res Rep**, United States, v. 5, n. 4, p. 100655, fev. 2024.

TSAO, M. N.; RADES, D.; WIRTH, A.; LO, S. S.; DANIELSON, B. L.; GASPAR, L. E.; SPERDUTO, P. W.; VOGELBAUM, M. A.; RADAWSKI, J. D.; WANG, J. Z.; GILLIN, M. T.; MOHIDEEN, N.; HAHN, C. A.; CHANG, E. L. Radiotherapeutic and surgical management for newly diagnosed brain metastasis(es): An american society for radiation oncology evidence-based guideline. **Pract Radiat Oncol**, United States, v. 2, n. 3, p. 210–225, jan. 2012.

VARIAN. **Eclipse 13 RapidArc/VMAT Planning Operations**. Palo Alto, CA 94304-1030, U.S.A., 2014.

VARIAN. **Instruções de uso de fótons e elétrons do Eclipse**. Palo Alto, CA 94304, U.S.A., 2022.

VOULODIMOS, A.; DOULAMIS, N.; DOULAMIS, A.; PROTOPAPADAKIS, E. Deep learning for computer vision: A brief review. **Computational Intelligence and Neuroscience**, Hindawi, v. 2018, p. 7068349, Feb 2018. ISSN 1687-5265. Disponível em: <<https://doi.org/10.1155/2018/7068349>>.

WANG, J.-Y.; QU, V.; HUI, C.; SANDHU, N.; MENDOZA, M. G.; PANJWANI, N.; CHANG, Y.-C.; LIANG, C.-H.; LU, J.-T.; WANG, L.; KOVALCHUK, N.; GENSHEIMER, M. F.; SOLTYS, S. G.; POLLOM, E. L. Stratified assessment of an FDA-cleared deep learning algorithm for automated detection and contouring of metastatic brain tumors in stereotactic radiosurgery. **Radiat. Oncol.**, Springer Science and Business Media LLC, v. 18, n. 1, p. 61, abr. 2023.

XUE, J.; REN, L.; SONG, B.; GUO, Y.; LU, J.; LIU, X.; GONG, G.; LI, D. Hypergraph-based numerical neural-like p systems for medical image segmentation. **IEEE Transactions on Parallel and Distributed Systems**, v. 34, n. 4, p. 1202–1214, 2023.

XUE, J.; WANG, B.; MING, Y.; LIU, X.; JIANG, Z.; WANG, C.; LIU, X.; CHEN, L.; QU, J.; XU, S.; TANG, X.; MAO, Y.; LIU, Y.; LI, D. Deep learning-based detection and

segmentation-assisted management of brain metastases. **Neuro Oncol**, England, v. 22, n. 4, p. 505–514, abr. 2020.

YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, v. 9, n. 4, p. 611–629, Aug 2018. ISSN 1869-4101. Disponível em: <<https://doi.org/10.1007/s13244-018-0639-9>>.

YANG, Z.; CHEN, M.; KAZEMIMOGHADAM, M.; MA, L.; STOJADINOVIC, S.; TIMMERMAN, R.; DAN, T.; WARDAK, Z.; LU, W.; GU, X. Deep-learning and radiomics ensemble classifier for false positive reduction in brain metastases segmentation. **Phys. Med. Biol.**, IOP Publishing, v. 67, n. 2, p. 025004, jan. 2022.

YANG, Z.; LIU, H.; LIU, Y.; STOJADINOVIC, S.; TIMMERMAN, R.; NEDZI, L.; DAN, T.; WARDAK, Z.; LU, W.; GU, X. A web-based brain metastases segmentation and labeling platform for stereotactic radiosurgery. **Med. Phys.**, Wiley, v. 47, n. 8, p. 3263–3276, ago. 2020.

YI, D.; GRØVIK, E.; TONG, E.; IV, M.; EMBLEM, K. E.; NILSEN, L. B.; SAXHAUG, C.; LATYSHEVA, A.; JACOBSEN, K. D.; HELLAND, Å.; ZAHARCHUK, G.; RUBIN, D. MRI pulse sequence integration for deep-learning-based brain metastases segmentation. **Med Phys**, United States, v. 48, n. 10, p. 6020–6035, ago. 2021.

YOO, S. K.; KIM, T. H.; CHUN, J.; CHOI, B. S.; KIM, H.; YANG, S.; YOON, H. I.; KIM, J. S. Deep-Learning-Based automatic detection and segmentation of brain metastases with small volume for stereotactic ablative radiotherapy. **Cancers (Basel)**, Switzerland, v. 14, n. 10, maio 2022.

YOO, Y.; CECCALDI, P.; LIU, S.; RE, T. J.; CAO, Y.; BALTER, J. M.; GIBSON, E. Evaluating deep learning methods in detecting and segmenting different sizes of brain metastases on 3D post-contrast t1-weighted images. **J Med Imaging (Bellingham)**, United States, v. 8, n. 3, p. 037001, maio 2021.

YOON, B. C.; SAAD, A. F.; REZAIL, P.; WINTERMARK, M.; ZAHARCHUK, G.; IV, M. Evaluation of Thick-Slab overlapping MIP images of Contrast-Enhanced 3D T1-Weighted CUBE for detection of intracranial metastases: A pilot study for comparison of lesion detection, interpretation time, and sensitivity with nonoverlapping CUBE MIP, CUBE, and Inversion-Recovery-Prepared Fast-Spoiled gradient recalled brain volume. **AJNR Am J Neuroradiol**, United States, v. 39, n. 9, p. 1635–1642, ago. 2018.

YOUSEF, A. F.; ELKHARBOTLY, A.; SETTIN, M.; MOUSA, Y. Role of diffusion-weighted mr imaging in discrimination between the intracranial cystic masses. **The Egyptian Journal of Radiology and Nuclear Medicine**, v. 45, n. 3, p. 869–875, Sep 2014. ISSN 0378-603X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378603X14000734>>.

ZAKARIA, R.; DAS, K.; BHOJAK, M.; RADON, M.; WALKER, C.; JENKINSON, M. D. The role of magnetic resonance imaging in the management of brain metastases: diagnosis to prognosis. **Cancer Imaging**, England, v. 14, n. 1, p. 8, abr. 2014.

ZHOU, Z.; SANDERS, J. W.; JOHNSON, J. M.; GULE-MONROE, M.; CHEN, M.; BRIERE, T. M.; WANG, Y.; SON, J. B.; PAGEL, M. D.; MA, J.; LI, J. MetNet: Computer-aided segmentation of brain metastases in post-contrast t1-weighted magnetic resonance imaging. **Radiother Oncol**, Ireland, v. 153, p. 189–196, set. 2020.

ZHOU, Z.; WANG, M.; ZHAO, R.; SHAO, Y.; XING, L.; QIU, Q.; YIN, Y. A multi-task deep learning model for EGFR genotyping prediction and GTV segmentation of brain metastasis. **J. Transl. Med.**, Springer Science and Business Media LLC, v. 21, n. 1, p. 788, nov. 2023.

Apêndices

APÊNDICE A – Contagem de *clusters* em Python

```

1 import numpy as np
2 import nibabel as nib
3 from scipy.ndimage import label
4
5 def identify_clusters(mask, min_cluster_size = 6):
6     labeled_mask, num_clusters = label(mask)
7     filtered_clusters = labeled_mask.copy()
8
9     for cluster_id in range(1, num_clusters + 1):
10        cluster_size = np.sum(labeled_mask == cluster_id)
11        if cluster_size < min_cluster_size:
12            filtered_clusters[labeled_mask == cluster_id] = 0
13
14        filtered_clusters, num_filtered_clusters = label(
15            filtered_clusters > 0)
16    return filtered_clusters, num_filtered_clusters
17
18 def count_clusters(mask1_path, mask2_path, dataset, modelo,
19 min_clustersize):
20
21     mask1 = nib.load(mask1_path).get_fdata()
22     mask2 = nib.load(mask2_path).get_fdata()
23
24     if dataset == 'OcanaTienda':
25         mask2 = np.where(mask2 > 0, 1, 0)
26     if dataset == 'TN_TCIA' or dataset == 'BrainMetShare' or
27         dataset == 'BMS_NIfTI':
28         mask2 = mask2
29     if modelo == 'AURORA':
30         mask1 = np.where(mask1 > 1, 1, 0)
31
32     mask1_clusters, num_mask1_clusters = identify_clusters(mask1,
33         min_clustersize = 6)
34     mask2_clusters, num_mask2_clusters = identify_clusters(mask2,
35         min_clustersize)
36
37     return num_mask1_clusters, num_mask2_clusters

```

Listing A.1 – Código em Python para a contagem de *clusters*