

UNIVERSIDADE FEDERAL DO PARANÁ

FLÁVIA SPRENGER

SEGMENTAÇÃO SEMIAUTOMÁTICA E CORRELAÇÃO COM ACHADOS
CIRÚRGICOS E HISTOPATOLÓGICOS EM GLIOMAS DE ALTO GRAU

CURITIBA

2023

FLÁVIA SPRENGER

SEGMENTAÇÃO SEMIAUTOMÁTICA E CORRELAÇÃO COM ACHADOS
CIRÚRGICOS E HISTOPATOLÓGICOS EM GLIOMAS DE ALTO GRAU

Dissertação de Mestrado apresentado ao curso de Pós-Graduação em Medicina Interna e Ciências da Saúde, setor de Ciências da Saúde, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Medicina Interna e Ciências da Saúde.

Orientador: Prof. Dr. Bernardo Corrêa de Almeida
Teixeira

CURITIBA

2023

S768 Sprenger, Flávia

Segmentação semiautomática e correlação com achados cirúrgicos e histopatológicos em gliomas de alto grau [recurso eletrônico] / Flávia Sprenger. – Curitiba, 2023.

Dissertação (mestrado) – Programa de Pós-Graduação em Medicina Interna e Ciências da Saúde. Setor de Ciências da Saúde. Universidade Federal do Paraná.

Orientador: Prof. Dr. Bernardo Corrêa de Almeida Teixeira

1. Títulometria. 2. Glioma. 3. Antígeno Ki-67. 4. Neoplasias. I. Teixeira, Bernardo Corrêa de Almeida II. Programa de Pós-Graduação em Medicina Interna e Ciências da Saúde. Setor de Ciências da Saúde. Universidade Federal do Paraná. III. Título.

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE BIBLIOTECAS/UFPR
BIBLIOTECA DE CIÊNCIAS DA SAÚDE, BIBLIOTECÁRIA: RAQUEL PINHEIRO COSTA
JORDÃO CRB 9/991



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS DA SAÚDE
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MEDICINA INTERNA E
CIÊNCIAS DA SAÚDE - 40001016012P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação MEDICINA INTERNA E CIÊNCIAS DA SAÚDE da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **FLÁVIA SPRENGER** intitulada: **"SEGMENTAÇÃO SEMIAUTOMÁTICA E CORRELAÇÃO COM ACHADOS CIRÚRGICOS E HISTOPATOLÓGICOS EM GLIOMAS DE ALTO GRAU."**, sob orientação do Prof. Dr. BERNARDO CORRÊA DE ALMEIDA TEIXEIRA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

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

Curitiba, 09 de Fevereiro de 2023.

Assinatura Eletrônica
09/02/2023 11:06:05.0
BERNARDO CORRÊA DE ALMEIDA TEIXEIRA
Presidente da Banca Examinadora

Assinatura Eletrônica
09/02/2023 15:13:21.0
DANTE LUIZ ESCUISSATO
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
09/02/2023 17:05:55.0
LEANDRO TAVARES LUCATO
Avaliador Externo (FACULDADE DE MEDICINA DA UNIVERSIDADE DE SÃO PAULO)

AGRADECIMENTOS

Aos professores David Menotti, do Departamento de Informática da Universidade Federal do Paraná, professores Alexandre Xavier Falcão e Samuel Botter Martins do Laboratório de Informática e *Data Science* da Universidade de Campinas pela concessão do *software* e colaboração na realização deste trabalho.

À Clínica CETAC pela concessão do espaço físico, recursos humanos e *softwares* para o processamento das amostras.

Ao Instituto de Neurologia de Curitiba, em especial ao serviço de neurocirurgia, pela concessão dos registros cirúrgicos para a seleção da amostra.

À Dra Marcella Cavalcanti do Neopath Curitiba pelo auxílio com os dados histológicos e imunohistoquímicos da amostra.

À equipe do SOPHiA Genetics™ pelo apoio ao expandir os horizontes desta pesquisa.

Ao Center for *Biomedical Image Computing & Analysis* da *Perelman School of Medicine - University of Pennsylvania* pela parceria internacional e oportunidade de contribuição em estudo multicêntrico.

RESUMO

Os gliomas malignos são as neoplasias cerebrais primárias mais prevalentes. A imagem pré-operatória desempenha um papel importante no manejo clínico e cirúrgico, e seus achados guiam as subsequentes intervenções. Nosso objetivo foi correlacionar a volumetria tumoral por segmentação semiautomática em imagens de ressonância magnética (RM) pré-operatória com parâmetros clínicos e histopatológicos em uma coorte submetida à ressecção cirúrgica assistida por 5-ALA. Foram incluídos 137 pacientes com glioblastomas e astrocitomas grau 4 com RM pré-operatória disponível, submetidos a cirurgia entre 2015 e 2021. Achados clínicos, cirúrgicos e histopatológicos também foram obtidos (sexo, idade, tipo de ressecção, fluorescência 5-ALA residual, diagnóstico histológico, IDH, MGMT, ATRX, Ki67 e sobrevida). As imagens pré-operatórias de RM foram pré-processadas e segmentadas no VISVA para tumor inteiro (WT) em 3D FLAIR, realce (ET) e necrose (NC) em 3D pós-gadolínio T1. Os volumes de edema (ED) foram obtidos pela subtração de ET e NC de WT. A amostra final foi de 91 pacientes. Não houve correlação entre a volumetria tumoral e tipo de ressecção ou sobrevida. Maiores índices Ki-67 correlacionaram-se positivamente com maiores volumes de WT ($p = 0,008$), ET ($p = 0,03$) e ED ($p < 0,001$), sem correlação com volumes de NC ($p = 0,96$). Os índices Ki67 também foram maiores em pacientes com fluorescência 5-ALA residual no leito operatório ($p = 0,01$; *cutoff* de 20%). Não houve correlação entre a volumetria tumoral, tipo de ressecção e parâmetros de sobrevida. Encontramos associação positiva entre os índices de Ki67, os volumes tumorais metabolicamente ativos (WT, ET, ED) e fluorescência residual ao 5-ALA no leito operatório.

Palavras-chave: segmentação, volumetria, Ki67, glioma, 5-ALA

ABSTRACT

Malignant gliomas are the most prevalent primary brain neoplasms. Preoperative imaging plays an important role in clinical and surgical management, and its findings guide further interventions. Our goal was to correlate tumor volumetry in semiautomated segmentation on preoperative magnetic resonance imaging with clinical and histopathological findings in a 5-ALA-assisted resection cohort. We initially included 137 patients with glioblastomas and grade 4 astrocytomas with available preoperative imaging that underwent surgery between 2015 and 2021. Clinical, surgical and histomolecular findings were also obtained (sex, age, type of resection, residual 5-ALA fluorescence, histological diagnosis, IDH, MGMT, ATRX, Ki67 and survival). Preoperative MRI studies were preprocessed and segmented in VISVA for whole tumor (WT) in 3D FLAIR, enhancement (ET) and necrosis (NC) in 3D post-gadolinium T1. Edema volumes (ED) were obtained by subtracting ET and NC from WT. The final sample consisted of 91 patients. There was no correlation between tumor volumetry and type of resection or survival. Higher Ki-67 scores were positively correlated with higher volumes of WT ($p = 0.008$), ET ($p = 0.03$) and ED ($p < 0.001$), with no correlation with NC volumes ($p = 0.96$). Ki67 scores were also higher in 5-ALA compromised margins ($p = 0.01$). No correlation was observed between tumor volume, type of resection and survival parameters. We found a positive association between Ki67 indexes, metabolically active tumor volumes (WT, ET, ED) and residual 5-ALA fluorescence.

Keywords: segmentation, volumetry, Ki67, glioma, 5-ALA

LISTA DE ILUSTRAÇÕES

GRÁFICO 1 – PREVALÊNCIA DOS GLIOMAS	13
FIGURA 1 – RESUMO DAS CLASSIFICAÇÕES DE GLIOMAS DA OMS	15
FIGURA 2 – CLASSIFICAÇÃO DE 2021 DA OMS DOS GLIOMAS	16
FIGURA 3 – EXEMPLO DE GLIOBLASTOMA À TC	23
FIGURA 4 – ASPECTO DE IMAGEM DO GLIOBLASTOMA	24
FIGURA 5 – ASPECTO PÓS-CONTRASTE DO GLIOBLASTOMA.....	25
FIGURA 6 – EXEMPLO DE ESPECTROSCOPIA	26
FIGURA 7 – EXEMPLO DE TRACTOGRAFIA.....	27
FIGURA 8 – EXEMPLO DE PERFUSÃO T2*	28
FIGURA 9 – ASPECTO MACROSCÓPICO DO GLIOBLASTOMA.....	29
FIGURA 10 – PEÇA CIRÚRGICA DE RESSECÇÃO DE GLIOBLASTOMA.....	29
FIGURA 11 – ASPECTO MICROSCÓPICO DO GLIOBLASTOMA	30
FIGURA 12 – EXEMPLO DE KI67 BAIXO E ALTO	31
FIGURA 13 – IMAGEM INTRAOPERATÓRIA COM 5-ALA	34
FIGURA 14 – DIAGRAMA DAS ESPECIALIDADES DA IA	36
FIGURA 15 – REPRESENTAÇÃO GRÁFICA DAS REDES NEURAI E SUAS CAMADAS.....	36
FIGURA 16 – REPRESENTAÇÃO DO <i>GRAPH CUT</i>	42
FIGURA 17 – CORREÇÃO VIA DIFT NO VISVA.....	44
QUADRO 1 – DESCRITORES UTILIZADOS NA REVISÃO BIBLIOGRÁFICA.....	46
FIGURA 18 – PRÉ-PROCESSAMENTO DAS IMAGENS.....	53
FIGURA 19 – EXEMPLO DE SEGMENTAÇÃO NO VISVA.....	54
FIGURA 20 – AVALIAÇÃO DE MARGENS POR RM E 5-ALA	55
FIGURA 21 – AMOSTRA FINAL	57
GRÁFICO 2 – DIAGNÓSTICOS DOS PACIENTES EXCLUÍDOS	58
FIGURA 22 – DISTRIBUIÇÃO POR LOBOS DAS LESÕES.....	59
GRÁFICO 4 – TIPO DE ABORDAGEM CIRÚRGICA	61
GRÁFICO 5 – CURVA DE SOBREVIVÊNCIA POR TIPO DE RESSECÇÃO CIRÚRGICA	62
GRÁFICO 7 – CURVA DE SOBREVIVÊNCIA E WT	67
GRÁFICO 8 – CURVA DE SOBREVIVÊNCIA E ET	67
GRÁFICO 9 – CURVA DE SOBREVIVÊNCIA E NC	68

GRÁFICO 10 – CURVA DE SOBREVIVÊNCIA E ED68

LISTA DE TABELAS

TABELA 1 – RESUMO DAS PRINCIPAIS MUDANÇAS NAS CLASSIFICAÇÕES DA OMS	15
TABELA 2 – GRADUAÇÃO DOS GLIOMAS DIFUSOS DO ADULTO	17
TABELA 3 – RESUMO DAS TÉCNICAS DE SEGMENTAÇÃO	43
TABELA 4 – RESUMO DOS ARTIGOS PUBLICADOS SOBRE SEGMENTAÇÃO 3D EM GLIOMAS	47
TABELA 5 – REVISÃO DA LITERATURA EM PROGNÓSTICO E VOLUMETRIA ...	49
TABELA 6 – REVISÃO DA LITERATURA SOBRE CORRELAÇÃO KI67 E VOLUMETRIA	50
TABELA 7 – PARÂMETROS DE AQUISIÇÃO DAS IMAGENS DE RM	52
TABELA 8 – DISTRIBUIÇÃO DA AMOSTRA POR IDADE	58
TABELA 9 – NÚMERO DE EXAMES	59
TABELA 10 – CARACTERÍSTICAS DE IMAGEM EM SEQUÊNCIAS AVANÇADAS	59
TABELA 11 – VOLUMETRIAS TUMORAIS EM MM ³	60
TABELA 12 – CORRELAÇÃO ENTRE VOLUMETRIAS	60
TABELA 13 – TIPO DE RESSECÇÃO <i>VERSUS</i> PROGNÓSTICO E VOLUMETRIA TUMORAL	61
TABELA 14 – MARGENS CIRÚRGICAS 5-ALA E SOBREVIDA EM MESES	62
TABELA 15 – MARGENS CIRÚRGICAS 5-ALA E VOLUMETRIA EM MM ³	63
TABELA 16 – PERFIL IMUNOHISTOQUÍMICO	63
TABELA 17 – IDH <i>VERSUS</i> PROGNÓSTICO E VOLUMETRIA	64
TABELA 18 – MGMT <i>VERSUS</i> PROGNÓSTICO E VOLUMETRIA	64
TABELA 19 – ATRX <i>VERSUS</i> PROGNÓSTICO E VOLUMETRIA	65
TABELA 20 – KI-67 <i>VERSUS</i> VOLUMETRIA	65
TABELA 21 – KI-67 E MARGENS 5-ALA	65
TABELA 22 – SOBREVIDA EM MESES	66
TABELA 23 – VOLUMETRIA <i>VERSUS</i> PROGNÓSTICO	66
TABELA 24 – REVISÃO DA LITERATURA EM PROGNÓSTICO E VOLUMETRIA ..	72

LISTA DE SIGLAS

DL	-	<i>Deep learning</i>
ED	-	Edema
ET	-	<i>Enhancing tumor</i>
FL	-	<i>Federated learning</i>
GPU	-	Graphics processing unit
IA	-	Inteligência artificial
IDH	-	Isocitrato desidrogenase
ML	-	<i>Machine learning</i>
MRI	-	<i>Magnetic resonance imaging</i>
NC	-	<i>Necrotic core</i>
OMS	-	Organização Mundial da Saúde
RM	-	Ressonância magnética
SNC	-	Sistema nervoso central
SWI	-	<i>Susceptibility weighted imaging</i>
TC	-	Tomografia computadorizada
VISVA	-	<i>Volumetric Image Segmentation for Visualization and Analysis</i>
WT	-	<i>Whole tumor</i>

SUMÁRIO

1 INTRODUÇÃO	12
2 REVISÃO DA LITERATURA	13
2.1 NEOPLASIAS PRIMÁRIAS DO SISTEMA NERVOSO CENTRAL	13
2.1.1 Introdução	13
2.1.2 Classificação	14
2.1.3 Gliomas de alto grau	17
2.1.3.1 Epidemiologia e fatores de risco	18
2.1.3.2 Fisiopatologia	19
2.1.3.3 Manifestações clínicas	22
2.1.3.4 Avaliação por imagem	22
2.1.3.5 Diagnóstico histopatológico e molecular	29
2.1.3.6 Modalidades terapêuticas e prognóstico	32
2.2 INTELIGÊNCIA ARTIFICIAL	35
2.2.1 Aplicações em neuro-oncologia	38
2.2.2 Segmentação	40
2.2.2.1 Histórico da segmentação em neuro-oncologia	44
2.2.2.2 Estado da arte em segmentação e gliomas	46
3 MATERIAIS E MÉTODOS	51
3.1 DESENHO DO ESTUDO	51
3.2 OBJETIVOS.....	51
3.3 CRITÉRIOS DE INCLUSÃO	51
3.4 CRITÉRIOS DE EXCLUSÃO	51
3.5 MÉTODOS.....	52
3.5.1 Imagens.....	52
3.5.2 Dados clínicos, cirúrgicos e histomoleculares	54
3.6 ANÁLISE ESTATÍSTICA.....	56
3.7 PARECER ÉTICO	56
4 RESULTADOS.....	57
4.1 IMAGEM E VOLUMETRIA	58
4.2 ASPECTOS CIRÚRGICOS E TERAPÊUTICOS	60
4.3 DIAGNÓSTICO HISTOLÓGICO	63
4.4 PROGNÓSTICO	66

4.5 ARTIGO PUBLICADO.....	69
4.6 APRESENTAÇÃO DOS RESULTADOS PARCIAIS	69
4.7 ARTIGO SUBMETIDO	69
5 DISCUSSÃO	70
CONCLUSÃO	79
REFERÊNCIAS.....	80
ANEXO I – APROVAÇÃO DO COMITÊ DE ÉTICA	90
ANEXO II – ARTIGO PUBLICADO	91
ANEXO III – APRESENTAÇÃO DOS RESULTADOS PARCIAIS	108
ANEXO IV – CONFIRMAÇÃO DE SUBMISSÃO	109
ANEXO V – ARTIGO SUBMETIDO (TEXTO NA ÍNTEGRA)	110

1 INTRODUÇÃO

Os gliomas malignos são as neoplasias primárias cerebrais mais prevalentes, com uma incidência estimada entre dois a quatro casos por 100.000 pessoas. Apesar dos avanços recentes e busca incessante por terapias-alvo, os gliomas de alto grau continuam tendo prognóstico sombrio, com uma sobrevida média de 35% após o primeiro ano do diagnóstico (BONDY et al., 2008; OMURO; DEANGELIS, 2013; THAKKAR et al., 2014; TONN; WESTPHAL; RUTKA, 2010).

Segundo dados do INCA, são estimados 810 novos casos de neoplasias do sistema nervoso central no estado do Paraná e de 90 novos casos apenas em Curitiba por ano (INSTITUTO NACIONAL DO CÂNCER INCA, 2019).

A avaliação por imagem tem um papel fundamental no diagnóstico e manejo dos glioblastomas. Características de imagem à ressonância magnética (RM) têm valor prognóstico, mas o fator mais relacionado ao aumento da sobrevida é a ressecção cirúrgica máxima com preservação funcional (HANIF et al., 2017; OMURO; DEANGELIS, 2013; THAKKAR et al., 2014). Cirurgias guiadas por fluorescência 5-ALA têm sido uma ferramenta valiosa para alcançar ressecção macroscópica total de gliomas de alto grau, facilitando a visualização da lesão e de suas margens (KLINIK et al., 2006a; STUMMER et al., 2011).

Diversos estudos almejaram correlacionar a volumetria tumoral com taxas de sobrevida e aspectos histomoleculares, porém, a grande maioria com métodos inconsistentes e de difícil reprodutibilidade (ANDREOU et al., 1983a; GRABOWSKI et al., 2014; HAMMOUD et al., 1996a; ILIADIS et al., 2012; LACROIX et al., 2001; LEVIN et al., 1980a; WANGARYATTAWANICH et al., 2015; WOOD; GREEN; SHAPIRO, 1988a; XUE; ALBRIGHT, 1999a; ZINN et al., 2011).

Sendo assim, o presente estudo tem como objetivo a correlação dos volumes das subáreas tumorais a partir de segmentação semiautomática com aspectos clínico-prognósticos (sobrevida) e histológicos (IDH, MGMT, ATRX e Ki67), bem como com parâmetros cirúrgicos, tendo como diferencial se tratar de uma coorte submetida à ressecção com auxílio do 5-ALA.

Iniciamos o texto por uma extensa revisão da literatura sobre gliomas, inteligência artificial e métodos computacionais de segmentação. Em seguida abordamos a metodologia do estudo e os resultados, seguido da discussão com achados da literatura e conclusões do estudo.

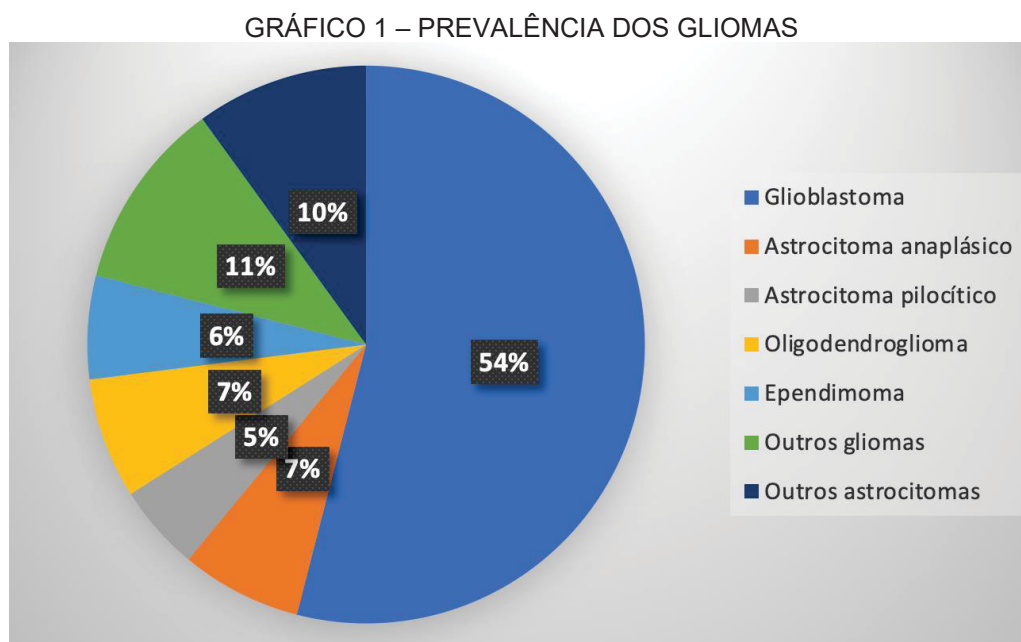
2 REVISÃO DA LITERATURA

2.1 NEOPLASIAS PRIMÁRIAS DO SISTEMA NERVOSO CENTRAL

2.1.1 Introdução

As metástases representam metade dos tumores do SNC. A outra metade, composta pelos tumores primários, são um grupo amplamente heterogêneo de lesões, com distribuição variável nos diversos compartimentos anatômicos, faixas etárias e tipos histológicos (OSBORN, 2012).

Os gliomas são as neoplasias malignas primárias mais prevalentes do SNC, e consistem em um conjunto complexo de lesões originadas nas células da glia, constituída por astrócitos, oligodendrócitos, células do epêndima e do plexo coroide (GRÁFICO 1) (BONDY et al., 2008; TONN; WESTPHAL; RUTKA, 2010).



FONTE: Adaptado de Agnihotri (2012).

NOTA: Foi utilizada a classificação da OMS de 2016 devido a escassez de dados epidemiológicos da nova classificação.

Os gliomas são divididos em linhagens astrocítica e não-astrocítica. Acreditava-se que os astrocitomas emergiam diretamente das células estreladas, os astrócitos. Sabe-se hoje, que não têm origem em astrócitos normais que sofreram degeneração maligna, mas sim, de células tronco pluripotentes com capacidade de

diferenciação entre fenótipos neuronais, astrocíticos e oligodendrogliais (OSBORN, 2012).

Sendo assim, o estudo do genoma humano revolucionou o entendimento das neoplasias do SNC, postulando que a maioria das malignidades primárias do SNC emergem de células tronco doentes, e não que partem exclusivamente da linhagem celular à qual pertencem. Desse modo, as características genotípicas devem ser somadas ao fenótipo das lesões, permitindo sua adequada classificação (OSBORN, 2012).

Neste contexto entra em cena a classificação da Organização Mundial da Saúde (OMS). Como o escopo do presente estudo são os gliomas de alto grau, estes serão o foco da revisão.

2.1.2 Classificação

A ausência de uma classificação uníssona compreendida e aceita por neurologistas, neurocirurgiões, neuropatologistas e radiologistas impedia a adequada investigação etiológica e prognóstica dos gliomas (ZÜLCH; KÖLN, 1980).

Datam da década de 1950 as primeiras iniciativas para a elaboração de uma normatização internacionalmente aceita, para a nomenclatura e classificação das neoplasias do SNC (SCHEITHAUER, 2009).

Em 1979 foi publicada a primeira edição da Classificação da OMS dos tumores do SNC nos consagrados livros azuis, fiando-se principalmente em características histomorfológicas das lesões, com escassos recursos imunohistoquímicos e moleculares (SCHEITHAUER, 2009; ZÜLCH; KÖLN, 1980).

Dado o grande sucesso e aceitação da primeira edição, novos livros azuis sucederam, com pequenas mudanças centradas principalmente em parâmetros histológicos, até a quarta revisão em 2016 (KLEIHUES; BURGER; SCHEITHAUER, 1993; LOUIS et al., 2007, 2016).

A partir desse momento, houve quebra de paradigma, e passou-se a caminhar em direção ao melhor entendimento molecular e genético dessas condições, permitindo avaliação prognóstica com maior distinção entre as diferentes categorias, previamente agrupadas (BRITO et al., 2019).

Tais mudanças foram então paulatinamente incorporadas nas classificações subsequentes, chegando em sua quinta edição no ano de 2021 (TABELA 1, FIGURA 1) (LOUIS et al., 2021).

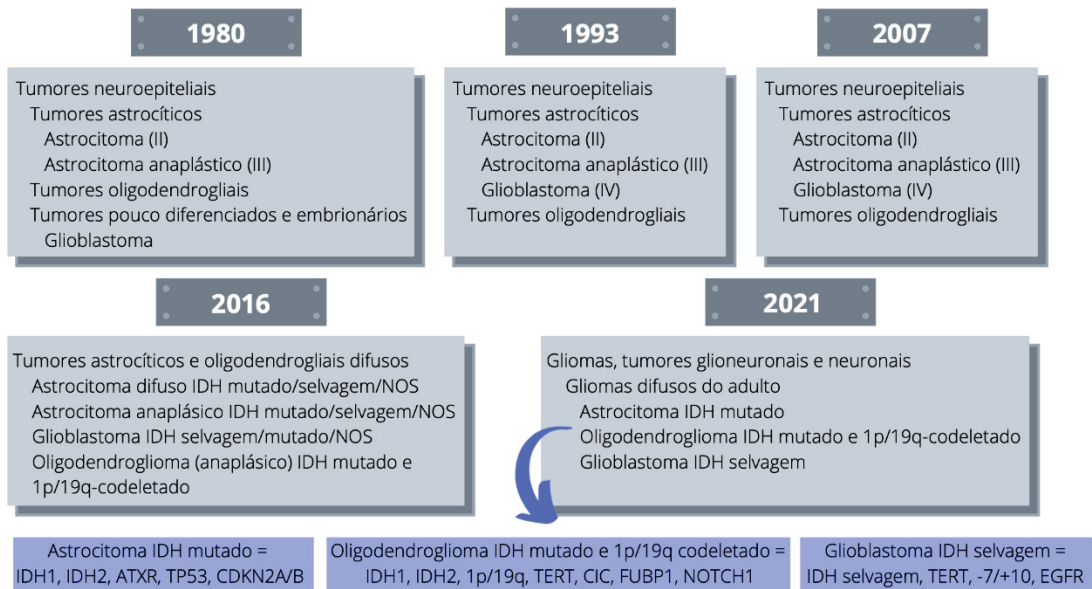
TABELA 1 – RESUMO DAS PRINCIPAIS MUDANÇAS NAS CLASSIFICAÇÕES DA OMS

Ano (Edição)	Principais mudanças incorporadas
1993 (2ª)	Glioblastoma passa a ser incluído nos astrocitomas e retirado do grupo de tumores pouco diferenciados e embrionários
2000 (3ª)	Sem mudanças significativas no tocante aos glioblastomas
2007 (4ª)	Sem mudanças significativas no tocante aos glioblastomas
2016 (4ª – revisão)	Importante reestruturação dos gliomas difusos, incluindo características genéticas Classificação dos astrocitomas em IDH mutado ou selvagem Inclusão de glioblastoma IDH selvagem e mutado como entidades
2021 (5ª)	A graduação deve ser informada em algarismos arábicos Astrocitomas de menor grau IDH selvagem +7/-10 ou amplificação EGFR ou mutação TERT devem ser classificados como glioblastomas Astrocitomas IDH mutados com deleção CDKN2AB devem ser classificados como astrocitomas grau 4 Glioblastomas devem obrigatoriamente possuir IDH selvagem Classificações que utilizam em seu nome localização, faixa etária ou padrões histogenéticos foram retirados (ex: mixoide, xantomatoso, anaplásico)

FONTE: Adaptado de Scheithauer *et al* (2009).

Nota: OMS – Organização Mundial da Saúde.

FIGURA 1 – RESUMO DAS CLASSIFICAÇÕES DE GLIOMAS DA OMS

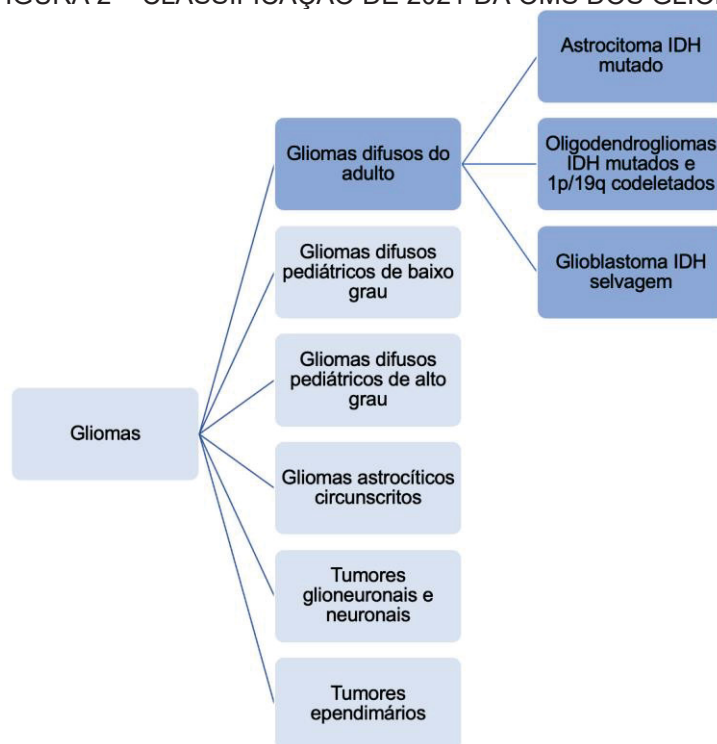


FONTE: A autora (2022).

Nota: OMS – Organização Mundial da Saúde; NOS – *not otherwise specified*.

Atualmente, os gliomas, tumores glioneuronais e neuronais passam a ser divididos em seis grandes grupos, representados na figura abaixo (FIGURA 2) (LOUIS et al., 2021).

FIGURA 2 – CLASSIFICAÇÃO DE 2021 DA OMS DOS GLIOMAS



FONTE: A autora (2022).

Nota: OMS – Organização Mundial da Saúde.

O grupo dos gliomas difusos do adulto, foco do presente estudo, é composto pelos astrocitomas, oligodendrogliomas e glioblastomas, cada um com sua assinatura genética definidora de classe (FIGURAS 1 e 2) (LOUIS et al., 2021).

Os astrocitomas são gliomas astrocíticos difusos, com mutação do gene IDH1 ou IDH2, com graus que variam entre 2 à 4 (TABELA 2). Destaca-se que astrocitomas grau 2 ou 3 com amplificação do gene EGFR, mutação TERT ou +7/-10 passam a ser considerados glioblastomas. A presença de mutação CDKN2AB também determina *upgrade* para astrocitoma grau 4 (LOUIS et al., 2021).

Na presente classificação os glioblastomas são obrigatoriamente gliomas astrocíticos difusos grau 4, com IDH do tipo selvagem, que apresentem proliferação microvascular ou amplificação do gene EGFR, mutação TERT ou +7/-10 (LOUIS et al., 2021).

TABELA 2 – GRADUAÇÃO DOS GLIOMAS DIFUSOS DO ADULTO

Glioma	Graus
Astrocitoma IDH mutado	2, 3 e 4
Oligodendroglioma IDH mutado e 1p/19q codeletado	2 e 3
Glioblastoma IDH selvagem	4

FONTE: A autora (2021).

Nota: OMS – Organização Mundial da Saúde.

Quando o status IDH é desconhecido, os tumores devem ser nomeados astrocitomas difusos não especificado ou *not otherwise specified (NOS)*. Em contrapartida, quando o perfil molecular é conhecido, mas não permite adequação à nenhuma das categorias conhecidas, deve-se referir à lesão como *not elsewhere classified (NEC)* (LOUIS et al., 2021).

2.1.3 Gliomas de alto grau

Rudolf Virchow foi o primeiro pesquisador a descrever as células da glia do SNC, atuantes como neurocimento e participando da homeostase extracelular. Quase uma década depois, em 1865, foi o mesmo a utilizar o termo glioma pela primeira vez ao relatar neoplasias malignas com origem nas células gliais. Também foi o primeiro a propor uma classificação dos gliomas em lesões de baixo grau de acordo com a celularidade (AGNIHOTRI et al., 2013; STOYANOV; DZHENKOV, 2018).

Devido a alta agressividade e indiferenciação desse tipo de lesão, mais de cinquenta anos depois, Bailey e Cushing questionaram a origem glial dos glioblastomas, devido a ausência de semelhanças com astrócitos normais (AGNIHOTRI et al., 2013; STOYANOV; DZHENKOV, 2018).

Tal questionamento foi refutado pelo neuropatologista alemão Scherer, que determinou que o diagnóstico deve ser feito com base no total da amostra e não apenas em grupos celulares isolados. Sendo assim, determinou-se que o glioblastoma e as demais lesões astrocíticas têm precursores comuns (AGNIHOTRI et al., 2013; STOYANOV; DZHENKOV, 2018).

A partir da segunda metade do século 20, com a maior disponibilidade de métodos imunohistoquímicos e posteriormente moleculares, a definição e classificação dos gliomas foi revolucionada. Tais mudanças foram paulatinamente incorporadas, até as classificações atuais, e atualmente estes parâmetros regem o

diagnóstico e manejo dos gliomas (AGNIHOTRI et al., 2013; STOYANOV; DZHENKOV, 2018).

2.1.3.1 Epidemiologia e fatores de risco

Os gliomas malignos são as neoplasias cerebrais primárias mais prevalentes, estimando-se uma incidência de aproximadamente que varia entre dois e quatro casos por 100.000 pessoas, com cerca de 20.000 novos diagnósticos anuais nos Estados Unidos, apresentando uma discreta predileção pelo sexo masculino, sendo 1,6 vezes mais frequente em homens (OMURO; DEANGELIS, 2013; THAKKAR et al., 2014).

Os glioblastomas representam cerca de 83% dos gliomas malignos, mais frequentes em caucasianos, com pico de incidência entre os 55 e 85 anos (OMURO; DEANGELIS, 2013; THAKKAR et al., 2014).

A maioria é de ocorrência esporádica, contudo, existe associação com neurofibromatose tipo 1 e 2, síndromes de Li-Fraumeni e Turcot, Cowden, esclerose tuberosa e schwannomatose familiar (OSBORN, 2012).

Quando existe história familiar positiva, o risco é cerca de duas vezes maior para o desenvolvimento de glioma. O principal fator de risco ambiental sabidamente estabelecido é a radiação ionizante, especialmente crianças que receberam radioterapia com irradiação do crânio. Ínfimas doses de radiação usadas para o tratamento de *Tinea capitis* e hemangiomas também são capazes de elevar o risco em até três vezes (HANIF et al., 2017; OMURO; DEANGELIS, 2013).

Existe relação inversa com desordens atópicas, como asma, rinite e eczema. Sabe-se que não são fatores de risco para o desenvolvimento de gliomas malignos o uso de telefones celulares e outros campos eletromagnéticos, trauma cerebral, alimentos contendo N- nitroso, como em embutidos ou defumados, aspartame, riscos ocupacionais, exposição à pesticidas (OMURO; DEANGELIS, 2013).

Medidas preventivas baseadas em estilo de vida não são capazes de reduzir a incidência dos gliomas. Ademais, o diagnóstico precoce e tratamento ainda não demonstraram melhorar o desfecho, sendo assim, não se justifica rastreio populacional (OMURO; DEANGELIS, 2013).

Quanto ao local acometido, a imensa maioria afeta o sistema nervoso supratentorial, com predileção pelo lobo frontal. As lesões infratentoriais são raras,

representando menos de três por cento dos glioblastomas, e ocorrendo em idade menor, com pico de incidência entre 50 e 56 anos (THAKKAR et al., 2014).

2.1.3.2 Fisiopatologia

Os gliomas malignos emergem de desordens genéticas sequenciais e cumulativas, decorrentes de diferentes fatores intrínsecos e ambientais (OMURO; DEANGELIS, 2013).

Cada mutação conhecida possui fisiopatologia bem estabelecida para culminar com a gliomagenese.

IDH diz respeito a um grupo de enzimas com três isoformas (IDH1, IDH2 e IDH3) responsáveis por etapas metabólicas importantes no ciclo de Krebs e no metabolismo da glutamina. Essas enzimas através da afinidade de seu sítio catalítico pelo isocitrato, convertem NADP^+ em NADPH, resultando na formação do alfa-cetoglutarato, em um ambiente normal, sendo processo fundamental da homeostase celular (HAN et al., 2020).

Mutações IDH são descritas em diversos tipos de neoplasia além dos gliomas, como colangiocarcinomas, condrossarcomas, carcinomas de tireoide leucemias. Quando mutadas, estas enzimas consomem NADPH para a formação de NADP^+ , suprimindo a detoxificação das moléculas de peróxido de hidrogênio, e culminando com o acúmulo de D-2-hidroxiglutarato, um oncometabólito importante da gliomagenese. Na imensa maioria dos casos, a mutação ocorre na enzima IDH1, e sua ocorrência determina melhor prognóstico em relação aos gliomas IDH selvagens (HAN et al., 2020; THAKKAR et al., 2014).

Sabendo-se das nuances do mau funcionamento do complexo IDH no desenvolvimento de neoplasias, é foco de extensa pesquisa a possibilidade de terapia-alvo, havendo representantes aprovados pela agência regulamentadora estadunidense, a *Food and Drug Administration* (FDA), como o ivosidenib e o vorasidenib, aprovados para o tratamento de leucemias. Caminha-se em grandes passos para a aprovação dessas drogas no manejo de gliomas (HAN et al., 2020).

ATRX refere-se a proteínas que existem em duas isoformas, codificadas pelo gene ATRX, descoberto no estudo de pacientes com retardo mental ligado ao cromossomo X e alfa-talassemia. Essas proteínas são fundamentais na manutenção da estabilidade genômica, através da regulação da transcrição, remodelamento da

cromatina e deposição das histonas nos telômeros (NANDAKUMAR; MANSOURI; DAS, 2017).

Em contraste com a grande clareza do papel do IDH na gliomagenese, o papel molecular do gene ATRX nos gliomas ainda é incerto. Sabe-se que as mutações cursam com a inativação deste gene, através de deleções, fusões ou ambos, consumando uma instabilidade do genoma e cursando com pior prognóstico (NANDAKUMAR; MANSOURI; DAS, 2017).

Localizado no cromossomo 17, o gene TP53 abriga o código para a produção da proteína p53, grande maestra da regulação da proliferação celular, integridade genômica, determinando fim do ciclo celular e apoptose em células imperfeitas (GUO et al., 2020; ZHANG et al., 2018).

Mutações do gene TP53 estão presentes nos mais diversos tipos de câncer, não sendo diferente para os gliomas, neste caso, associadas à fenótipos mais invasivos e proliferativos, com pouca apoptose e maior progressão. Outro aspecto que contribui para o pior prognóstico nestes casos é a resistência à quimioterápicos danificadores do DNA, como a cisplatina (GUO et al., 2020; ZHANG et al., 2018).

A deleção homozigótica do *locus* CDKN2A é o componente mais frequentemente danificado na via p53 (ZHANG et al., 2018).

Os genes CDKN2A/B constituem as quinases dependentes de ciclina, divididas em CDKN2A E CDKN2B, responsáveis por inibir a degradação da proteína p53. Em células onde os genes CDKN2A/B são homozigoticamente deletados há déficit na produção de proteínas supressoras de tumor, admitindo a ligação entre ciclinas e suas quinases, e consequentemente ativação de vias pró-mitóticas descontroladas (APPAY et al., 2019; LU et al., 2020).

Sabe-se que o prognóstico é tenebroso na presença dessa mutação, e por esse motivo, a última classificação da OMS determina *upgrade* para grau 4 nos astrocitomas IDH e CDKN2 mutados, independente da morfologia histológica (APPAY et al., 2019; LOUIS et al., 2021).

Até 2016 os glioblastomas podiam ser classificados de acordo com o *status* IDH, onde as lesões IDH mutadas eram consideradas secundárias, evoluindo de astrocitomas de menor grau, com melhor prognóstico, e os glioblastomas IDH selvagens, considerados primários (LOUIS et al., 2016).

Atualmente, os glioblastomas são obrigatoriamente gliomas de alto grau com IDH selvagem. Outros marcos moleculares dessa condição são as mutações TERT, +7/-10 e EGFR (LOUIS et al., 2021).

TERT é a porção catalítica da enzima telomerase, responsável pela manutenção do comprimento dos telômeros, constituídos de repetições de ácidos nucleicos que não codificam proteínas, nas extremidades dos cromossomos. Mutações do complexo TERT ocorrem em diversos tipos de câncer, como melanomas, carcinomas uroteliais e gliomas, cuja carcinogênese ocorre através da amplificação TERT, que permite a imortalização de células tumorais, também conferindo pior prognóstico e menor sobrevida nestes pacientes (YUAN et al., 2016).

Ganhos no cromossomo 7, no *locus* do EGFR, associado a perdas genéticas no cromossomo dez, no *locus* do gene de supressão tumoral conhecido como fosfatase e tensina (PTEN) também são características marcantes dos glioblastomas. Deleções nesse cromossomo, seja completa ou de apenas um dos braços estão associadas a pior prognóstico (DEL MAR INDA et al., 2003).

O EGFR é uma tirosinoquinase transmembrana do cromossomo 7, que no glioblastoma promove a invasibilidade tumoral, divisão celular e resistência a radioquimioterapia. Pode encontrar-se amplificado através de mutações, sendo a variante III (vIII) a mais comumente encontrada, ou ainda, com atividade aumentada através da inibição ou deleção de vias regulatórias. A amplificação EGFR é presente em cerca de 40% dos glioblastomas, e a vIII está associada a pior prognóstico (THAKKAR et al., 2014).

A MGMT é uma enzima de reparo do DNA que remove agentes alquilantes da posição O6 da guanina. Sua metilação torna a enzima silenciada, aumentando a sensibilidade à quimioterapia com temozolamida. Em contrapartida, um promotor não metilado desta enzima promove sua superexpressão e resistência à quimioterapia (THAKKAR et al., 2014).

O promotor MGMT é metilado em 50% dos casos de astrocitomas grau 4, notando-se maior sobrevida livre de progressão e sobrevida geral (HEGI et al., 2002; KATSIKIANNIS et al., 2021).

Estudos do genoma identificaram ainda diversos outros genes adicionais relacionados à doença, como o RTEL (20q13.33), CDKN2BAS (9p21.3), CCDC26 (8q24.21) e PHLDB1 (11q23.3), embora cada um destes com associação fraca,

provavelmente refletindo a natureza multimolecular da lesão (OMURO; DEANGELIS, 2013).

2.1.3.3 Manifestações clínicas

O sintoma mais prevalente dos pacientes com gliomas é a cefaleia, presente em cerca de 50% dos casos. O desafio maior reside na suspeição de etiologia maligna, visto que a dor geralmente é inespecífica. Deve-se atentar quando a cefaleia é progressivamente pior, unilateral ou de instalação recente em pacientes com mais de 50 anos (OMURO e DEANGELIS, 2013).

Papiledema pode estar presente nos casos em que há aumento da pressão intracraniana, mais raro atualmente devido ao maior acesso ao diagnóstico por imagem (OMURO e DEANGELIS, 2013).

Sintomas cognitivos, mudanças da personalidade e distúrbios da fala devem ser cuidadosamente avaliados em idosos, evitando presunção do diagnóstico de demência ou *dellirium* (OMURO e DEANGELIS, 2013).

Outros sintomas podem estar presentes, como distúrbios da marcha e incontinência, geralmente em lesões com efeito de massa exuberante. Os déficits neurológicos focais também devem ser cuidadosamente avaliados, evitando-se confusão com infartos (OMURO e DEANGELIS, 2013).

Por fim, crises convulsivas também são prevalentes, acometendo 20 a 40% dos pacientes, ressaltando-se as aquelas com origem focal (OMURO e DEANGELIS, 2013).

2.1.3.4 Avaliação por imagem

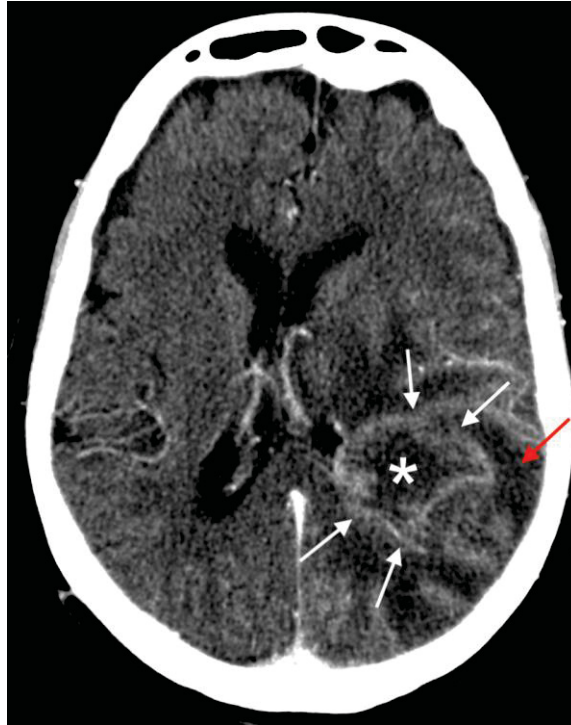
Os gliomas de alto grau acometem a substância branca subcortical e profunda dos hemisférios cerebrais, comumente estendendo-se ao longo dos tratos de fibras compactas do corpo caloso e corticoespinhais. O tamanho é variável, e devido a rápida disseminação pela substância branca até 20% dos tumores são multifocais ao diagnóstico. Vale ressaltar, que até cinco por cento dos glioblastomas multifocais são verdadeiramente sincrônicos, e não decorrentes de disseminação (OSBORN, 2012).

A tomografia computadorizada (TC) geralmente consiste no método inicial. Os glioblastomas manifestam-se como uma massa heterogênea a hipodensa circundada

por halo iso a hiperdenso, relacionado ao componente sólido, com efeito de massa e edema peritumoral.

Hemorragias são frequentes, e calcificações não são esperadas. Nas fases pós-contraste observa-se realce periférico irregular, com vasos proeminentes adjacentes à lesão (FIGURA 3) (OSBORN, 2012).

FIGURA 3 – EXEMPLO DE GLIOBLASTOMA À TC



FONTE: A autora (2022).

LEGENDA: Imagem de TC axial pós-contraste de um glioblastoma. Observa-se lesão expansiva parietal esquerda (setas brancas), com importante efeito de massa, realce irregular e periférico, com centro hipodenuante (asterisco), inferindo necrose. Há também halo hipodenuo de edema peritumoral (seta vermelha).

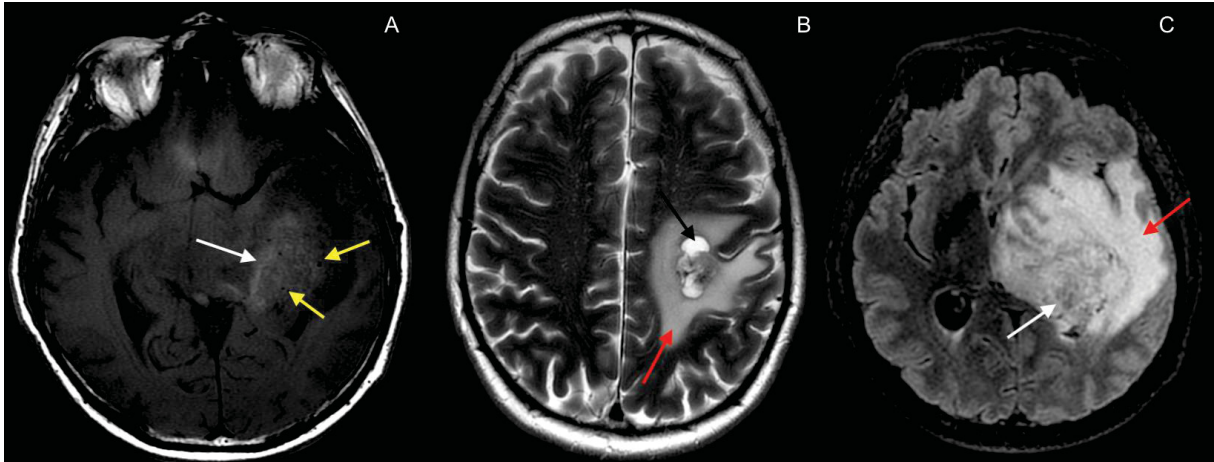
NOTA: TC – Tomografia computadorizada.

À RM os glioblastomas se apresentam como massas heterogêneas e mal delimitadas. Nas sequências ponderadas em T1 podem ser observadas áreas de hipersinal, traduzindo focos de hemorragia subaguda (FIGURA 4A). Dada a vascularização abundante, podem ser identificados também artefatos de fluxo ou *flow voids* (FIGURA 4A) (OSBORN, 2012; SHUKLA et al., 2017).

Nas sequências sensíveis a líquido, como T2 e FLAIR as lesões são difusamente heterogêneas, com áreas centrais de hipersinal traduzindo necrose ou degeneração cística (FIGURA 4B). Estas sequências demonstram ainda normalmente extenso hipersinal perilesional, relacionado a edema vasogênico, ou eventualmente, invasão microscópica. É fundamental avaliação meticulosa destas áreas, visto que

mais de 90% das recorrências ocorrem dentro de seus limites. Isso se deve ao aumento da permeabilidade vascular provocada pelo edema associada a particularidades infiltrativa do tumor (FIGURA 4C) (OSBORN, 2012; SHUKLA et al., 2017).

FIGURA 4 – ASPECTO DE IMAGEM DO GLIOBLASTOMA



FONTE: A autora (2022).

LEGENDA: Exemplo de glioblastoma à RM. A: Axial T1 pré-contraste evidenciando focos de hipersinal espontâneo na lesão, relacionado a hemorragia (seta branca) e *flow voids* (setas amarelas) em seu interior, relacionados a vascularização proeminente. B: Axial T2 demonstrando áreas de degeneração cística e necrose (seta preta) e de edema perilesional (seta vermelha). C: Axial T2 FLAIR demonstrando lesão heterogênea e predominantemente hiperintensa (seta branca), com extenso edema perilesional (seta vermelha).

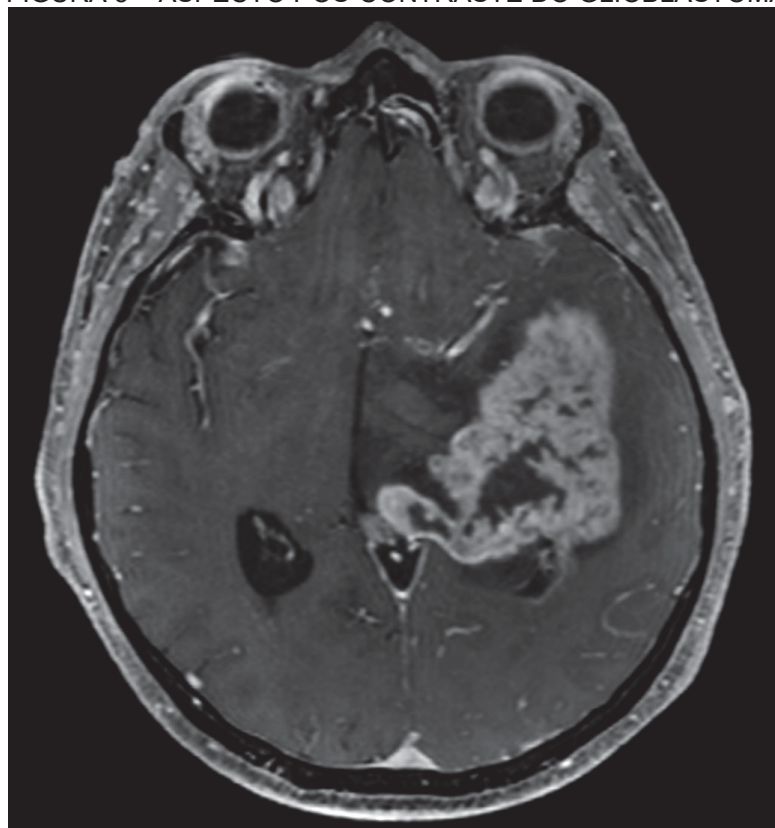
NOTA: RM – ressonância magnética.

Hipersinal em DWI (*diffusion weighted imaging*) associado a baixo sinal em ADC (*apparent diffusion coefficient*) pode ser observado no componente sólido realçante, traduzindo hiper celularidade (OSBORN, 2012; SHUKLA et al., 2017).

Nas imagens ponderadas em susceptibilidade magnética, como T2* *gradiente echo* ou SWI, pontos de marcado hipossinal demonstram produtos de degradação da hemoglobina relacionados a hemorragia intratumoral (OSBORN, 2012).

Nas sequências pós-gadolínio nota-se um realce heterogêneo, predominantemente periférico, grosseiro e irregular, circundando um centro hiporrealçante, representativo de necrose (FIGURA 5) (OSBORN, 2012).

FIGURA 5 – ASPECTO PÓS-CONTRASTE DO GLIOBLASTOMA



FONTE: A autora (2022).

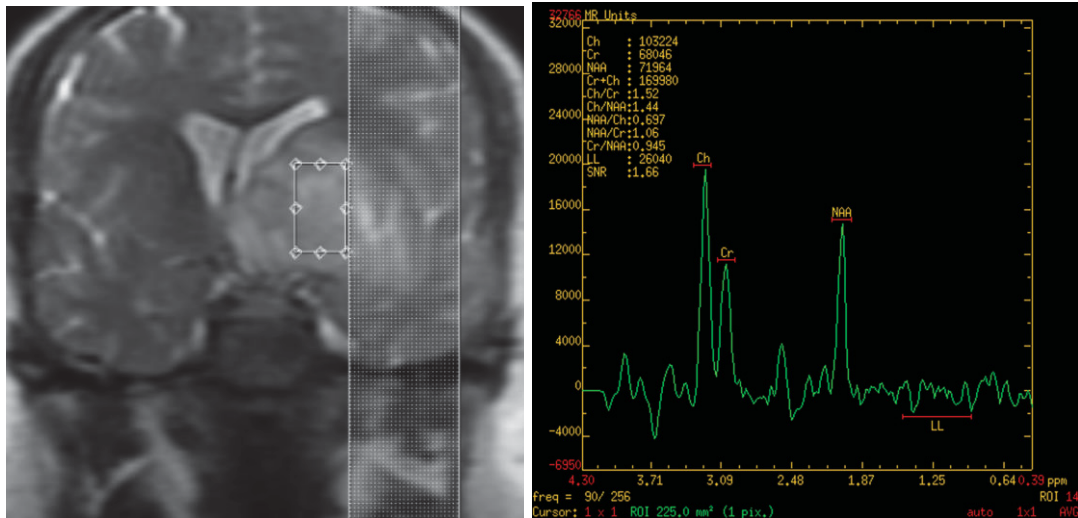
LEGENDA: Imagem de RM axial T1 pós-gadólíneo. Nota-se realce irregular e periférico pelo agente de contraste, um centro com ausência de realce, refletindo necrose, e sinais de invasão da superfície ependimária do ventrículo lateral esquerdo.

NOTA: RM – ressonância magnética.

Além do protocolo convencional, é exponencial o crescimento das sequências avançadas. São elas a espectroscopia de prótons, sequências dinâmicas pós-contraste, como perfusão e permeabilidade, tractografia e RM funcional (SHUKLA et al., 2017).

A espectroscopia de prótons é uma técnica da RM que permite a identificação e concentração de diferentes metabólitos em um tecido. Os glioblastomas exibem elevação do pico da colina, denotando proliferação celular, com redução do N-acetilaspártato (NAA), traduzindo despopulação neuronal, além de aumento do pico de lipídio e lactato, refletindo anaerobiose. Pode-se observar também redução do pico de mioinositol, ou com ferramentas avançadas de pós-processamento, um pico de 2-hidroxi-glutarato (FIGURA 6) (OSBORN, 2012; ÖZ et al., 2014; SHUKLA et al., 2017).

FIGURA 6 – EXEMPLO DE ESPECTROSCOPIA



FONTE: A autora (2021).

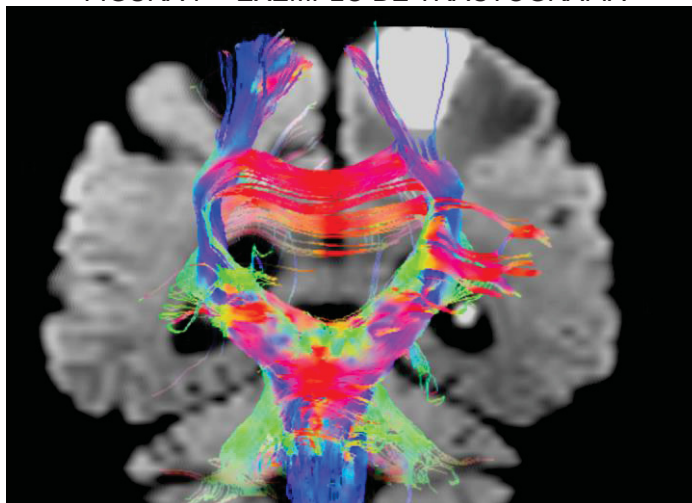
LEGENDA: Exemplo de espectroscopia de prótons single voxel de glioblastoma, com voxel de amostragem à esquerda. À direita nota-se aumento do pico de colina, com aumento das relações colina/creatina e colina/NAA.

NOTA: ppm – partes por milhão. Ch – colina. Cr – creatina. NAA – N-acetil-aspartato. LL – lipídios e lactato.

RM funcional é um método utilizado para visualização de atividade cortical a partir de alterações sutis do fluxo sanguíneo em resposta a estímulos pré-determinados. No contexto dos gliomas, permite planejamento cirúrgico detalhado, evitando ressecção de áreas eloquentes, mas também permitindo abordagens mais agressivas através da identificação dos limites de áreas importantes (SHUKLA et al., 2017).

Diffusion tensor imaging (DTI) é um recurso de RM empregado para estimar a organização axonal do cérebro a partir da difusão da anisotropia. Seu mapa pode ser reconstruído tridimensionalmente em tractografia. Seu emprego em neuro-oncologia gera ricas imagens que guiam a terapia cirúrgica, mas também tem papel no pós-operatório, auxiliando na identificação de tumor residual (FIGURA 7) (SHUKLA et al., 2017).

FIGURA 7 – EXEMPLO DE TRACTOGRAFIA



FONTE: A autora (2023).

LEGENDA: Reconstrução da tractografia evidencia a íntima relação do trato corticoespinal esquerdo (fibras roxas) com a massa parietal esquerda.

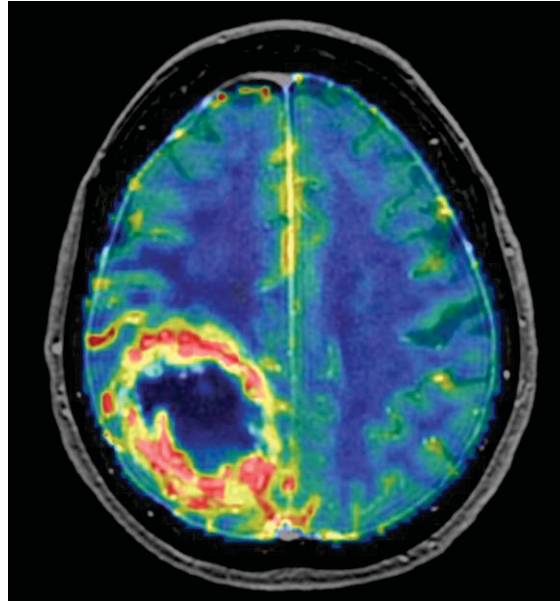
As imagens dinâmicas pós-contraste expõem a farmacocinética da captação de contraste, achado este, relacionado com o prognóstico e também associação com subtipos moleculares e mutações através de métodos computacionais avançados (SHUKLA et al., 2017).

O estudo perfusional pode ser realizado através da avaliação do encurtamento T1 provocado pelo gadolínio ou da queda de sinal provocada pela susceptibilidade magnética em imagens em gradiente (ESSIG et al., 2013).

A primeira, conhecida como permeabilidade ou perfusão T1 avalia a permeabilidade capilar através do coeficiente *k-trans* que tem como princípio o encurtamento do tempo T1 provocado pela chegada do gadolínio num tecido de interesse (ESSIG et al., 2013).

Por sua vez, a perfusão T2* utiliza imagens ponderadas em gradiente para estimar a neoangiogênese através do volume cerebral relativo numa área de interesse a partir da queda de sinal provocada pela susceptibilidade magnética do contraste utilizado (FIGURA 8). Esta, possui íntima relação com o grau tumoral e prognóstico, tendo papel fundamental no diagnóstico dos gliomas de alto grau (ESSIG et al., 2013).

FIGURA 8 – EXEMPLO DE PERFUSÃO T2*



FONTE: A autora (2021).

LEGENDA: Mapa de cores da perfusão em fusão com T1 pós-gadolinio. Há aumento do rCBV no mapa de cores na periferia da lesão parietal direita, correspondendo ao componente sólido.

NOTA: rCBV – *Relative cerebral blood volume*.

Os glioblastomas possuem diversas vias de disseminação, podendo ocorrer através da substância branca, do líquido cefalorraquidiano (LCR), do epêndima, por invasão direta da dura mater e crânio, e, por fim, através de metástases hematogênicas a distância, especialmente para medula óssea, fígado e pulmão (OSBORN, 2012).

O diagnóstico diferencial nos diferentes métodos de imagem são vastos. O maior diferencial é com metástases, que geralmente acometem a transição entre substância branca e cinzenta, com aspecto menos infiltrativo em comparação com o glioblastoma (OSBORN, 2012).

Outros tumores incluem o astrocitoma anaplásico, que geralmente não exhibe realce pelo agente de contraste paramagnético e o oligodendroglioma anaplásico, cujo diferencial pode ser difícil por imagem. O linfoma primário do sistema nervoso central também é um diagnóstico diferencial importante, embora, raramente seja necrótico em pacientes sem HIV (OSBORN, 2012).

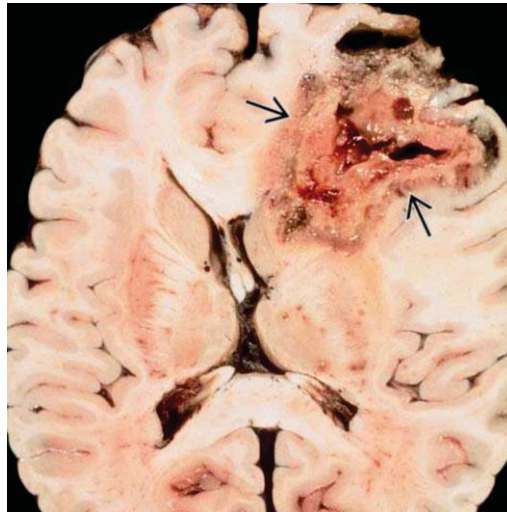
Além das neoplasias, deve-se sempre afastar a possibilidade de infarto, buscando-se relação com territórios vasculares, abscessos, estes, com restrição à difusão das moléculas da água central e as doenças desmielinizantes tumefativas,

que assumem aspecto em ferradura na substância branca subcortical (OSBORN, 2012).

2.1.3.5 Diagnóstico histopatológico e molecular

Os gliomas de alto grau, sobretudo o glioblastoma, macroscopicamente consistem em lesões heterogêneas, com áreas sólidas, contendo hemorragia e necrose, geralmente centradas na substância branca (FIGURAS 9 e 10) (HANIF et al., 2017; OSBORN, 2012).

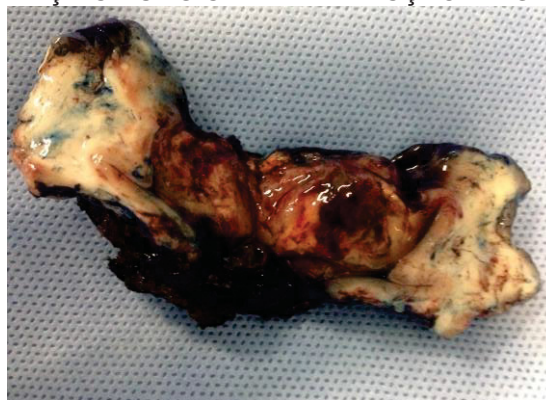
FIGURA 9 – ASPECTO MACROSCÓPICO DO GLIOBLASTOMA



FONTE: Adaptado de Osborn (2018).

LEGENDA: Produto de autópsia de paciente com glioblastoma primário. Nota-se o tumor heterogêneo (setas pretas), circundando o centro hemorrágico e necrótico.

FIGURA 10 – PEÇA CIRÚRGICA DE RESSECÇÃO DE GLIOBLASTOMA



FONTE: A autora (2023).

LEGENDA: Peça macroscópica de ressecção de glioblastoma, demonstrando lesão heterogênea, com áreas de hemorragia e necrose.

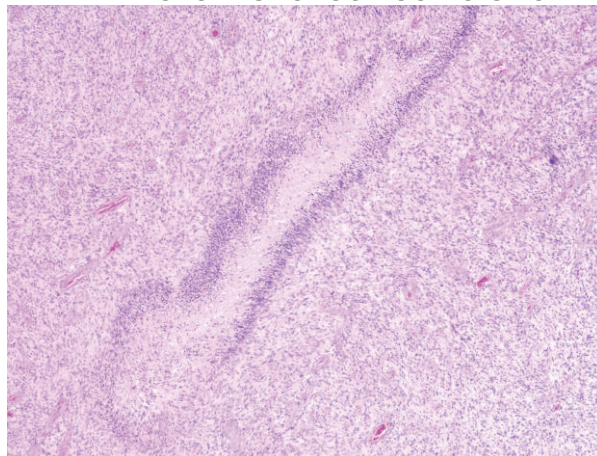
Os gliomas difusos possuem padrão de crescimento à histologia extremamente particular, raro nos demais tipos de neoplasia, caracterizado por infiltração difusa do parênquima, formando uma rede ao longo do neurópilo e agregando-se ao redor de neurônios, vasos sanguíneos e dos revestimentos piais, com fácil extensão ao longo dos feixes de substância branca (PERRY; WESSELING, 2016).

A matriz tumoral pode variar desde áreas paucicelulares com focos de degeneração cística e gliose, à áreas hipercelulares com proliferação microvascular e necrose (PERRY; WESSELING, 2016).

A classificação parte inicialmente da comparação das células tumorais com as demais células normais. Os astrocitomas são caracterizados por núcleos irregulares e hipercromasia, associado a um fundo fibrilar que se mescla ao citoplasma, conferindo aspecto de núcleo despido. O grande pleomorfismo do microambiente histológico e celular do glioblastoma conferiu inicialmente o nome de glioblastoma multiforme, atualmente em desuso (PERRY; WESSELING, 2016).

Os gliomas difusos contendo aumento da atividade mitótica, proliferação microvascular glomeruloide e necrose, especialmente quando cercadas por células tumorais (pseudo-paliçadas) agrupadas, são classificados como grau 4. Quando há apenas aumento da atividade mitótica, atribui-se grau 3 (antigo astrocitoma anaplásico). Neoplasias grau 2 não possuem nenhuma dessas características supradescritas (FIGURA 11) (PERRY; WESSELING, 2016; RONG et al., 2006).

FIGURA 11 – ASPECTO MICROSCÓPICO DO GLIOBLASTOMA



FONTE: A autora (2023).

LEGENDA: Imagem de microscopia óptica de glioblastoma corada com hematoxilina-eosina. Nota-se acúmulo de células neoplásicas ao redor de áreas de necrose e de células apoptóticas. A este aspecto se dá o nome de pseudo-paliçadas.

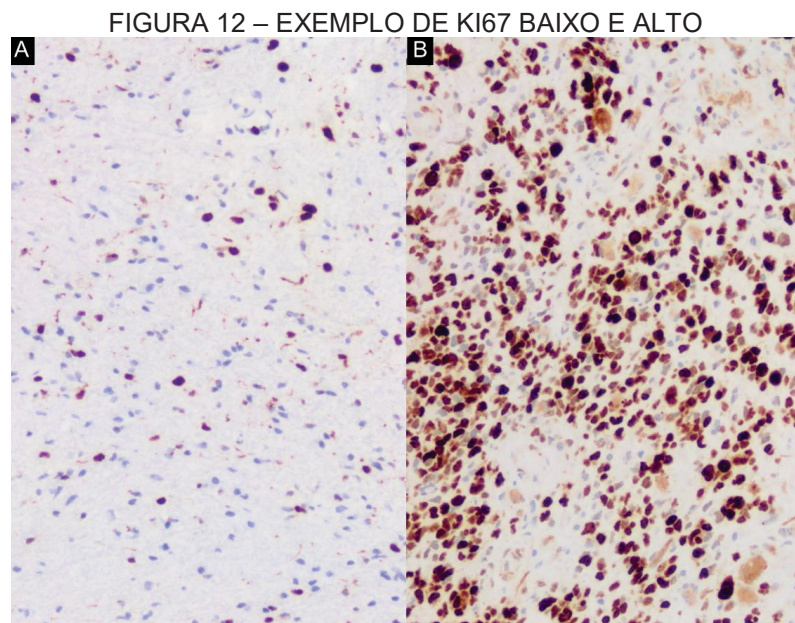
A avaliação imunohistoquímica é fundamental para o diagnóstico patológico. Geralmente inicia-se pelo anticorpo GFAP (*glial fibrillary acidic protein*), uma proteína do citoesqueleto das células gliais, fortemente positivo nos astrocitomas (PERRY; WESSELING, 2016).

A vimetina é uma proteína citoplasmática, expressada mais precocemente na astrogênese, e portanto, pode ser positiva em astrocitomas pobremente diferenciados e de alto grau (PERRY; WESSELING, 2016).

Outros marcadores utilizados são a citoqueratina, para excluir a possibilidade de metástase de linhagem epitelial, e S-100, outra proteína positiva em células gliais (PERRY; WESSELING, 2016).

Existem diversas técnicas histológicas para determinar a atividade mitótica do tecido estudado. A mais simples e acessível consiste na coloração hematoxilina e eosina e a contagem manual das mitoses no campo de visão. Tem como principal limitação avaliar apenas as mitoses em fase M (PERRY; WESSELING, 2016).

Outra técnica consiste na marcação imunohistoquímica de Ki67 que também estima a mitose, sendo capaz de avaliar as fases G1, S, G2 e M, aumentando, portanto, a sensibilidade (FIGURA 12) (PERRY; WESSELING, 2016).



FONTE: A autora (2023).

LEGENDA: Exemplo de Ki67 (imunohistoquímica, aumento de 200x) de aproximadamente 5-10% das células tumorais (A) e de 80% (B).

Também são avaliados por imunohistoquímica o *status* p53, EFGR IDH e BRAF (PERRY; WESSELING, 2016).

Apesar de importante no diagnóstico dos gliomas, a imunohistoquímica possui uma série de limitações, especialmente a falta de especificidade dos marcadores para os tipos de neoplasia e a necessidade de múltiplos marcadores, encarecendo o método. Além disso, não existem diretrizes padronizando as técnicas a serem empregadas, havendo grande variabilidade entre os serviços assistentes (NANDAKUMAR; MANSOURI; DAS, 2017; PERRY; WESSELING, 2016).

Sendo assim, deve ser sempre utilizada no contexto clínico completo, em associação com achados de imagem e cirúrgicos, para a adequada definição etiológica, terapêutica e prognóstica (PERRY; WESSELING, 2016).

O diagnóstico das demais mutações, como CDKN2A/B, TERT e +7/-10 ampara-se nas técnicas de biologia molecular e isolamento do DNA, com sequenciamento genético e reação da polimerase em cadeia em tempo real (rtPCR) (WOOD; HALFPENNY; MOORE, 2019).

2.1.3.6 Modalidades terapêuticas e prognóstico

Gliomas de alto grau são neoplasias agressivas, incuráveis, e de prognóstico ruim, com sobrevida média de 15 meses após o diagnóstico. A sobrevida após o primeiro ano é de 35%, de 13,7% para o segundo ano e de menos de cinco por cento em cinco anos. O pico de mortalidade ocorre no quinto trimestre após o diagnóstico. Pacientes que sobrevivem dois anos após o diagnóstico tendem a ter maior sobrevida geral. Em pacientes que não recebem nenhuma modalidade terapêutica a sobrevida é de aproximadamente três meses (HANIF et al., 2017; THAKKAR et al., 2014).

O tratamento sintomático inclui o uso de corticoesteroides, visando alívio dos sintomas neurológicos causados pelo edema perilesional, e antiepiléticos em pacientes com crises convulsivas. O agente de escolha é o levetiracetam, mas outros agentes como topiramato, lamotrigina e ácido valproico podem ser empregados. Deve-se evitar anticonvulsivantes indutores de enzimas hepáticas como fenitoína, carbamazepina e fenobarbital por reduzirem a efetividade de alguns quimioterápicos (OMURO; DEANGELIS, 2013).

Diante da suspeita de glioma maligno o paciente deve ser referido à centro terciário com perspectiva de ressecção cirúrgica, visando alívio do efeito de massa, citorredução e obtenção de amostra significativa para caracterização histomolecular.

Tumores inoperáveis devem ser amostrados por biópsia para adequada classificação (OMURO; DEANGELIS, 2013).

O tratamento cirúrgico de gliomas envolve uma série de etapas meticulosamente planejadas e executadas por uma equipe médica especializada. O procedimento começa com uma craniotomia, na qual uma porção do crânio é removida para acessar o cérebro.

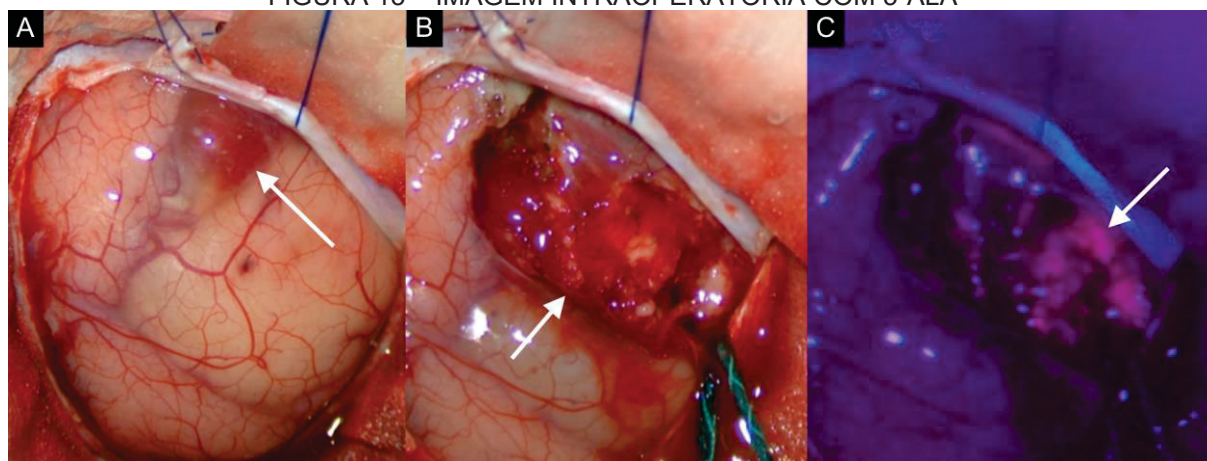
Durante a cirurgia, são utilizados equipamentos avançados, como microscópios cirúrgicos e neuroendoscópios, para ampliar a visão e permitir uma remoção precisa do tumor. Além disso, o mapeamento cerebral intraoperatório é frequentemente realizado para identificar áreas funcionais críticas e preservá-las durante a remoção do glioma. O cirurgião trabalha cuidadosamente para remover o máximo de tecido tumoral possível, respeitando as estruturas vitais do cérebro.

O neuronavegador é uma ferramenta tecnológica avançada utilizada na neurocirurgia para auxiliar os cirurgiões na navegação precisa e segura durante os procedimentos. Ele combina informações detalhadas de imagens pré-operatórias, como ressonância magnética e tomografia computadorizada, com informações em tempo real do campo cirúrgico. Com o uso de técnicas de rastreamento eletromagnético ou óptico, o sistema de neuronavegação fornece ao cirurgião uma representação tridimensional do cérebro do paciente, permitindo a identificação precisa de estruturas anatômicas, como tumores, vasos sanguíneos e áreas funcionais críticas. Durante a cirurgia, o cirurgião pode visualizar em tempo real sua posição em relação ao alvo cirúrgico e ajustar a trajetória e a profundidade dos instrumentos com base nas informações fornecidas pelo neuronavegador. Isso ajuda a melhorar a precisão, a segurança e os resultados dos procedimentos neurocirúrgicos, permitindo uma abordagem mais precisa, minimizando o risco de danos a estruturas cerebrais saudáveis e auxiliando na ressecção tumoral máxima.

O microscópio em neurocirurgia é essencial para ampliação, iluminação e orientação precisas. Permite visualizar detalhes minuciosos, especialmente em cirurgias complexas. Com iluminação intensa, facilita a visualização mesmo em áreas de difícil acesso. A integração com sistemas de neuronavegação permite orientação precisa em tempo real. É crucial em procedimentos delicados, minimizando danos a estruturas cerebrais saudáveis. Na remoção de tumores, auxilia na visualização das margens tumorais, melhorando a ressecção e reduzindo a recorrência. Em resumo, o microscópio melhora a precisão, segurança e resultados na neurocirurgia.

O ácido 5-aminolevulínico (ALA-5) é um metabólito natural da via da hemoglobina, que pode se acumular em tecidos com permeabilidade vascular aumentada, como nos gliomas e em suas adjacências. Usualmente, é convertido em protoporfirina IX, contudo, a expressão reduzida da enzima ferroquelatase em células neoplásicas promove seu acúmulo, que pode ser detectado com excitação sob luz azul, permitindo melhor determinação intraoperatória da área a ser ressecada. A intensidade da fluorescência também pode ajudar a prever o grau de celularidade. Sendo assim, cirurgias com emprego do 5-ALA auxiliam ressecções mais completas, aumentando a sobrevida quando comparadas às ressecções microscópicas com luz branca convencional (FIGURA 13) (COBURGER; WIRTZ, 2019; HADJIPANAYIS; STUMMER, 2019; HERVEY-JUMPER; BERGER, 2016; KLINIK et al., 2006a; MAHMOUDI et al., 2019; PICART et al., 2019; RAMINA et al., 2016).

FIGURA 13 – IMAGEM INTRAOPERATÓRIA COM 5-ALA



FONTE: A autora (2023).

LEGENDA: Imagens intraoperatórias de ressecção de lesão glial demonstrando em A o aspecto macroscópico com cisto periférico (seta branca), em B a porção sólida da lesão (seta branca) e em C a forte fluorescência 5-ALA à luz azul do componente sólido do tumor (seta branca).

Terapia adjuvante com quimiorradioterapia deve ser considerada para todos os pacientes após a cirurgia. Desde sua publicação em 2005, o protocolo Stupp é o de escolha para a adjuvância, consistindo numa dose total de radioterapia de 60 Gy realizado de segunda a sexta-feira por seis semanas com 2Gy diariamente, associado a temozolamida na dose de 75 mg/m² de superfície corporal todos os dias da semana durante a radioterapia. Após o término da radioterapia realizam-se seis ciclos com dose de 150-200 mg/m² de superfície corporal por cinco dias durante cada ciclo de 28 dias (STUPP et al., 2005)

Pacientes com metilação da MGMT (O6-metilguanina-DNA-metiltransferase) têm resposta superior à temozolomida. Polímeros contendo o alquilante carmustina também podem ser depositados no leito cirúrgico (OMURO; DEANGELIS, 2013; THAKKAR et al., 2014).

O bevacizumab (Avastin®), um anticorpo monoclonal antifator de crescimento vascular endotelial (VEGF), está em estudo para emprego na primeira linha de tratamento, com resultados promissores até o presente momento (OMURO; DEANGELIS, 2013)

Praticamente todos os tumores progridem após um período livre de progressão, que vai de sete a dez meses e tristemente nenhuma das terapias disponíveis é capaz de melhorar a sobrevida como um todo, com benefícios limitados em casos seletos. Dentre os recursos, pode-se lançar mão de nova cirurgia para citorredução e alívio do efeito de massa além de atualização do *status* molecular e histológico. A quimioterapia de resgate inclui os agentes supracitados, além de outros alquilantes como nitrosureias e carboplatina (OMURO; DEANGELIS, 2013).

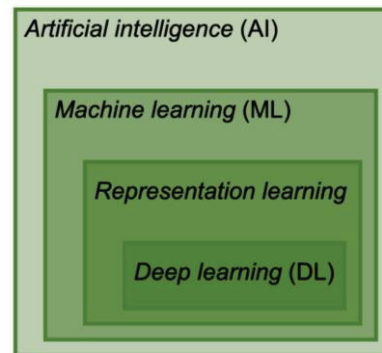
2.2 INTELIGÊNCIA ARTIFICIAL

Inteligência artificial é o ramo da ciência computacional dedicada ao desenvolvimento de sistemas que normalmente necessitariam da inteligência humana para sua solução (CHARTRAND et al., 2017).

A ciência da IA não é recente. As primeiras descrições datam da década de 50, e o principal limitante para tamanho atraso na sua propagação foi a falta de computadores capazes de processar este tipo de rede, além da baixa disponibilidade dados (CHARTRAND et al., 2017).

A IA possui diversas subespecialidades. A mais abrangente delas, o aprendizado de máquinas ou *machine learning* (ML), desenvolve algoritmos que desempenham tarefas a partir do aprendizado de padrões, ao invés de determiná-los na programação, com suas subespecialidades *representation learning* (RL) e *deep learning* (DL) (FIGURA 14) (CHARTRAND et al., 2017).

FIGURA 14 – DIAGRAMA DAS ESPECIALIDADES DA IA



FONTE: Adaptado de Chartrand (2017).

NOTA: IA – Inteligência artificial

O cerne do DL reside no desenvolvimento de redes neurais. De maneira limitada, a atividade de um neurônio pode parecer relativamente simples, mas se torna progressivamente complexas quando se leva em conta as diversas vias e conexões entre os bilhões de neurônios do cérebro humano. Funcionam, de maneira semelhante, as redes neurais, onde múltiplos nós neurais artificiais estão interligados, produzindo um desfecho isolado, mas baseado numa somatória de evidências providenciada pelo restante da rede neural (CHARTRAND et al., 2017).

O aspecto profundo das redes neurais diz respeito às camadas situadas entre a camada de entrada, ou *input*, e a saída, ou *output*. O empilhamento de camadas de neurônios permite estabelecer uma hierarquia de características, que se tornam progressivamente mais complexas. Foge à compreensão, inclusive dos cientistas da computação os processos que ocorrem nas camadas profundas, e que produzem os desfechos desejados. Este fenômeno é conhecido como a caixa preta das redes neurais (FIGURA 15) (CHARTRAND et al., 2017).

FIGURA 15 – REPRESENTAÇÃO GRÁFICA DAS REDES NEURAI E SUAS CAMADAS



FONTE: Adaptado de Chartrand (2017).

Uma vez criada, a rede neural é treinada com o ajuste dos parâmetros de peso e viés de cada nó. Numa rede moderna, podem existir milhões destes parâmetros para ajuste. Sendo assim, parte-se de uma configuração inicial aleatória e segue-se com a otimização do algoritmo a partir do gradiente descendente (CHARTRAND et al., 2017).

A cada passagem dos dados na rede, é estabelecida a função de perda, que consiste na inacurácia do método, ou seja, o quão distante ele encontra-se do valor de regressão. Com base nessa função, os parâmetros são reajustados, e os dados processados diversas vezes, na direção de minimizar a função de perda e de maximizar a acurácia do modelo (CHARTRAND et al., 2017).

Após o desenvolvimento da rede neural, é necessário treiná-la. Para isso, duas abordagens são possíveis – a supervisionada e a não supervisionada (SOTOUDEH et al., 2019).

Na metodologia supervisionada, o modelo é treinado com imagens demarcadas e tenta encontrar tais características em imagens cruas. Emprega como principais técnicas para o seu desenvolvimento regressão linear e logística, *support vector machines* (SVMs), *random forest* e Bayes (SOTOUDEH et al., 2019).

Na metodologia não supervisionada o modelo tenta encontrar características sem treinamento prévio com imagens segmentadas através de técnicas como propagação de afinidade, mapas auto-organizantes, *K-means*, e *clustering* hierárquico (SOTOUDEH et al., 2019).

Na prática, métodos totalmente não supervisionados ainda não apresentam resultados satisfatórios, e por isso, têm crescido a importância do método semi-supervisionado, consistindo na combinação do uso de informações etiquetadas e não etiquetadas (CHARTRAND et al., 2017).

O tamanho da amostra de treinamento para produzir resultado satisfatório depende de essencialmente da natureza e da complexidade da tarefa. Reside nesse ponto a maior limitação do método, pois a disponibilidade de imagens de boa qualidade segmentadas e etiquetadas é limitada (SOTOUDEH et al., 2019).

Uma alternativa para aumentar a amostra é o aumento artificial através de pequenas transformações das imagens que dispensam nova demarcação. Isso pode ser obtido através de inversão, rotação, translação, zoom e deformação elástica (CHARTRAND et al., 2017).

Por fim, a performance do modelo pode ser aferida de diversas formas. É possível fazer análise bioestatística convencional com sensibilidade, especificidade e acurácia, bem como empregar métodos mais complexos, como a curva ROC. Para a comparação entre segmentações feitas pelo operador e pelo algoritmo, pode-se lançar mão do escore Dice. É importante destacar, que quanto maior a complexidade do algoritmo, mais complexa é a avaliação de sua performance (SOTOUDEH et al., 2019).

Diversos softwares são disponíveis para o desenvolvimento de redes neurais, a maioria destes abertos ao público e gratuitos. Dentre estes estão Theano, Torch, TensorFlow, CNTK, Caffe e Keras. O compartilhamento de algoritmos e *datasets* tem permitido o avanço acelerado do ML (CHARTRAND et al., 2017).

2.2.1 Aplicações em neuro-oncologia

No que concerne os gliomas, a IA tem ganhado espaço em todas as esferas do manejo, desde as etapas pré-tratamento, com o diagnóstico, localização, segmentação, genômica, planejamento cirúrgico e demais características extraíveis da imagem, bem como no pós-tratamento, no diagnóstico de progressão de doença ou recorrência (SOTOUDEH et al., 2019).

A volumetria dos diversos componentes da neoplasia, como edema, realce e necrose representa informação vital para o planejamento cirúrgico, mas a segmentação manual é uma técnica demasiadamente trabalhosa. Visando suprir essa necessidade, a IA também entra em campo, com diversos tipos de algoritmos de segmentação dos componentes do tumor, com grande qualidade, sobretudo nas redes neurais convolucionais (LOTAN et al., 2019; SOTOUDEH et al., 2019).

Modelos semelhantes também têm sido empregados na avaliação por imagem intraoperatória, dado o comportamento altamente infiltrativo desses graus de neoplasia, guiando a real interface entre parênquima normal e tecido encefálico infiltrado (SOTOUDEH et al., 2019).

Classicamente, os neurorradiologistas reconhecem padrões à imagem de RM que são capazes de inferir com certo grau de confiabilidade o grau tumoral, como a presença de realce, necrose e comportamento perfusional. Algoritmos de IA podem extrair além destas, outras características de imagem imperceptíveis aos olhos humanos, ciência à qual se dá o nome de *radiomics* (SOTOUDEH et al., 2019).

A maioria das ferramentas de *radiomics* atualmente são consideradas limitadas, por extrair informações valiosas, mas com baixa capacidade de inferir o desfecho desejado. Desse modo, a maioria dos pesquisadores utilizam a extração das características radiômicas dos tumores como parte do algoritmo, que com a adição de camadas se torna mais preciso em prever mutações. Contudo, nenhuma ferramenta possui ainda validação para aplicação clínica (SOTOUDEH et al., 2019).

A IA também tem sido empregada na avaliação histológica dos gliomas, com o advento do escaneamento de lâminas. São descritas aplicações no diagnóstico histológico, detecção de tipos celulares e predição do prognóstico, com bons resultados. Tais redes neurais prometem auxiliar e facilitar o trabalho do patologista, delimitando áreas de interesse ou realizando a contagem de mitoses por campo, por exemplo (SOTOUDEH et al., 2019).

IA também visa impactar positivamente o planejamento da radioterapia, delimitando o tecido de interesse e determinando as isodoses em estruturas adjacentes relevantes (SOTOUDEH et al., 2019).

Apesar dos grandes avanços no campo da radiologia, com espectroscopia, perfusão e PET, permanece sendo desafio à toda a equipe assistente a diferenciação entre pseudoprogressão, relacionada a alterações pós-tratamento, como radionecrose, e progressão de doença. Entra em campo, então, outra aplicação da IA, ainda precoce e com poucos estudos de amostras limitadas (SOTOUDEH et al., 2019).

Por fim, e de grande valia é a determinação do prognóstico do paciente, classicamente feito a partir da união de aspectos de imagem, epidemiológicos e histomoleculares. Diversos algoritmos de IA têm sido desenvolvidos nesse sentido, utilizando desde imagens convencionais de RM à sequências avançadas, como rCBV, DTI e fs-MRI (SOTOUDEH et al., 2019).

Muitos desafios limitam a implantação em larga escala da IA atualmente. O primeiro e maior deles sendo a escassez de dados padronizados e uniformes, visto que cada serviço possui protocolo específico de diagnóstico, imagem e manejo, limitando a uniformização de uma grande população para a realização de *trials* de validação (JIN et al., 2020; SOTOUDEH et al., 2019).

Junto à escassez de dados, residem os aspectos ético e legais de compartilhamento de dados de pacientes entre serviços, uma vez que vinculadas às imagens existem informações pessoais do paciente, além da possibilidade de

reconstrução tridimensional dos *voxels* da face, quebrando o sigilo (SOTOUDEH et al., 2019).

Outro grande limitante é a dificuldade de generalização dos algoritmos, pois as camadas de *black box* das redes neurais fogem ao entendimento do desenvolvedor, e consequentemente dos que tentam reproduzi-lo (JIN et al., 2020; SOTOUDEH et al., 2019).

Por fim, há imensa falta de interoperabilidade entre a equipe assistente e a equipe de engenharia computacional, e na maioria das vezes, não se compreende os anseios e objetivos do outro (JIN et al., 2020; SOTOUDEH et al., 2019).

Por ser o escopo do presente estudo, a revisão se concentrará nas técnicas de segmentação.

2.2.2 Segmentação

Segmentação diz respeito à técnica computacional de extrair regiões de interesse, ou *regions of interest* (ROIs) de uma determinada imagem qualquer. Sendo assim, consiste em atribuir rótulos a partes de imagens que podem ser identificadas como de interesse devido a suas cores, formas, brilho, intensidade ou margens (DAR; PADHA, 2019; NOROUZI et al., 2014).

No contexto da radiologia, a segmentação resume-se em denominar ROIs em imagens dos diferentes métodos, como radiografias, ultrassonografias, TCs e RMs.

Somente na maior base de dados de indexação de publicações médicas existem mais de 400.000 publicações com o descritor “*segmentation*”, sendo mais de cem mil destas datadas dos últimos cinco anos.

A segmentação requer inicialmente o pré-processamento das imagens. Etapas de preparação e tratamento são comuns a todos os tipos de imagens, sejam elas naturais ou imagens médicas, podendo-se citar como exemplo etapas de remoção de ruídos, normalização e ajuste de intensidade.

Imagens médicas possuem como qualquer outro tipo de dados, as suas especificidades, sejam eles oriundos de diferentes aparelhos, imperfeições relativas ao método de amostragem e até a necessidade de usar um espaço de coordenadas padrão. Há ainda necessidades especiais relativas à região do corpo, que para imagens da região cerebral engloba a remoção de olhos, pescoço, coluna, etc.

Dito isso, alguns métodos possuem maior prevalência que outros, como o caso da correção de inomogeneidades, onde o método N4 é amplamente utilizado em trabalhos de segmentação de tumores cerebrais. Já para a remoção de partes que não sejam o cérebro há diversas abordagens, tais como métodos baseados em redes neurais ou por um atlas probabilístico (FONOV et al., 2011; HU et al., 2019; KLEESIEK et al., 2016; MARTINS et al., 2019; TUSTISON et al., 2010; ZHONG et al., 2020).

A segunda etapa de pré-processamento tem como objetivo o processamento das imagens originais sem a influência e degradação do nível de intensidade das lesões devido a aplicação de filtros. Assim, é feito o alinhamento do plano sagital, bem como a transformação para o espaço MNI. Em seguida é feita a segmentação do cérebro utilizando a máscara obtida anteriormente. Finalmente o nível de intensidade das imagens é padronizado, garantindo que não ocorra discrepâncias do nível de intensidade entre diferentes imagens ou diferentes pacientes.

Após o pré-processamento, procede-se com os métodos computacionais de segmentação, que se baseiam em propriedades da região de interesse (NOROUZI et al., 2014).

O mais conhecido deles, *thresholding*, consiste em uma das técnicas mais simples, partindo do pressuposto que imagens são compostas por um histograma com pixels com diferentes intensidades de cinza, sendo os pixels de intensidade de sinal inferiores ao de interesse o fundo, ou *background*, e os pixels de intensidade semelhante ou superior ao de interesse correspondem à região a ser segmentada (DAR; PADHA, 2019; NOROUZI et al., 2014).

Particularmente no caso das imagens médicas, onde os *backgrounds* não são constantes, o *thresholding* local, ou seja, de uma área de interesse, onde se define a região a ser estudada e o *background*, tem papel maior (DAR; PADHA, 2019; NOROUZI et al., 2014).

Existem ainda algoritmos para determinar automaticamente o valor de *threshold*, ou seja, a intensidade de cinza a partir de qual os pixels passam a ser definidos como de interesse (NOROUZI et al., 2014).

Dentre as técnicas *region* based, a mais difundida é a de *region growing*, onde se define uma semente em região de interesse, e a partir de homogeneidades das adjacências o método realiza a segmentação. Tem como principal desvantagem a necessidade de estabelecer uma semente, que pode ser definida manualmente ou automaticamente (DAR; PADHA, 2019; NOROUZI et al., 2014).

As estratégias baseadas no reconhecimento das margens, ou *edge based*, são capazes de detectar descontinuidades significativas dos valores de intensidade dos pixels em uma imagem, capazes de gerar contornos mais sutis (DAR; PADHA, 2019; NOROUZI et al., 2014).

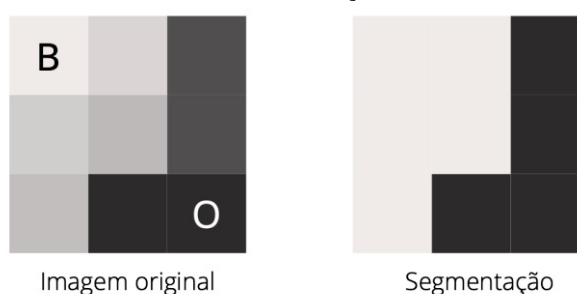
Visando sanar as segmentações que superestimavam os ROIs, ou *over-segmentation*, a técnica de *watershed* foi criada, permitindo separar objetos sobrepostos em uma imagem, sendo de fácil aplicação em imagens com escalas de cinza (DAR; PADHA, 2019; NOROUZI et al., 2014).

Dentre os métodos não-supervisionados, encontra-se o *k-nearest neighbour* onde *k* é o número de pixels vizinhos de maior semelhança com a região de interesse a serem incluídos. Sendo assim, independe das propriedades estatísticas dos pixels. Tem como maior desvantagem a necessidade de um *dataset* grande e segmentado para o treinamento do algoritmo (DAR; PADHA, 2019; NOROUZI et al., 2014).

Outro método de reconhecimento de padrão é o *maximum likelihood*, que é capaz estimar um grande volume de dados a partir de uma amostra, de acordo com um modelo estatístico, sendo o mais comum deles a distribuição Gaussiana, quando não se tem acesso ao *dataset* completo (NOROUZI et al., 2014).

Os métodos supracitados utilizam ou informações da região de interesse ou dos contornos. Existem ainda, os métodos híbridos, que utilizam ambos, sendo o *graph cut* o mais empregado nas imagens médicas. Consiste em técnica interativa onde o usuário identifica o objeto de interesse e o *background* em alguns pixels, e o algoritmo é capaz de segmentar levando em conta a detecção de informações da região de interesse, mas também de seus contornos (FIGURA 16) (NOROUZI et al., 2014).

FIGURA 16 – REPRESENTAÇÃO DO *GRAPH CUT*



FONTE: Adaptado de Norouzi et al (2014).

LEGENDA: Na imagem original à esquerda com diferentes intensidades de cinza dos pixels define-se através de sementes o objeto de interesse (O) e o *background* (B), e à direita demonstra-se o resultado da segmentação.

Para a avaliação do desempenho da segmentação através de um determinado algoritmo, utiliza-se o coeficiente Dice, que quantifica a sobreposição entre a segmentação manual e a realizada pela técnica a ser testada (NOROUZI et al., 2014).

TABELA 3 – RESUMO DAS TÉCNICAS DE SEGMENTAÇÃO

Tipo	Metodologia	Vantagens	Desvantagens
<i>Thersholding</i>	<i>Thersholding local</i>	Fácil implementação Não requer treinamento	Margens indefinidas e ruidosas
<i>Region-based</i>	<i>Region growing</i>	Baseado na similaridade dos pixels vizinhos confere certa resistência ao ruído Método fácil e conveniente	Necessidade de estabelecer semente
<i>Edge based</i>	<i>Edge detection</i>	Seleciona uma grande região da imagem	Aplicação em backgrounds simplificados
	<i>Watershed</i>	Reduz oversegmentation Separa objetos sobrepostos	Tempo

FONTE: Adaptado de Bar et al (2019).

Com a dificuldade de implementação em larga escala dos métodos automáticos, é crescente o papel das técnicas semiautomáticas, que têm como grande vantagem a fácil utilização, com inúmeras ferramentas gratuitas e amplamente disponíveis.

A mais difundida delas é o itk-SNAP, disponível para todos os sistemas operacionais, com interface simples e tendo como principal vantagem a capacidade de importar imagens em formato DICOM, sem necessidade de conversão para NiFTI.

Outra ferramenta gratuita e disponível para os principais sistemas operacionais é o 3D Slicer, que também importa formato DICOM para segmentação, que pode ser manual ou semiautomática.

Existe ainda uma alternativa brasileira, o InVesalius, desenvolvido há mais de vinte anos, que também permite a segmentação semiautomática de imagens médicas em formato DICOM.

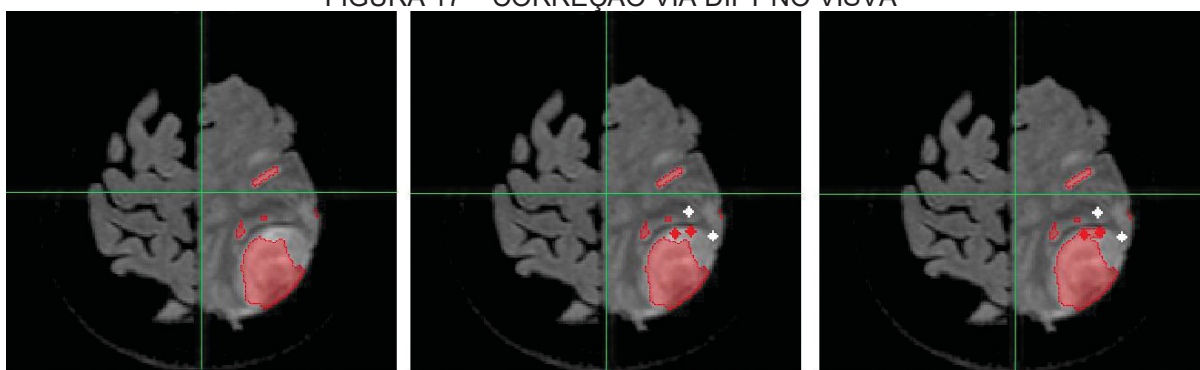
O VISVA (*Volumetric Image Segmentation, Visualization and Analysis*) é uma ferramenta desenvolvida pelo Laboratório de Informática e Data Science da Universidade de Campinas (LIDS Unicamp), que permite a visualização das imagens em três diferentes eixos (axial, coronal e sagital), além da visualização 3D da segmentação, utilizando técnica descrita por Falcão em 2004 (FALCÃO; BERGO, 2004).

O VISVA conta ainda com diversos módulos de segmentação, como por limiar, segmentação automática do cérebro e segmentação interativa manual. Para o processo de anotação volumétrica utiliza o módulo de segmentação interativa, o qual necessita de apenas alguns marcadores manuais para a segmentação de todo o processo, a partir do estabelecimento de uma semente em uma área de interesse.

Além disso, é possível com a inserção de novos marcadores corrigir a segmentação de forma rápida, sem a necessidade de começar o processo do zero, através da ferramenta IFT (*Image Foresting Transform*) em sua forma diferencial (DIFT), cujo algoritmo para segmentação ocorre da mesma forma que o *watershed*, com o crescimento de novas regiões a partir de uma representação de bordas da imagem original (FIGURA 17).

Com o VISVA pode-se então anotar as diferentes regiões dos tumores cerebrais de forma rápida e interativa, seguindo padrões já estabelecidos de anotação.

FIGURA 17 – CORREÇÃO VIA DIFT NO VISVA



FONTE: A autora (2022).

Legenda: Exemplo de uso da ferramenta VISVA, a qual usa a IFT diferencial para segmentação interativa de uma lesão cerebral. A primeira figura mostra o exemplo de uma segmentação feita anteriormente com a IFT, porém observe que há uma região que precisa ser corrigida. A segunda figura mostra a inserção de novas sementes (sementes de objeto e fundo, vermelhas e brancas, respectivamente) e finalmente a terceira figura ilustra a correção via DIFT, com aumento da área segmentada. Sendo assim, múltiplas correções podem ser feitas.

2.2.2.1 Histórico da segmentação em neuro-oncologia

Datam de mais de quarenta anos as iniciativas que buscam correlacionar o tamanho dos gliomas com parâmetros clínicos dos pacientes acometidos, sobretudo prognósticos.

Em 1979 Reeves e Marks publicaram o primeiro estudo tentando correlacionar as dimensões do glioblastoma em imagem e o prognóstico. Para isso, realizaram a

medida da área das lesões em imagens de TC pré-operatórias a partir do maior comprimento e largura de 56 pacientes (REEVES; MARKS, 1979).

Surpreendentemente, não foi encontrada correlação entre a área do tumor e a sobrevida ($p > 0,18$) quando comparados os pacientes com lesões pequenas ($< 300 \text{ cm}^2$) ou grandes ($\geq 300 \text{ cm}^2$) (REEVES; MARKS, 1979).

Reeves concluiu em seu estudo que se existe correlação entre o tamanho da lesão e a sobrevida, deve ser pequena e demonstrável apenas em amostras maiores, sendo assim, um estadiamento baseado em medidas é de baixo valor prognóstico (REEVES; MARKS, 1979).

Em 1980, Levin utilizou as medidas dos maiores eixos do realce, necrose e edema em imagens de TC para prever a probabilidade de progressão e sobrevida (LEVIN et al., 1980b).

Três anos depois, Andreou *et al* correlacionou as medidas dos dois maiores eixos das lesões em imagens de TC com a sobrevida, sem bons resultados. Porém, encontrou relação inversa do tamanho de tumor residual ($p < 0,01$) e da extensão do edema ($p < 0,01$) em imagens pós-operatórias com a sobrevida (ANDREOU et al., 1983b).

Quase uma década depois, Wood *et al* estudaram a correlação entre a área de realce dos tumores e o prognóstico de 510 pacientes. No segmento pré-operatório, não houve diferença significativa de sobrevida entre os grupos de lesões pequenas ($< 16 \text{ cm}^2$) ou grandes ($\geq 16 \text{ cm}^2$, $p = 0,48$) (WOOD; GREEN; SHAPIRO, 1988b).

Ainda em 1988, Albright Jr desenvolveu um método computacional baseado em planimetria e geometria esférica para calcular o volume das lesões a partir de imagens de TC, bem como as hipodensidades centrais, relacionadas a necrose, e perilesionais, relacionadas a edema, e as relações entre estes componentes, comprovando a maior acurácia do método quando comparado com as técnicas geométricas elipsoides ou esféricas (ALBRIGHT JR; FRAM EK, 1988a, 1988b).

Em 1999, Xue também estudou a volumetria através da planimetria de imagens de TC de 99 pacientes com gliomas, mas desta vez, correlacionando-a com a sobrevida. Os pacientes com lesões pequenas ($< 32 \text{ cm}^2$) tiveram sobrevida maior que os que apresentavam tumores grandes ($\geq 32 \text{ cm}^2$) (XUE; ALBRIGHT, 1999b).

Além disso, nos pacientes com uma performance clínica baixa, com índices de KPS menores ou iguais a 60, os que apresentavam tumores pequenos tiveram sobrevida significativamente maior ($p = 0,01$) (XUE; ALBRIGHT, 1999b).

Xue ainda comparou o método planimétrico com o método geométrico, sendo que no segundo, as correlações acima citadas não foram encontradas, concluindo assim, que no estudo dos gliomas, as medidas geométricas podem não ter valor prognóstico (XUE; ALBRIGHT, 1999b).

Com o advento e propagação da RM, Hammoud estudou a volumetria dos componentes tumorais de edema, realce e necrose, embasado no argumento de que são traduções radiológicas de variáveis biológicas do próprio tumor. Para isso, utilizou imagens de RM pré-operatórias de 48 pacientes a partir das quais foram calculados os volumes das lesões, a partir das medidas dos três eixos e assumindo-se formato elipsoide, e criadas escalas de zero a três para realce, necrose e edema (HAMMOUD et al., 1996b).

O volume das lesões não teve correlação com a sobrevida dos pacientes. No entanto, houve diferença estatisticamente significativa quanto maior o grau de necrose ($p < 0,001$) e realce ($p = 0,003$). Já para o edema, houve um efeito quadrático, com a pior sobrevida no grau II ($p = 0,001$) (HAMMOUD et al., 1996b).

Foi o primeiro estudo a inferir o valor prognóstico da quantidade de necrose em RMs pré-operatórias de pacientes com glioblastoma.

Desde então, diversos autores buscaram a correlação da volumetria tumoral com aspectos clínicos, cirúrgicos, histomoleculares e prognósticos em gliomas de alto grau, e nesse contexto, foi crescente o papel das técnicas semiautomáticas.

2.2.2.2 Estado da arte em segmentação e gliomas

Foi realizada busca avançada da literatura com os descritores expostos no quadro abaixo (QUADRO 1) e com o auxílio de operadores booleanos.

DESCRITORES UTILIZADOS	
Glioma	<i>Segmentation</i>
Glioblastoma	<i>Volumetry</i>

QUADRO 1 – DESCRITORES UTILIZADOS NA REVISÃO BIBLIOGRÁFICA
Fonte: A autora (2022).

Foram encontrados 86 artigos na língua inglesa com os descritores acima no título, dos quais 71 tinham como escopo a segmentação e suas aplicações em gliomas, sendo a maioria relacionado à técnicas automáticas e publicação de algoritmos de segmentação (TABELA 4).

TABELA 4 – RESUMO DOS ARTIGOS PUBLICADOS SOBRE SEGMENTAÇÃO 3D EM GLIOMAS
(continua)

Ano	Autor	Objetivo	Técnica	N
1995	Phillips WE <i>et al</i>	Proposta de algoritmo	Automática	1
2005	Hata <i>et al</i>	Proposta de ferramenta de segmentação	Semiautomática	15
2009	Emblem KE <i>et al</i>	Proposta de algoritmo	Automática	50
2010	Galldiks N <i>et al</i>	Segmentação semiautomática do PET-RM	Semiautomática	12
2010	Assefa D <i>et al</i>	Proposta de algoritmo	Automática	27
2011	Iliadis G <i>et al</i>	Volumetria e prognóstico	Semiautomática	50
2012	Gooya A <i>et al</i>	Proposta de algoritmo	Automática	122
2012	Zhu Y <i>et al</i>	Proposta de ferramenta de segmentação	Semiautomática	26
2013	Schmitt P <i>et al</i>	Comparação segmentação manual x medidas lineares	Manual	65
2014	Porz N <i>et al</i>	Comparação segmentação manual x algoritmo	Automática	25
2014	Kwon D <i>et al</i>	Proposta de ferramenta de segmentação	Semiautomática	Não especifica
2014	Cordova JS <i>et al</i>	Proposta de algoritmo	Automática	37
2014	Lee J <i>et al</i>	Proposta de algoritmo	Automática	32
2014	Tran AN <i>et al</i>	Predição de resposta tratamento adjuvante	Semiautomática	39
2014	Grabowski MM <i>et al</i>	Volumetria residual e prognóstico	Semiautomática	128
2015	Odland A <i>et al</i>	Comparação segmentação manual x semiautomática	Semiautomática	20
2015	Njeh I <i>et al</i>	Segmentação	Automática	Não usou dataset próprio
2015	Juan-Albarracín J <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2015	Riley GT <i>et al</i>	Predição da progressão pela volumetria	Semiautomática	22
2015	Amelot A <i>et al</i>	Proposta de atlas para prever crescimento glioma	Automática	152
2015	Kazerooni AF <i>et al</i>	Proposta de algoritmo	Automática	13
2015	Steed TC <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2015	Wangaryattawanich P <i>et al</i>	Volumetria e prognóstico	Semiautomática	94
2016	Fyllingen <i>et al</i>	Comparação entre ferramentas	Semiautomática	20
2016	Dunn Jr W <i>et al</i>	Comparação entre ferramentas	Semiautomática	67
2016	Sreenivasan SA <i>et al</i>	Comparação segmentação manual x medidas lineares	Manual	44
2016	Abdallah MB <i>et al</i>	Variabilidade segmentação	Manual	12
2016	Cordier N <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2016	Singh R <i>et al</i>	Proposta de algoritmo	Automática	8
2016	Bakas S <i>et al</i>	Proposta de algoritmo	Automática	186
2017	Bakas S <i>et al</i>	Publicação database e algoritmo	Automática	243

2017	Huber T <i>et al</i>	Predição da progressão pela volumetria	Semiautomática	30
2017	Wu YP <i>et al</i>	Segmentação em aplicativo de celular	Semiautomática	129
2017	Bø <i>et al</i>	Variabilidade segmentação	Semiautomática	23
2017	Li Z <i>et al</i>	Proposta de algoritmo	Automática	160
2017	Lee M <i>et al</i>	Comparação entre ferramentas	Semiautomática	Não usou dataset próprio
2017	Petersen J <i>et al</i>	Comparação entre ferramentas	Semiautomática	Não usou dataset próprio
2018	Zhao Z <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2018	Selvapandian A <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2018	Perkuhn M <i>et al</i>	Proposta de algoritmo	Automática	64
2018	Thust SC <i>et al</i>	Segmentação 2D x 3D para prever mutações	Semiautomática	44
2019	Wu Y <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2019	Tixier F <i>et al</i>	Confiabilidade da segmentação	Semiautomática	90
2019	Sun R <i>et al</i>	Proposta de algoritmo	Automática	5
2019	Henker C <i>et al</i>	Correlação volumetria x Ki-67	Semiautomática	152
2019	Canalini L <i>et al</i>	Segmentação ultrassom intraoperatório	Automática	Não usou dataset próprio
2019	Ozaki T <i>et al</i>	Segmentação em PET-RM	Automática	45
2020	Naser MA <i>et al</i>	Proposta de algoritmo	Automática	110
2020	Berntsen EM <i>et al</i>	Predição da progressão pela volumetria	Manual	106
2020	Su R <i>et al</i>	Proposta de algoritmo	Automática	26
2020	Khoshnavanian A <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2020	Zhong L <i>et al</i>	Proposta de algoritmo	Automática	62
2020	Tang F <i>et al</i>	Proposta de algoritmo	Automática	59
2020	Choi Y <i>et al</i>	Comparação segmentação manual x algoritmo	Automática	45
2020	Eijgelaar RS <i>et al</i>	Proposta de algoritmo	Automática	634
2020	Wan Y <i>et al</i>	Proposta de algoritmo	Automática	120
2021	Bouget D <i>et al</i>	Comparação segmentação manual x algoritmo	Automática	1887
2021	Zhihua L <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2021	Decuyper M <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2021	Aziz MJ <i>et al</i>	Segmentação	Automática	Não usou dataset próprio
2021	Zhang Z <i>et al</i>	Proposta de algoritmo	Automática	Não especifica
2021	Khalaf NA <i>et al</i>	Variabilidade segmentação	Automática	20
2021	Van Garderen KA <i>et al</i>	Implementação de ferramenta de segmentação	Automática	55
2021	Park JE <i>et al</i>	Proposta de algoritmo	Automática	238
2021	Van Kempen EJ <i>et al</i>	Metanálise	Não se aplica	Não se aplica

2021	Zheijang J	Proposta de algoritmo	Automática	Não usou dataset próprio
2021	Fu J <i>et al</i>	Proposta de algoritmo	Automática	258
2021	Ali MJ <i>et al</i>	Proposta de algoritmo	Automática	Não especifica
2021	Carver EN <i>et al</i>	Proposta de algoritmo	Automática	Não usou dataset próprio
2022	Lotan E <i>et al</i>	Proposta de algoritmo	Automática	437
2022	Cheng J <i>et al</i>	Segmentação e predição IDH	Automática	Não especifica
2022	Zhang TC <i>et al</i>	Proposta de algoritmo	Automática	18
2022	Thomas MF <i>et al</i>	Segmentação em sequências faltando sintetizadas	Automática	68

FONTE: A autora (2022)

No tocante ao prognóstico, os principais estudos com objetivos semelhantes encontram-se resumidos na tabela abaixo (TABELA 5).

TABELA 5 – REVISÃO DA LITERATURA EM PROGNÓSTICO E VOLUMETRIA

Autor	N	ROI	Técnica	Variáveis	Resultados
Lacroix M <i>et al</i> (2001)	416	Realce (0 a II), necrose (0 a III), edema (0 a III)	Seleção da área no axial e extrapolação para volume através da espessura da lesão (MedVision 1.14)	Idade, sexo, KPS, tipo de cirurgia, tratamento adjuvante, status vital, porcentagem de ressecção	Graus maiores de realce ($p = 0,007$), edema ($p = 0,04$) e necrose ($p < 0,001$) ligados a sobrevida inferior.
Iliadis G <i>et al</i> (2011)	65	Volumes de realce, edema e necrose	Semiautomática (Volumio)	Idade, sexo, KPS, tipo de cirurgia, MGMT	Menor sobrevida relacionada a maiores volumes de realce e necrose, sem relação com edema. Mutação MGMT ligada a valores menores de necrose
Grabowski MM <i>et al</i> (2014)	128	Volume de realce (T1Gd+) pré e pós-operatório	Semiautomática (BrainLab)	Idade, sexo, KPS	Menor sobrevida relacionada a maiores volumes de realce pré e pós-operatório
Wangaryattawanich P <i>et al</i> (2015)	94	Volumes de realce, edema e necrose	Semiautomática (3D Slicer)	Idade, sexo, KPS	Menor sobrevida relacionada a maiores volumes de edema (<i>cutoff</i> de 85.000 mm ³ , $p = 0,019$) e realce (<i>cutoff</i> de 35.000 mm ³ , $p = 0,013$)

FONTE: A autora (2022)

NOTA: O estudo de Lacroix *et al* não foi incluído na tabela 4 por não se tratar de segmentação 3D.

Além do prognóstico, alguns autores estudaram a correlação da volumetria com mutações e sua capacidade em predizê-las.

Lee *et al* avaliou o volume dos componentes tumorais em gliomas grau 4 e identificou volumes menores de realce e necrose em pacientes com mutação IDH (LEE et al., 2019).

Para mutações MGMT os resultados são mais escassos, destacando-se o estudo de Iliadis *et al*, que encontrou volumes menores de necrose em pacientes mutados (ILIADIS et al., 2012).

Com relação à associação dos índices Ki67 e volumetria em glioma, a literatura é escassa e controversa (TABELA 6).

TABELA 6 – REVISÃO DA LITERATURA SOBRE CORRELAÇÃO KI67 E VOLUMETRIA

Autor	Ano	N	Técnica	Segmentação	Achados
Henker <i>et al</i>	2019	152	Semiautomática (<i>contour expansion - SmartBrush by Brainlabs</i>)	Edema no 2D FLAIR; Realce tumor e necrose no 3D T1Gd+; subtrai áreas císticas	Ausência de correlação do Ki67 com volumetria
Armocida <i>et al</i>	2020	127	<i>Manual tracer</i> (Osirix)	Não específica (área de sinal alterado no T1, sem especificar 2D/3D)	Correlação entre Ki67 e volume tumoral em lesões > 45 cm ³
Caramanti <i>et al</i>	2022	70	<i>Manual tracer</i> (Horos)	Edema no FLAIR; realce no T1Gd+ (não especifica 2D/3D)	Correlação entre edema (p < 0,004) e Ki67. Sem correlação com realce ou tumor total

FONTE: A autora (2022).

3 MATERIAIS E MÉTODOS

3.1 DESENHO DO ESTUDO

Trata-se de estudo observacional transversal retrospectivo.

3.2 OBJETIVOS

O presente estudo tem como objetivo primário a correlação da volumetria das subáreas tumorais com variáveis clínicas (sobrevida) e histopatológicas (IDH, ATRX, MGMT e Ki67).

Os objetivos secundários concentram-se na correlação da volumetria com aspectos cirúrgicos (tipo de ressecção) e margens operatórias ao 5-ALA (fluorescência residual ou não).

3.3 CRITÉRIOS DE INCLUSÃO

A partir dos registros do serviço de Neurocirurgia do Instituto de Neurologia de Curitiba, foram incluídos todos os pacientes submetidos à abordagem cirúrgica (biópsia, ressecções parciais, subtotais e completas) de astrocitomas grau 4 e glioblastomas, com diagnóstico histológico confirmado, entre os anos de 2015 e 2021 e com exames de RM pré-operatórios disponíveis, contendo obrigatoriamente sequências volumétricas T2 FLAIR e T1 pós-gadolínio.

Cento e trinta e sete pacientes foram inicialmente incluídos, sem predileção por sexo, faixa etária, cor ou grupo social específicos.

3.4 CRITÉRIOS DE EXCLUSÃO

Foram excluídos os pacientes com imagens de RM pré-operatória indisponíveis, ou aqueles com protocolo de aquisição de imagem inadequado, sinais de manipulação neurocirúrgica prévia ou de artefatos que impossibilitassem a análise das imagens.

3.5 MÉTODOS

3.5.1 Imagens

Os exames pré-operatórios foram adquiridos em dois aparelhos da *General Electric* (GE) *Healthcare* de 1.5 Tesla (Signa HDxt) com bobina de cabeça de oito canais, e um de 3.0 Tesla da GE *Healthcare* (Discovery MR750W) com bobina de cabeça e pescoço de 28 canais.

As imagens incluíram obrigatoriamente sequências volumétricas 3D T2 FLAIR (FLAIR CUBE) e 3D T1 pós-gadolínio, segundo os parâmetros descritos abaixo (TABELA 7).

TABELA 7 – PARÂMETROS DE AQUISIÇÃO DAS IMAGENS DE RM

Parâmetro	FLAIR CUBE	3D T1 FSPGR pós-gadolínio
FOV (mm)	24,00	24,00
Phase FOV	1,00	0,80
NEX	1,0	1,0
Espessura de corte (mm)	1,20	1,40
TR (ms)	5.000	Mínimo
TE (ms)	130	<i>In phase</i>
Tamanho do pixel (mm ²)	1,1 x 1,1 mm	1,0 x 1,0 mm

FONTE: A autora (2023).

NOTA: RM – ressonância magnética; FOV – *field of view*; NEX – *number of excitations*; TR – tempo de repetição; TE – tempo de *echo*.

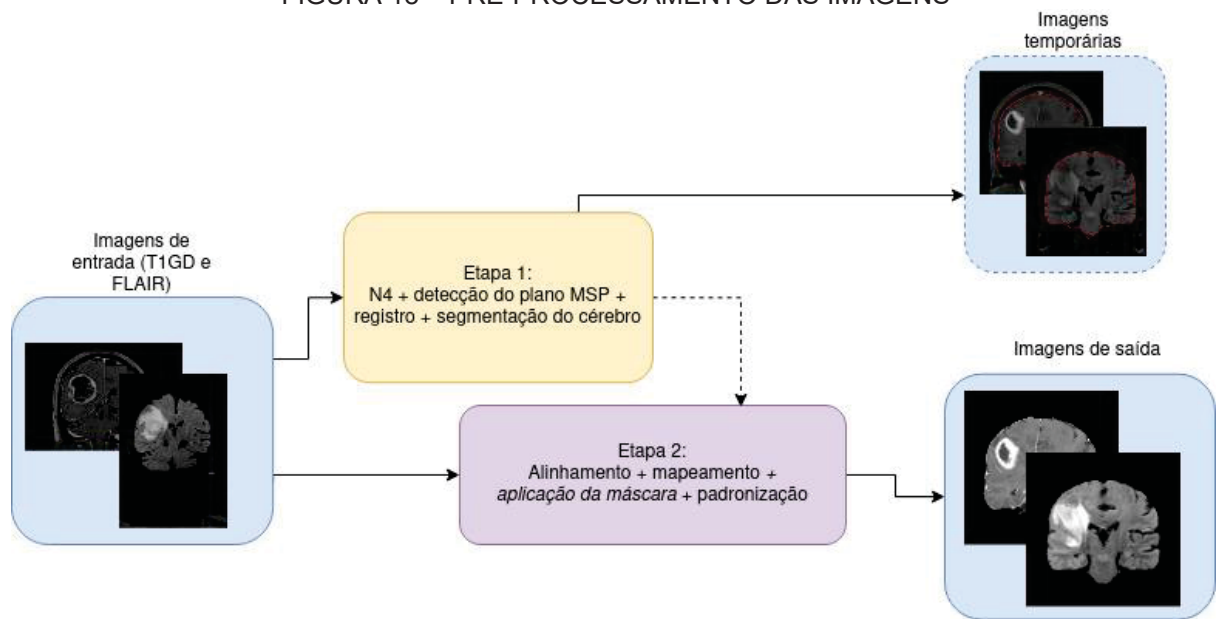
As imagens de RM foram salvas em formato DICOM em workstation Advantage Workstation 4.7 (GE Medical System, Milwaukee, WI) em *hard disk* (HD) externo.

Como pré-processamento, as imagens foram então submetidas a interpolação para conversão isométrica para *voxels* de 1 x 1 x 1 mm, e em seguida convertidas de formato DICOM para NIFTI.

Após a conversão em formato NIFTI, através de métodos digitais e de *stripping* foram removidas informações das imagens que permitam a identificação do paciente, como nome, registro e *voxels* da face.

Além disso, as imagens foram registradas em um mesmo eixo de coordenadas, foi realizada remoção de ruído, correção de inomogeneidade e *histogram matching* (FIGURA 18).

FIGURA 18 – PRÉ-PROCESSAMENTO DAS IMAGENS

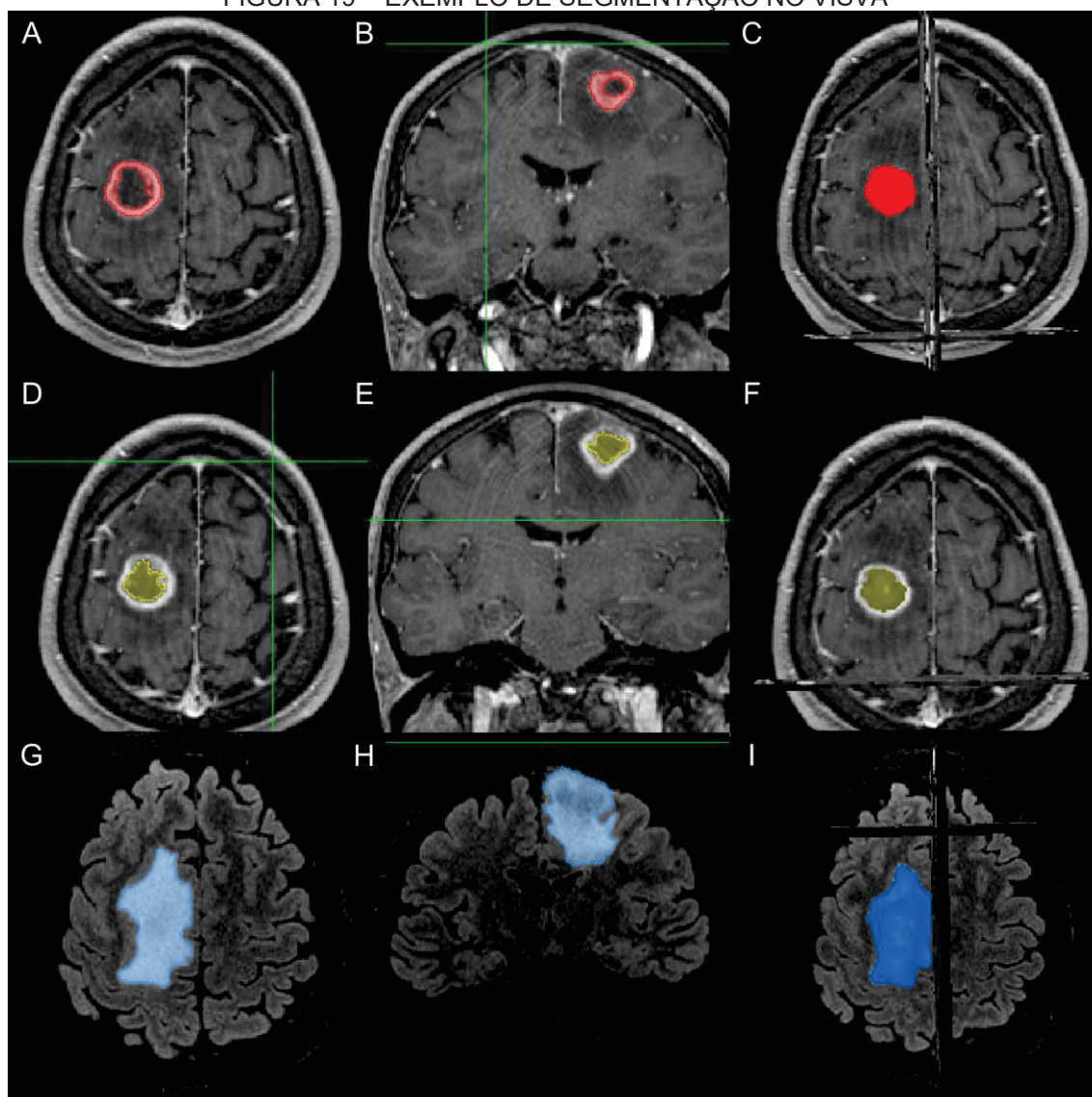


FONTE: A autora (2022).

LEGENDA: Pré-processamento de imagens de lesão cerebral em duas etapas. À esquerda é demonstrado o *input* com as imagens em T1 pós-contraste e FLAIR. A primeira etapa realiza ajustes espaciais e a segmentação do cérebro, gerando as imagens temporárias. A segunda etapa realiza o alinhamento e mapeamento das imagens, com aplicação da máscara e padronização às imagens originais, sem a degradação de filtros, e destas surgem as imagens de saída devidamente pré-processadas para segmentação.

As imagens devidamente pré-processadas foram importadas no *software* VISVA, em ambiente Linux, onde foram segmentadas de forma semi-automática para os *labels* WT em FLAIR CUBE e ET e NC em 3D T1 pós-gadolínio, e os volumes em mm³ foram inseridos na planilha de dados. Os volumes de ED foram obtidos através da subtração de ET e NC de WT (FIGURA 19).

FIGURA 19 – EXEMPLO DE SEGMENTAÇÃO NO VISVA



FONTE: A autora (2021).

LEGENDA: Exemplo de segmentação no VISVA. A-C: Segmentação para realce (ET, vermelho) em 3D T1Gd+ axial, coronal e *volume rendering*, respectivamente. D-F: Segmentação para necrose (NC, amarelo) em 3D T1Gd+ axial, coronal e *volume rendering*, respectivamente. G-I: Segmentação para tumor total (WT, azul) em 3D FLAIR axial, coronal e *volume rendering*, respectivamente.

Após a segmentação feita por médico residente do terceiro ano de Radiologia e conferida por médico neurorradiologista com cinco anos de experiência, os arquivos etiquetados foram exportados e encaminhados aos pesquisadores do departamento de informática.

3.5.2 Dados clínicos, cirúrgicos e histomoleculares

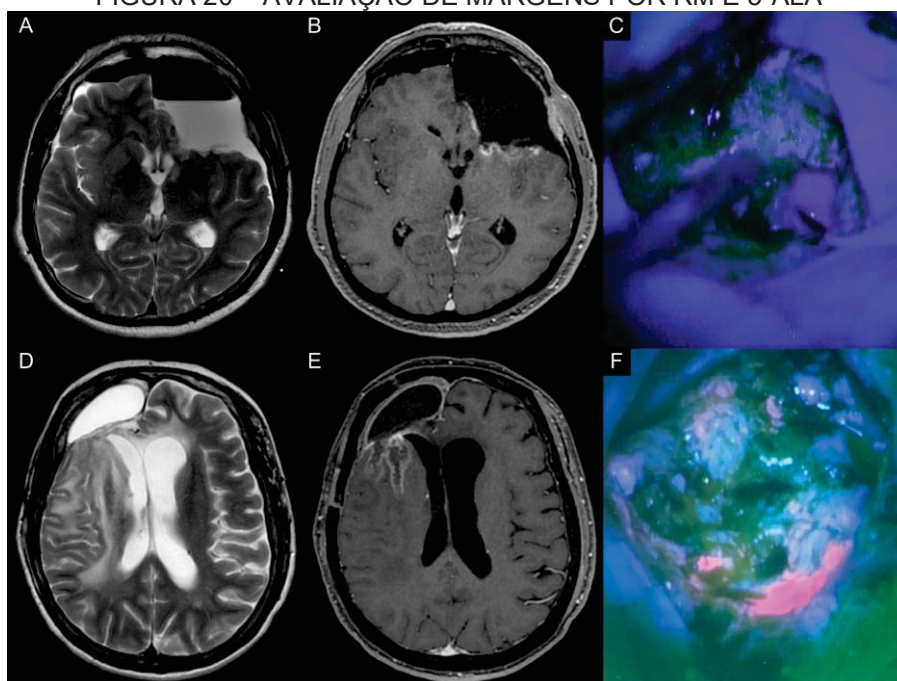
Em paralelo, foram consultados em prontuário e laudos médicos, e inseridos na planilha de dados os achados cirúrgicos (tipo de ressecção, fluorescência 5-ALA),

histopatológicos (diagnóstico por congelação, diagnóstico definitivo, IDH, MGMT, ATRX, TP53 e Ki67) e prognósticos (sobrevida livre de doença e geral). Todos os pacientes foram submetidos à mesma modalidade de terapia adjuvante segundo o protocolo Stupp.

Ressecções completas foram consideradas na ausência de realces sólidos pelo contraste em exame de RM pós-operatória imediata. Os pacientes que apresentassem lesão residual, com realce sólido ou nodular pelo contraste foram considerados como ressecções incompletas.

A ausência de fluorescência 5-ALA residual no leito operatório foi classificada como margens 5-ALA livres. Na presença de fluorescência residual no leito operatório, os pacientes foram classificados como margens 5-ALA comprometidas (FIGURA 20).

FIGURA 20 – AVALIAÇÃO DE MARGENS POR RM E 5-ALA



FONTE: A autora (2023).

LEGENDA: A-B: Imagens de RM pós-operatórias imediatas com axial T2 (A) e T1 pós-gadolinio (B) demonstrando ressecção de lesão frontal esquerda, com realce linear ao redor da cavidade cirúrgica em B. C: Imagem intraoperatória demonstrando ausência de fluorescência 5-ALA residual no leito operatório. Este caso foi classificado como ressecção completa com margens 5-ALA livres. D-E: Imagens de RM pós-operatórias imediatas com axial T2 (D) e T1 pós-gadolinio (E) demonstrando cavidade operatória frontal direita, com realce residual sólido e irregular na margem posterior. F: Imagem intraoperatória evidencia fluorescência 5-ALA residual em rosa. Este caso foi classificado como ressecção incompleta com fluorescência 5-ALA residual no leito operatório.

3.6 ANÁLISE ESTATÍSTICA

A análise estatística foi realizada com o auxílio de planilha eletrônica do Microsoft Excel e do *software* SPSS20, estabelecendo nível de significância de 0,05 (5%).

A normalidade da distribuição das variáveis foi testada a partir do teste de Kolmogorov-Smirnov, permitindo o emprego de testes paramétricos.

Foram empregados os testes T de Student para a análise de variáveis quantitativas contínuas, qui-quadrado e teste de Fischer para a comparação de dois grupos com variáveis discretas. O teste ANOVA foi utilizado para a análise de variáveis contínuas em três ou mais grupos, com a comparação múltipla de Turkey para comparação aos pares.

Para a avaliação da concordância entre laudos foi empregado o índice de concordância de Kappa. A correlação de Pearson foi utilizado para a relação entre as variáveis quantitativas contínuas.

Para a correlação do Ki67 com a fluorescência 5-ALA do leito operatório utilizou-se um *cutoff* de 20%, segundo os resultados de Henker *et al* (HENKER et al., 2019).

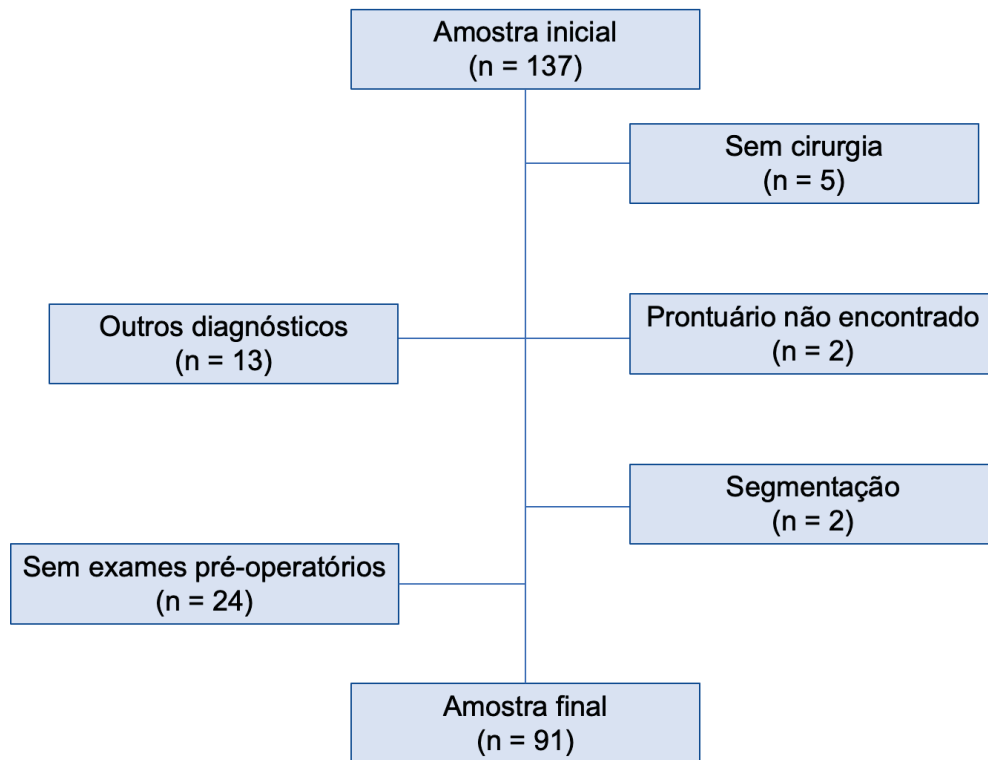
3.7 PARECER ÉTICO

O presente estudo tem aprovação no Comitê de Ética e Pesquisa em Seres Humanos (CEPSH) do INC sob o registro CAAE 40280320.7.0000.5227 datada de 16 de dezembro de 2020 (ANEXO I).

4 RESULTADOS

Foram selecionados inicialmente 137 pacientes, dos quais 46 foram excluídos (FIGURA 21). A amostra final foi composta por 91 pacientes.

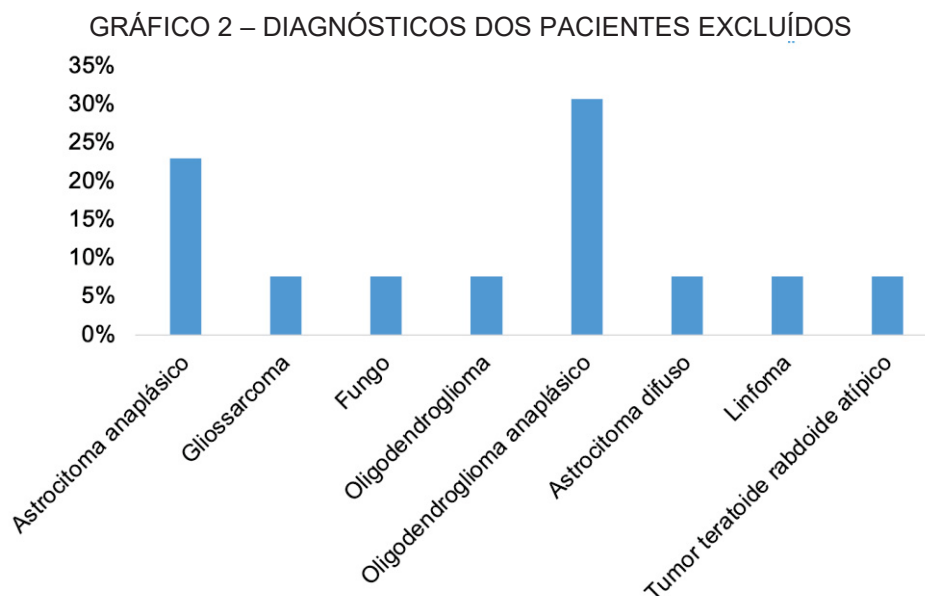
FIGURA 21 – AMOSTRA FINAL



FONTE: A autora (2022).

NOTA: Segmentação – erros no pré-processamento que ao remover os voxels do crânio removiam parte da lesão.

Dos pacientes excluídos por possuírem outros diagnósticos, a maioria possuía gliomas de grau 3 (GRÁFICO 2).



FONTE: A autora (2021).

NOTA: Segundo a classificação da OMS de 2016.

Mais da metade da amostra ($n = 56$, 61%) era composta por pacientes do sexo masculino. A idade dos pacientes variou entre 25 e 90 anos de idade, com uma média de 57 anos e mediana de 58 anos (TABELA 8), sem diferenças significativas entre os sexos ($p = 0,41$).

TABELA 8 – DISTRIBUIÇÃO DA AMOSTRA POR IDADE

Variável	Geral	Feminino ($n = 35$)	Masculino ($n = 56$)
Mínimo	25	25	27
Máximo	90	90	79
Média	$57 \pm 12,42$	$57,34 \pm 14,09$	$56,66 \pm 11,38$
Mediana	58	57	58

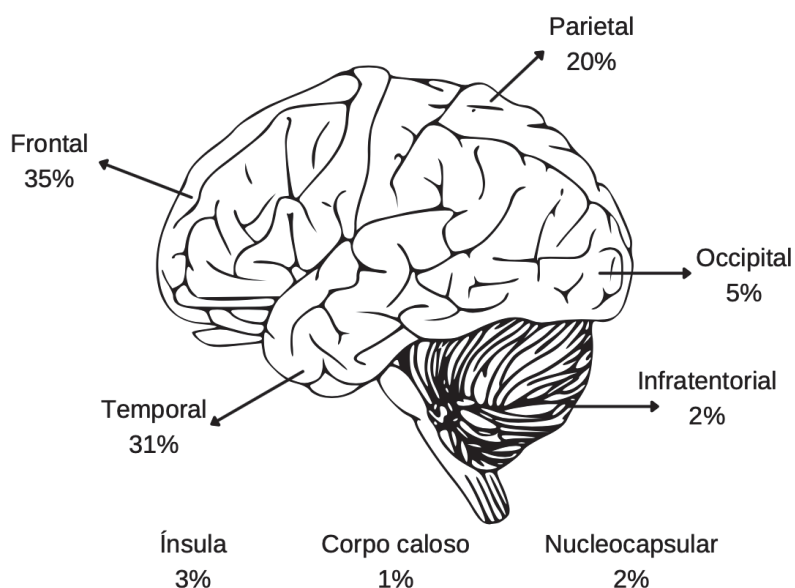
FONTE: A autora (2021).

4.1 IMAGEM E VOLUMETRIA

Os pacientes possuíam no mínimo uma e no máximo seis lesões, com uma média de 1,14 lesões por paciente.

As lesões eram mais frequentes nos lobos frontal ($n = 32$, 35%) e temporal ($n = 28$, 31%) (FIGURA 22), e pouco mais frequentes à esquerda ($n = 47$, 52%). Aproximadamente 5% das lesões ($n = 5$) acometiam ambos os hemisférios.

FIGURA 22 – DISTRIBUIÇÃO POR LOBOS DAS LESÕES



FONTE: A autora (2022).

O número de exames de RM pré, intra e pós-operatórios são descritos na tabela 9 (TABELA 9).

TABELA 9 – NÚMERO DE EXAMES

	Pré-operatório	Intraoperatório	Pós-operatório
Mínimo	1	0	0
Máximo	3	1	27
Média	1,33 ± 0,55	0,43 ± 0,5	6,66 ± 5,82
Mediana	1,00	0,00	6,00

FONTE: A autora (2021).

As características de imagem descritas em laudo médico são expostas nas tabelas abaixo (TABELA 10). Os laudos de três pacientes não estavam disponíveis.

TABELA 10 – CARACTERÍSTICAS DE IMAGEM EM SEQUÊNCIAS AVANÇADAS

	Redução NAA e creatina (n = 88)	Aumento de lípidos e lactato (n = 88)	Aumento colina (n = 88)	Perfusão aumentada (n = 88)
Sim	64	53	62	68
Não	0	5	2	20
Não citado	24	30	24	0

FONTE: A autora (2021).

As volumetrias dos componentes tumorais encontram-se descritas na tabela abaixo (TABELA 11).

TABELA 11 – VOLUMETRIAS TUMORAIS EM MM³

	Parâmetro	Volumes (mm ³)
WT	Mínimo	5.151,00
	Máximo	24.1853,00
	Média	96.044 ± 60.343,87
	Mediana	87.277,00
ET	Mínimo	0,00
	Máximo	171.860,00
	Média	24.312,20 ± 27.583,92
	Mediana	15.140,00
NC	Mínimo	0,00
	Máximo	73.487,00
	Média	8.078,55 ± 12.920,93
	Mediana	3.565,00
ED	Mínimo	0,00
	Máximo	199.626,00
	Média	61.891,64 ± 48.606,93
	Mediana	53.348,00

FONTE: A autora (2021).

NOTA: WT – *whole tumor*; ET – *enhancing tumor*; NC – *necrotic core*; ED – edema.

Volumes maiores de algum dos componentes tumorais estavam relacionados a maiores volumes das demais áreas, com exceção da correlação entre os volumes de edema, com os de realce e necrose (TABELA 12).

TABELA 12 – CORRELAÇÃO ENTRE VOLUMETRIAS

		WT	ET	NC
ET	Corr (r)	0,500		
	P-valor	<0,001		
NC	Corr (r)	0,433	0,484	
	P-valor	<0,001	<0,001	
ED	Corr (r)	0,855	0,094	-0,006
	P-valor	<0,001	0,399	0,960

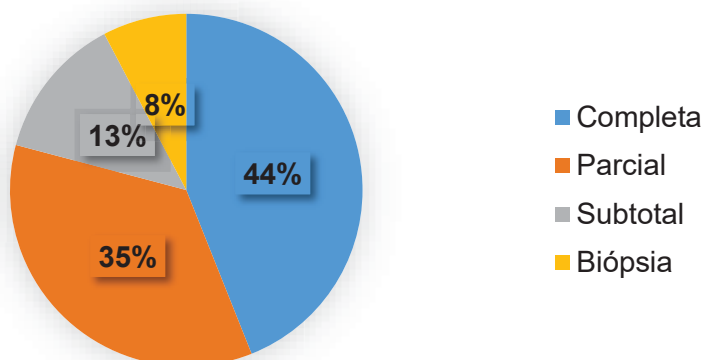
Fonte: A autora (2021).

NOTA: Corr (r) – Correlação de Pearson; WT – *whole tumor*; ET – *enhancing tumor*; NC – *necrotic core*; ED – edema.

4.2 ASPECTOS CIRÚRGICOS E TERAPÊUTICOS

Quanto ao tipo de abordagem cirúrgica, a maioria dos pacientes foi submetida à ressecção completa (n = 40, 44%) ou parcial (n = 32, 35%) das lesões, seguidas de ressecção subtotal (n = 12, 13%) e biópsia (n = 7, 8%) (GRÁFICO 4).

GRÁFICO 4 – TIPO DE ABORDAGEM CIRÚRGICA



FONTE: A autora (2021).

Os pacientes submetidos a ressecções incompletas (parcial e subtotal) tiveram sobrevida geral menor (13,9 *versus* 23,7 meses) em comparação com as ressecções completas ($p = 0,004$) (TABELA 13, GRÁFICO 5).

Não houve correlação do tipo de ressecção com a volumetria, nem com o número de lesões (TABELA 13).

Os pacientes submetidos a ressecções completas realizaram mais exames de RM pós-operatórios ($p < 0,001$) (TABELA 13).

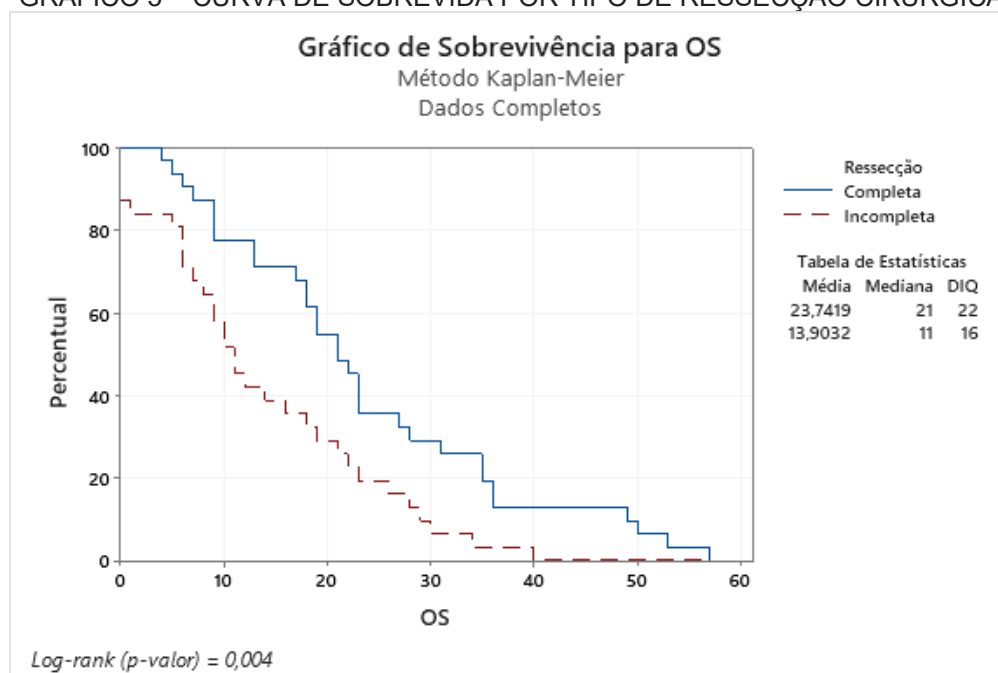
TABELA 13 – TIPO DE RESSECÇÃO *versus* PROGNÓSTICO E VOLUMETRIA TUMORAL

Parâmetro	Tipo de ressecção	Média	Mediana	Desvio Padrão	Mín	Max	IC	p
PFS (meses)	Completa	12,6	12	8,2	0	28	2,9	0,404
	Incompleta	10,6	6,5	10,1	1	40	3,9	
OS (meses)	Completa	23,7	21	14,5	4	57	5,1	0,004
	Incompleta	13,9	11	10,9	0	40	3,8	
WT (mm ³)	Completa	93.558	97.408	61.594	5.151	241.853	19.331	0,723
	Incompleta	98.354	86.231	59.813	10.246	222.636	18.089	
ET (mm ³)	Completa	24.261	15.892	22.330	0	80.419	6.920	0,987
	Incompleta	24.360	12.151	31.977	0	171.860	9.558	
NC (mm ³)	Completa	8.868	3.812	12.532	0	54.062	3.884	0,594
	Incompleta	7.344	3.297	13.378	0	73.487	3.999	
ED (mm ³)	Completa	55.323	46.268	53.686	0	176.300	16.236	0,604
	Incompleta	61.060	54.157	50.286	0	199.626	14.376	
Exames pós-operatórios	Completa	9,1	8	6,6	0	27	2,1	<0,001
	Incompleta	4,5	3	4,1	0	16	1,2	
Número de lesões	Completa	1,0	1	0,2	1	2	0,0	0,128
	Incompleta	1,3	1	1,0	1	6	0,3	

FONTE: A autora (2021).

NOTA: PFS – *progress free survival*; OS – *overall survival*; WT – *whole tumor*; ET – *enhancing tumor*; NC – *necrotic core*; ED – *edema*; Min – mínimo, Max – máximo; IC – intervalo de confiança. Teste T de Student.

GRÁFICO 5 – CURVA DE SOBREVIDA POR TIPO DE RESSECÇÃO CIRÚRGICA



FONTE: A autora (2022).

NOTA: OS – *overall survival*.

Das cirurgias que empregaram o ALA-5 ($n = 76$), em 41 delas as margens cirúrgicas encontravam-se livres de doença (54%). Nesse grupo, notou-se maior sobrevida geral em meses (14,7 *versus* 22,9 meses, $p = 0,016$) (TABELA 14).

TABELA 14 – MARGENS CIRÚRGICAS 5-ALA E SOBREVIDA EM MESES

	Parâmetro	Média (meses)	Mediana (meses)	Desvio Padrão	Min	Max	IC	p
PFS	5-ALA residual	11,4	7	10,4	1	40	4,0	0,824
	5-ALA livre	11,9	11	8,0	0	28	2,9	
OS	5-ALA residual	14,7	11	11,5	0	40	4,0	0,016
	5-ALA livre	22,9	21	14,6	4	57	5,1	

FONTE: A autora (2021).

NOTA: PFS – *progress free survival*; OS – *overall survival*; Min – mínimo, Max – máximo; IC – intervalo de confiança.

Teste T de Student.

Não encontramos correlação entre a volumetria tumoral e fluorescência 5-ALA residual no leito operatório (TABELA 15).

TABELA 15 – MARGENS CIRÚRGICAS 5-ALA E VOLUMETRIA EM MM³

		Mín	Máx	Média	Mediana	p
WT	5-ALA residual	17941,00	222636,00	101851,39	89586,00	0,421
	5-ALA livre	5151,00	241853,00	90766,26	91777,00	
ET	5-ALA residual	0,00	171860,00	25705,78	19326,50	0,823
	5-ALA livre	0,00	80419,00	24190,54	15140,00	
NC	5-ALA residual	0,00	73487,00	8931,35	4401,00	0,971
	5-ALA livre	0,00	54062,00	7573,89	3037,00	
ED	5-ALA residual	0,00	176300,00	55729,71	44444,00	0,470
	5-ALA livre	0,00	199626,00	63286,73	54566,50	

FONTE: A autora (2022).

NOTA: Min – mínimo, Max – máximo.

Teste T de Student.

Todos os pacientes da amostra foram submetidos à adjuvância segundo o protocolo Stupp. Quinze pacientes (16%) receberam bevacizumab em associação com a quimioterapia convencional.

4.3 DIAGNÓSTICO HISTOLÓGICO

O diagnóstico histológico final dos pacientes foi de glioblastoma (n = 55, 60%), glioblastoma NOS (n = 21, 23%) e astrocitoma grau 4 (n = 15, 16%), segundo a classificação de 2021 da OMS.

O perfil imunohistoquímico da amostra encontra-se descrito na tabela abaixo (TABELA 16).

TABELA 16 – PERFIL IMUNOHISTOQUÍMICO

	IDH (n = 70)	MGMT (n = 29)	ATRX (n = 40)	P53 (n = 39)
Positivo	15	12	38	37
Negativo	55	17	2	2

FONTE: A autora (2022).

Pacientes com mutação IDH (n = 15) apresentaram sobrevida geral maior (28 *versus* 19,5 meses, p = 0,049) em comparação com os pacientes IDH selvagem, sem diferenças entre os grupos para a sobrevida livre de doença (p = 0,107). Não houve correlação com os volumes das lesões (TABELA 17).

TABELA 17 – IDH *versus* PROGNÓSTICO E VOLUMETRIA

IDH		Média	Mediana	DP	Min	Max	IC	p
PFS	Negativo	13,4	14	8,9	1	27	3,5	0,107
	Positivo	21,0	25	14,5	1	40	11,6	
OS	Negativo	19,5	19	12,7	0	53	5,0	0,049
	Positivo	28,0	28	15,8	6	49	11,7	
WT	Negativo	94.485	89.586	64.884	7.003	241.853	22.138	0,458
	Positivo	112.655	114.196	42.313	59.909	179.867	29.321	
ET	Negativo	22.525	12.151	24.054	0	80.419	8.207	0,324
	Positivo	13.550	6.389	15.669	0	35.571	10.858	
NC	Negativo	9.376	2.048	17.228	0	73.487	5.878	0,956
	Positivo	9.736	2.139	13.376	0	35.738	9.269	
ED	Negativo	59.007	50.247	46.233	0	176.300	15.317	0,471
	Positivo	71.496	74.412	53.906	0	144.618	33.411	

FONTE: A autora (2022).

NOTA: PFS – *progress free survival*; OS – *overall survival*; WT – *whole tumor*; ET – *enhancing tumor*; NC – *necrotic core*; ED – edema; DP – desvio padrão; Min – mínimo, Max – máximo; IC – intervalo de confiança.

Teste T de Student.

O *status* MGMT não teve correlação com o prognóstico nem com valores de volumetria tumoral (TABELA 18).

TABELA 18 – MGMT *versus* PROGNÓSTICO E VOLUMETRIA

MGMT		Média	Mediana	DP	Min	Max	IC	p
PFS	Negativo	14,9	13	11,4	1	28	7,9	0,994
	Positivo	14,8	18,5	8,8	1	23	7,0	
OS	Negativo	21,6	25	17,1	0	53	11,8	0,959
	Positivo	21,2	18,5	14,4	9	49	11,5	
WT	Negativo	60.304	55.110	43.870	7.003	141.152	30.400	0,106
	Positivo	109.056	130.984	71.098	18.624	192.833	42.015	
ET	Negativo	13.697	10.232	17.225	0	52.596	11.936	0,110
	Positivo	34.627	35.571	31.712	0	80.419	18.740	
NC	Negativo	7.164	2.143	15.055	0	44.112	10.433	0,204
	Positivo	20.582	6.852	25.557	0	73.487	15.103	
ED	Negativo	35.060	44.444	28.013	0	69.469	18.302	0,550
	Positivo	45.564	29.287	46.019	0	130.984	25.016	

FONTE: A autora (2022).

NOTA: PFS – *progress free survival*; OS – *overall survival*; WT – *whole tumor*; ET – *enhancing tumor*; NC – *necrotic core*; ED – edema; DP – desvio padrão; Min – mínimo, Max – máximo; IC – intervalo de confiança.

Teste T de Student.

Os pacientes com mutação ATRX tiveram menor sobrevida livre de doença (14,8 *versus* 29,2 meses, $p = 0,008$) e menor sobrevida geral em comparação com os pacientes não mutados (15,2 *versus* 31,4 meses, $p = 0,003$) (TABELA 19). Não houve correlação com os parâmetros de volumetria

TABELA 19 – ATRX *versus* PROGNÓSTICO E VOLUMETRIA

ATRX		Média	Mediana	DP	Min	Max	IC	p
PFS	Negativo	29,2	27	6,1	25	40	5,4	0,008
	Positivo	14,8	18	9,8	1	27	5,3	
OS	Negativo	31,4	28	5,9	26	40	5,2	0,003
	Positivo	15,2	18	9,5	0	31	5,2	
WT	Negativo	108.762	107.448	54.923	39.727	186.801	43.946	0,657
	Positivo	94.718	87.386	70.023	7.003	241.853	30.688	
ET	Negativo	9.696	918	14.735	0	33.652	11.791	0,256
	Positivo	22.879	10.521	26.276	0	80.419	11.516	
NC	Negativo	2.281	126	4.062	0	10.139	3.250	0,327
	Positivo	10.671	1.848	20.139	0	73.487	8.826	
ED	Negativo	72.588	74.178	64.975	0	149.852	45.025	0,439
	Positivo	55.608	48.612	47.520	-11.863	176.300	19.857	

FONTE: A autora (2022).

NOTA: PFS – *progress free survival*; OS – *overall survival*; WT – *whole tumor*; ET – *enhancing tumor*; NC – *necrotic core*; ED – edema; DP – desvio padrão; Min – mínimo, Max – máximo; IC – intervalo de confiança.
Teste T de Student.

O Ki-67 encontrava-se disponível em 48 pacientes, com mínimo de 2%, máximo de 80% e média de 39% ($\pm 21\%$), sem relação com o KPS inicial ($p = 0,25$), PFS ($p = 0,388$), OS ($p = 0,174$) ou número de lesões ($p = 0,871$).

Índices maiores de Ki-67 estavam relacionados à maiores volumes tumorais totais, de realce e edema, sem correlação com os volumes de necrose (TABELA 20).

TABELA 20 – Ki-67 *versus* VOLUMETRIA

		Ki67
WT	Corr (r)	0,4302
	P-valor	0,008
ET	Corr (r)	0,1933
	P-valor	0,03
NC	Corr (r)	0,0073
	P-valor	0,96
ED	Corr (r)	0,4847
	P-valor	< 0,001

FONTE: A autora (2022).

NOTA: Corr (r) – correlação de Pearson.

Porcentagens maiores de Ki67 estavam relacionados à maiores índices de fluorescência residual das margens operatórias ao 5-ALA ($p = 0,01$, TABELA 21).

TABELA 21 – Ki-67 E MARGENS 5-ALA

Margens	Ki67 < 20% (n = 11)	Ki67 > 20% (n = 37)	p
5-ALA livre	9	17	0,01
Fluorescência residual 5-ALA	2	20	

FONTE: A autora (2021).

NOTA: *Cutoff* de 20% (segundo dados da literatura).

4.4 PROGNÓSTICO

O KPS inicial (n = 67) teve média de $76,42 \pm 7,92$, com mínimo de 30 e máximo de 80. Índices de KPS maiores eram relacionados à maior PFS (p = 0,049) e maior OS (p = 0,035).

Os demais parâmetros prognósticos de sobrevida encontram-se descritos na tabela abaixo (TABELA 22).

TABELA 22 – SOBREVIDA EM MESES		
	PFS (n = 64)	OS (n = 70)
Mínimo	0	0
Máximo	41	57
Média	$11,13 \pm 9,87$	$18,01 \pm 14,35$
Mediana	8	13,5

FONTE: A autora (2022).

NOTA: PFS – *progress free survival*; OS – *overall survival*.

O número de lesões não demonstrou repercussão sobre PFS (p = 0,395) ou OS (p = 0,462).

Não houve correlação dos volumes tumorais com a sobrevida (TABELA 23).

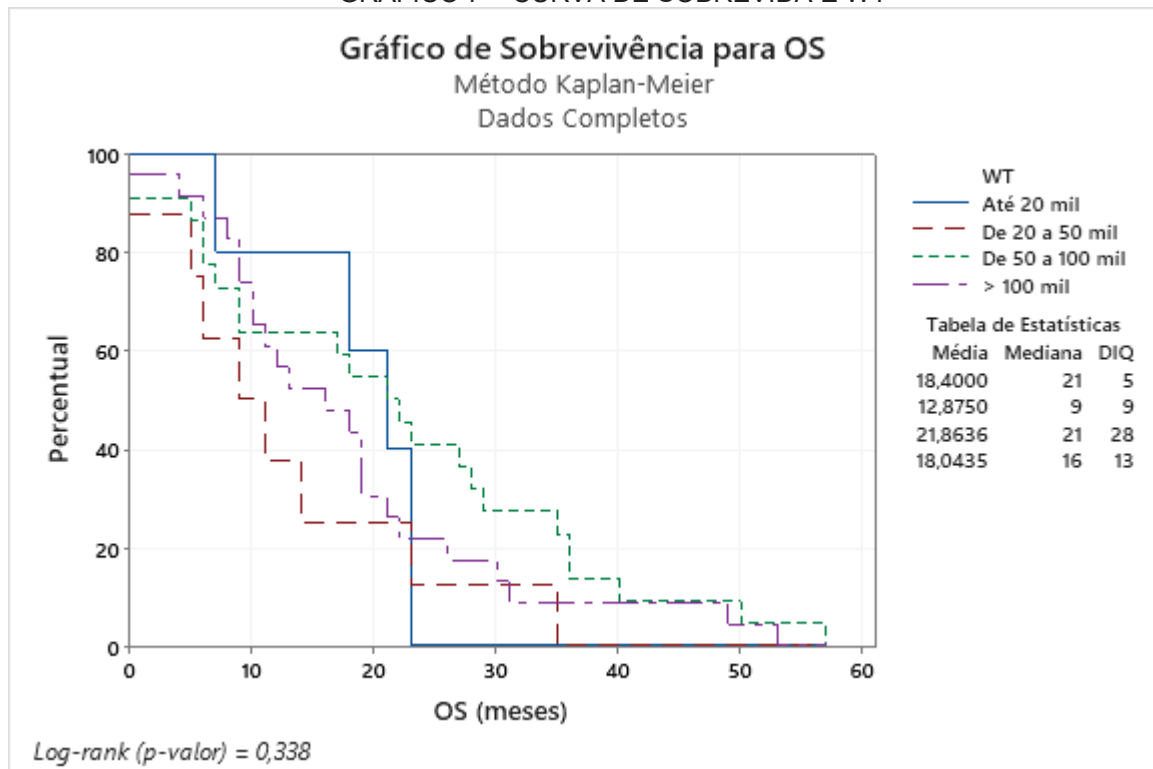
TABELA 23 – VOLUMETRIA VERSUS PROGNÓSTICO			
		PFS	OS
WT	Corr (r)	-0,106	-0,008
	P-valor	0,452	0,953
ET	Corr (r)	-0,236	-0,089
	P-valor	0,086	0,499
NC	Corr (r)	-0,023	0,171
	P-valor	0,868	0,192
ED	Corr (r)	-0,006	-0,062
	P-valor	0,967	0,636

Fonte: A autora (2022).

NOTA: Corr (r) – Correlação de Pearson; WT – *whole tumor*; ET – *enhancing tumor*; NC – *necrotic core*; ED – edema.

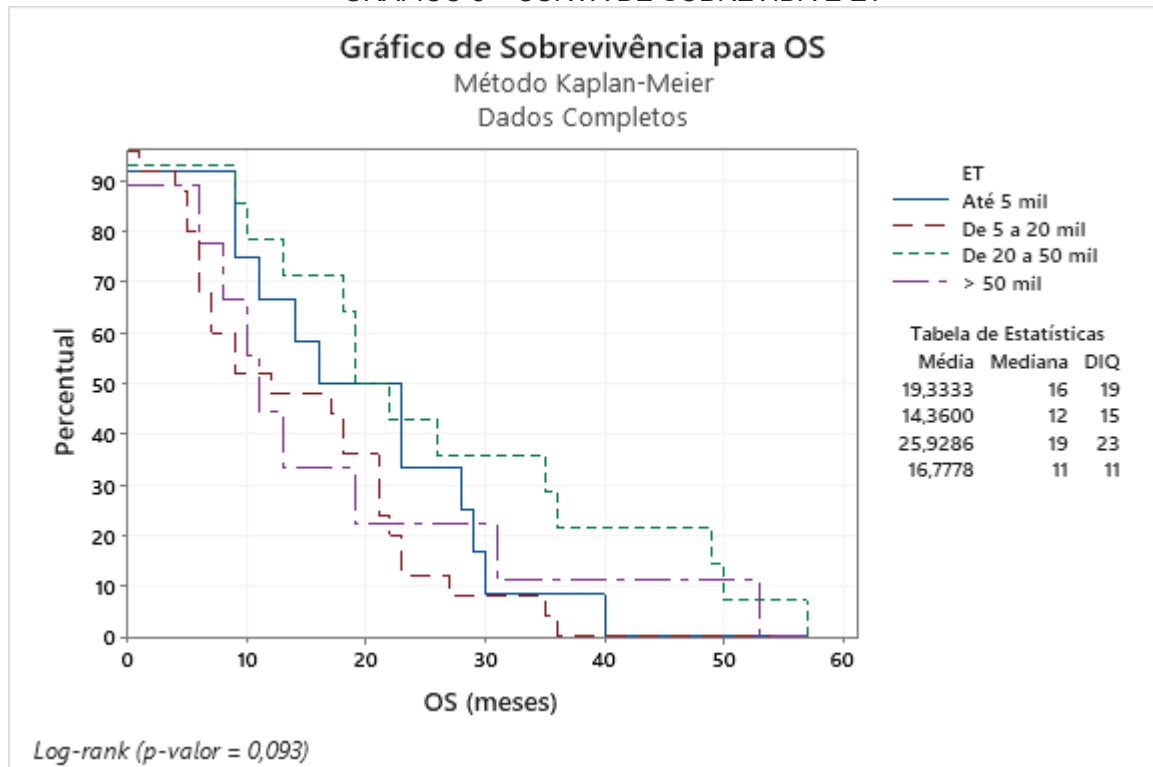
Além da análise segundo a correlação de Person, as volumetrias foram divididas em quatro grupos com número igual de pacientes e em ordem crescente de volume em mm³ para cada um dos parâmetros (WT, ET, ED e NC), sem diferença significativa de sobrevida entre os grupos (p = 0,338; p = 0,093; p = 0,328 e p = 0,659, respectivamente) (GRÁFICOS 7, 8, 9 e 10).

GRÁFICO 7 – CURVA DE SOBREVIDA E WT



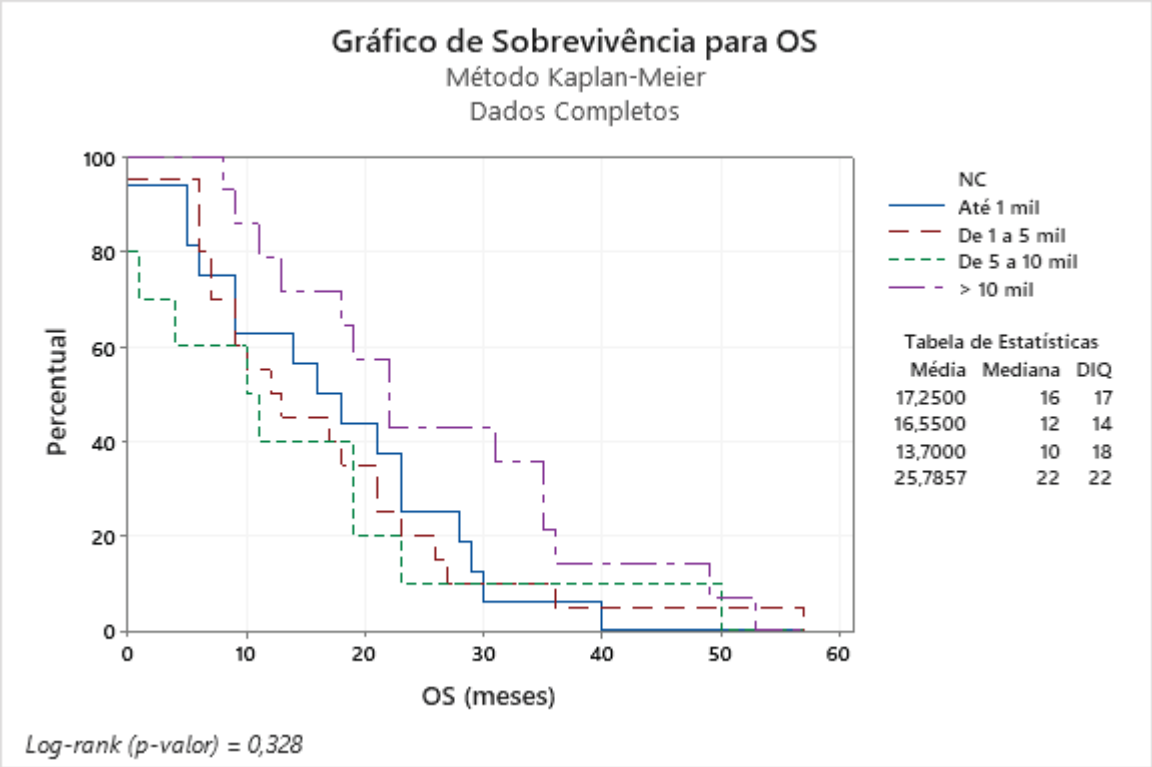
FONTE: A autora (2022).

GRÁFICO 8 – CURVA DE SOBREVIDA E ET



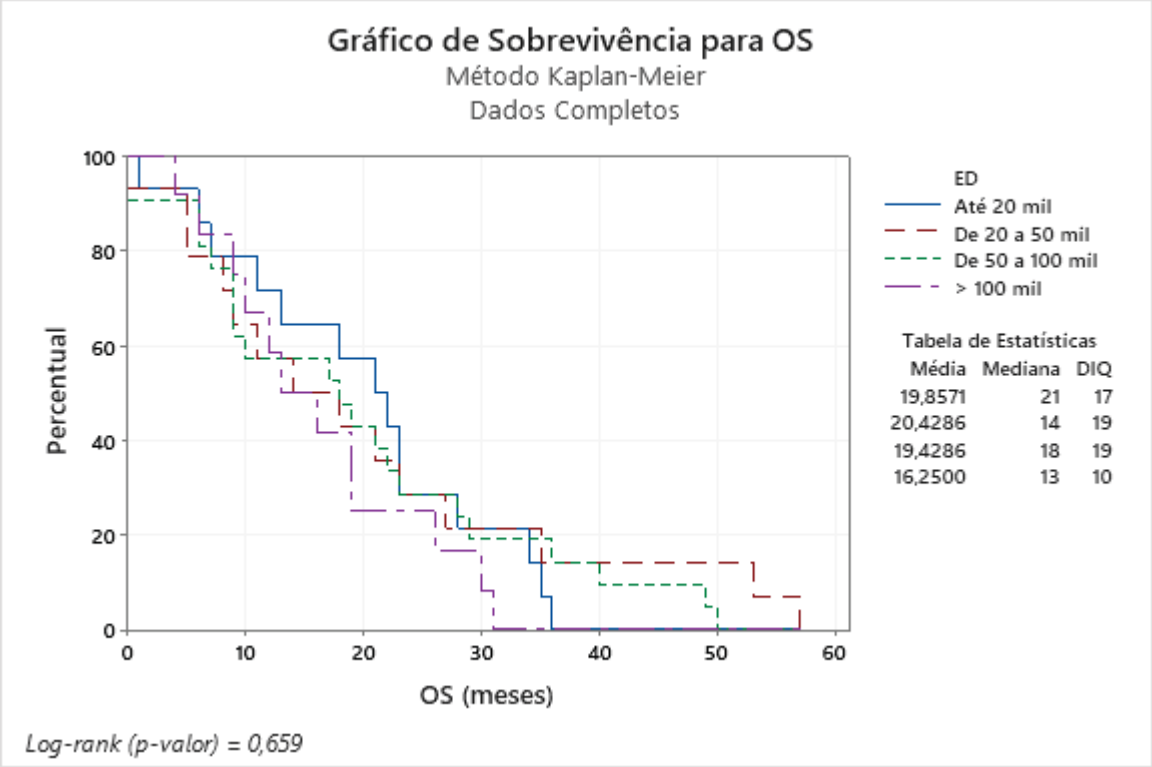
FONTE: A autora (2022).

GRÁFICO 9 – CURVA DE SOBREVIDA E NC



FONTE: A autora (2022).

GRÁFICO 10 – CURVA DE SOBREVIDA E ED



FONTE: A autora (2022).

4.5 ARTIGO PUBLICADO

A amostra deste estudo foi incluída em estudo multicêntrico liderado pelo *Center for Biomedical Image Computing & Analytics* (CBICA) da *Perelman School of Medicine* da *University of Pennsylvania*. O estudo foi inovador na modalidade de IA conhecida como *federated learning* em neuro-oncologia, baseando-se em descentralizar o treinamento de algoritmos, excluindo a necessidade de compartilhamento de dados e imagens de pacientes.

Trata-se da maior pesquisa de segmentação em gliomas, com uma amostra de 6314 casos de mais de 70 centros, sendo a nossa contribuição a única no Brasil.

Seus resultados foram publicados em dezembro de 2022 na *Nature Communications* (ANEXO II).

4.6 APRESENTAÇÃO DOS RESULTADOS PARCIAIS

Os resultados parciais desta pesquisa foram apresentados no *European Congress of Radiology* em Viena em julho de 2022, recebendo bolsa dentro do programa *Invest in the Youth*, que promove auxílio financeiro para jovens pesquisadores com trabalhos relevantes apresentados no congresso (ANEXO III).

4.7 ARTIGO SUBMETIDO

Os resultados deste estudo também foram submetidos à revista *World Neurosurgery* em dezembro de 2022 (ANEXOS IV e V). Conforme sugestões de revisores em submissões prévias, optamos por homogeneizar a amostra, excluindo os pacientes IDH mutados e os submetidos à terapia adjuvante com bevacizumabe.

5 DISCUSSÃO

Nossa amostra foi composta por 91 pacientes com astrocitomas grau 4 e glioblastomas, com predomínio do sexo masculino ($n = 56$, 62%), com mediana de idade de 58 anos. As lesões predominavam no compartimento supratentorial, sobretudo nos lobos frontal e temporal, apresentando-se à imagem com marcada heterogeneidade, áreas de necrose, sinais de proliferação celular à espectroscopia e com perfusão quente. Pouco menos da metade da amostra ($n = 40$, 44%) foi submetida à ressecção completa, estes pacientes apresentando maior sobrevida geral ($p = 0,004$). A maior parte dos pacientes contou com auxílio do 5-ALA ($n = 76$), a discreta maioria deles sem fluorescência residual no leito operatório ($n = 41$, 54%), este grupo também apresentando maior sobrevida ($p = 0,002$).

Quanto ao perfil histopatológico, lesões IDH selvagem foram mais prevalentes ($n = 55$), com pior sobrevida ($p = 0,049$). Mutações ATRX ($n = 38$) estavam ligadas à menor sobrevida livre de doença e geral ($p = 0,008$ e $0,003$, respectivamente). Índices Ki67 variaram amplamente, entre 2 e 80%, porém sem correlação com prognóstico.

No tocante aos objetivos deste estudo, não encontramos correlação entre as volumetrias tumorais e a sobrevida, ou com os diferentes tipos de mutação analisados (IDH, MGMT e ATRX). Entretanto, observamos associação entre índices Ki67 e os volumes de áreas metabolicamente ativas do tumor.

Sendo assim, rejeitamos a hipótese alternativa de que tumores maiores, com maiores componentes sólidos e de necrose cursam com menor sobrevida na presente amostra.

Quanto aos objetivos secundários, não observamos correlação da volumetria com o tipo de ressecção ou fluorescência 5-ALA residual no leito operatório.

Retomando os estudos resumidos na tabela 5 (TABELA 5), que analisaram a volumetria e sua correlação com o prognóstico, destacamos a grande heterogeneidade metodológica.

Lacroix *et al* utilizou uma amostra bastante significativa ($n = 416$) para correlacionar aspectos qualitativos e quantitativos de RM pré-operatória com o prognóstico, observando menor sobrevida em graus maiores de realce, edema e necrose. Entretanto, a segmentação foi realizada em sequências 2D, aferindo o maior eixo das lesões nos planos axial e coronal, com extrapolação para cálculo do volume através da espessura do corte. Apesar de realizada a volumetria, mesmo com

metodologia de acurácia questionável, o autor a empregou apenas para o cálculo da porcentagem de lesão residual em comparação com as imagens de RM pós-operatórias, e utilizou escala subjetiva de zero a dois ou três para realce, necrose e edema para correlacionar com a sobrevida (LACROIX et al., 2001). Sendo assim, acreditamos que os achados deste estudo não são comparáveis ao nosso.

Iliadis *et al* também utilizou sequências não volumétricas 2D, com volumetrias poligonais e pouco acuradas, encontrando menor sobrevida em pacientes com volumes maiores de realce e de necrose (ILIADIS et al., 2012).

Por sua vez, Grabowski *et al* empregou imagens volumétricas, com segmentação realizada por neurocirurgião, observando menor sobrevida em volumes maiores de realce (GRABOWSKI et al., 2014).

Finalmente, Wangaryattawanavich *et al* estudou os volumes de edema em FLAIR e realce e necrose em T1 pós-gadolínio, com método de segmentação semiautomática, encontrando menor sobrevida em pacientes com volumes maiores de edema (cutoff de 85.000 mm³, p = 0,019) e realce (cutoff de 35.000 mm³, p = 0,013). A autora não especifica se foram utilizadas sequências volumétrica (WANGARYATTAWANICH et al., 2015).

Ao comparar os resultados acima descritos com o nosso estudo, destacamos como principal diferença a não clareza das sequências utilizadas e seus parâmetros de aquisição, aspecto este, fundamental para a reprodutibilidade dos achados em imagens de RM. Lacroix e Iliadis utilizaram sequências 2D, onde a segmentação gera resultados poligonais e pouco precisos. Incluímos na tabela abaixo (TABELA 24) o resumo dos achados acima discutidos.

TABELA 24 – REVISÃO DA LITERATURA EM PROGNÓSTICO E VOLUMETRIA

Autor	N	ROI	Técnica	Variáveis	Resultados
Lacroix M <i>et al</i> (2001)	416	Realce (0 a II), necrose (0 a III), edema (0 a III)	Seleção da área no axial e extrapolação para volume através da espessura da lesão (MedVision 1.14)	Idade, sexo, KPS, tipo de cirurgia, tratamento adjuvante, status vital, porcentagem de ressecção	Graus maiores de realce ($p = 0,007$), edema ($p = 0,04$) e necrose ($p < 0,001$) ligados a sobrevida inferior.
Iliadis G <i>et al</i> (2011)	65	Volumes de realce, edema e necrose	Semiautomática (Volumio)	Idade, sexo, KPS, tipo de cirurgia, MGMT	Menor sobrevida relacionada a maiores volumes de realce e necrose, sem relação com edema. Mutação MGMT ligada a valores menores de necrose
Grabowski MM <i>et al</i> (2014)	128	Volume de realce (T1Gd+) pré e pós-operatório	Semiautomática (BrainLab)	Idade, sexo, KPS	Menor sobrevida relacionada a maiores volumes de realce pré e pós-operatório
Wangaryattawanich P <i>et al</i> (2015)	94	Volumes de realce, edema e necrose	Semiautomática (3D Slicer)	Idade, sexo, KPS	Menor sobrevida relacionada a maiores volumes de edema (<i>cutoff</i> de 85.000 mm ³ , $p = 0,019$) e realce (<i>cutoff</i> de 35.000 mm ³ , $p = 0,013$)
Nosso estudo	91	Volumes de tumor total (3D FLAIR), edema (3D FLAIR), realce (3D T1 Gd+) e necrose (3D T1 Gd+)	Semiautomática (VISVA by LIDS Unicamp)	Idade, sexo, KPS, tipo de ressecção, fluorescência 5-ALA residual, IDH, MGMT, ATRX e Ki67	Ausência de correlação dos volumes com KPS, tipo de ressecção, fluorescência 5-ALA residual, mutações IDH, MGMT ou ATRX. Volumes maiores de WT ($p = 0,008$), ET ($p = 0,03$), e ED ($p < 0,001$) relacionados a maiores valores de Ki67

FONTE: A autora (2023)

Levantamos como hipótese para a ausência de correlação entre as volumetrias e o prognóstico, o fato de que lesões volumosas nessa amostra não cursaram com maiores taxas de ressecções incompletas. Portanto, ressecções satisfatórias e com margens livres foram obtidas mesmo em tumores grandes, fator este, que sabidamente é o de maior valor prognóstico (COSTA et al., 2019; GANDHI et al., 2019).

Atribui-se em parte a esse resultado, o emprego do 5-ALA, um metabólito fotossensível natural da via da hemoglobina que se acumula em tecidos com aumento da permeabilidade capilar, como gliomas e seus entornos. Usualmente é convertido em protoporfirina IX, mas a expressão reduzida da enzima ferroquelatase em células neoplásicas promove seu acúmulo, que pode ser detectado sob excitação à luz azul, funcionando como guia intraoperatório para a remoção do tumor.

Stummer *et al* demonstraram em 2006 aumento significativo da probabilidade de se atingir ressecções completas (OR = 3.41 [2.03–5.71]; $p < 0,001$) sob fluorescência 5-ALA em comparação com a luz branca convencional, além de promover aumento da sobrevida (HR = 0.73 [0.57–0.94]; $p = 0,01$) (KLINIK et al., 2006b).

Desde a publicação de seus resultados, diversos autores ratificaram seus achados, sendo atualmente o 5-ALA uma ferramenta bem estabelecida para guiar ressecções mais extensas, promovendo melhora da sobrevida (COBURGER; WIRTZ, 2019; HADJIPANAYIS; STUMMER, 2019; HERVEY-JUMPER; BERGER, 2016; KLINIK et al., 2006a; MAHMOUDI et al., 2019; PICART et al., 2019).

O INC é pioneiro no emprego do 5-ALA em suas cirurgias, publicando em 2016 os relatos dos primeiros 41 casos submetidos à cirurgia com 5-ALA na América Latina, mesmo antes da aprovação pela agência americana *Food and Drug Administration* (FDA) em 2017, e acreditamos que este é um grande diferencial da nossa amostra (RAMINA et al., 2016).

Por não possuímos grupo controle ou termos realizado análise da eloquência das áreas ressecadas, este estudo não permite estabelecer definitivamente se a maior incidência de ressecções completas se deu devido ao 5-ALA. Além disso, atualmente, sabe-se que a nomenclatura qualitativa em ressecção completa ou incompleta tem grandes limitações, sendo mais apropriado o cálculo da porcentagem de volume residual a partir da RM pré e pós-operatória (SANAI et al., 2011). Reside neste aspecto

uma das limitações do trabalho, com perspectiva para futura exploração e validação dos achados.

O KPS, uma ferramenta que afere a funcionalidade do paciente, está também intimamente relacionada ao prognóstico. Pacientes com maior independência para as atividades diárias possuem prognóstico melhor, achado este, confirmado pela nossa amostra para a sobrevida geral ($p = 0,035$) e livre de doença ($p = 0,049$), (COSTA et al., 2019). No entanto, não encontramos correlação entre os valores de KPS e os volumes tumorais, diferentemente ao esperado, em que lesões maiores cursassem com menores performances de funcionalidade.

A classificação molecular das lesões também é fator prognóstico decisivo, e reside neste aspecto uma limitação deste trabalho, visto que o perfil imunohistoquímico estava indisponível em uma parte significativa da amostra. Além disso, durante o desenvolvimento do estudo foi publicada a nova classificação da OMS, impactando diretamente a definição de glioblastoma e demandando reestruturação metodológica.

Inicialmente incluímos lesões classificadas como glioblastoma segundo a classificação da OMS de 2016, apesar de conhecida a diferença prognóstica entre as lesões apresentando mutação IDH. A partir de junho de 2016, estas lesões passam a ser definidas astrocitoma grau 4, sendo que o glioblastoma deve ser obrigatoriamente IDH selvagem. Contudo, optamos por manter estes pacientes na amostra, adaptando-os à nova nomenclatura.

Conforme esperado, os pacientes IDH mutados apresentaram sobrevida superior na nossa amostra ($p = 0,049$) (HAN et al., 2020; THAKKAR et al., 2014). Não encontramos correlação entre o *status* IDH e as volumetrias tumorais, ao contrário dos achados descritos por Lee *et al*, em que lesões IDH mutadas apresentaram volumes menores de realce e necrose em comparação com os tumores IDH selvagens (LEE et al., 2019).

Sabe-se que a mutação MGMT está relacionada a melhor prognóstico devido a melhor resposta à quimioterapia, aspecto este, não observado na nossa amostra. A mutação MGMT também não demonstrou correlação com as volumetrias, diferentemente do observado por Iliadis *et al*, que demonstrou que pacientes MGMT mutados possuíam volumes menores de necrose (COSTA et al., 2019; ILIADIS et al., 2012). Contudo, ressalta-se que o autor utilizou sequências 2D e segmentação com *manual tracer*, resultando em labels poligonais e pouco acurados.

Assim como o IDH, sabe-se que a mutação ATRX atrela pior prognóstico, fato este, ratificado pela nossa amostra, que demonstrou que pacientes ATRX mutados possuíam sobrevidas livre de doença e geral menores ($p = 0,008$ e $0,003$, respectivamente), porém, sem correlação com a volumetria (NANDAKUMAR; MANSOURI; DAS, 2017).

O Ki67 é um marcador de proliferação celular, comumente elevado em gliomas. Na era da investigação genética dos gliomas, e devido aos achados deste estudo, decidimos revisitar o papel desta ferramenta, amplamente acessível, e sua correlação com volumetria e prognóstico em uma coorte submetida a ressecção com 5-ALA.

O marcador recebeu seu nome após a sua primeira descrição por Gerdes *et al* na cidade de Kiel, na Alemanha, com a descoberta de que suas 67 cópias estavam relacionadas a proliferação celular em ratos com linfoma de Hodgkin. Consiste em um anticorpo monoclonal IgG que opera como um marcador de proliferação celular. No contexto dos astrocitomas, sua porcentagem é relacionada a graus progressivamente maiores de malignidade e severidade da doença (GERDES *et al.*, 1983; HU *et al.*, 2013; LINN JOHANNESSEN; TORP, 2006).

Na nossa amostra, os índices de Ki67 não tiveram correlação com a idade ou escores de KPS, em consonância com estudos prévios (ARMOCIDA *et al.*, 2020; HENKER *et al.*, 2019)

Quanto à volumetria, encontramos uma correlação positiva entre o índice de Ki67 e os volumes de tumor total ($p = 0,008$), realce ($p = 0,03$) e edema ($p < 0,001$). Sendo assim, quanto maiores os níveis de proliferação celular, maiores os volumes dos componentes metabolicamente ativos da lesão.

Contudo, a literatura ainda é escassa no tocante ao impacto do Ki67 em achados quantitativos de RM pré-operatória.

Henker *et al* estudou os volumes de realce, necrose e edema peritumoral e não encontrou correlação com os índices de Ki67. O autor propôs que a ausência de correlação em sua amostra se deu devido ao tempo de aquisição das imagens de RM no momento das primeiras manifestações clínicas, e consequentemente muito antes da cirurgia, o que pode não ter transmitido de maneira fiel o aumento volumétrico ao longo do tempo (HENKER *et al.*, 2019). Nós nos atentamos à essa limitação, e utilizamos imagens pré-operatórias com no máximo três semanas antes da

manipulação cirúrgica, refletindo assim, de maneira mais próxima as características do espécime.

Outra grande diferença entre o estudo de Henker e o nosso, é que o primeiro subtraiu as áreas císticas do volume tumoral total e análise. A fisiopatologia da formação de cistos em gliomas é ainda questionável, com a maioria das teorias estabelecendo uma quebra da barreira hematoencefálica e efusão hiperproteica. Áreas císticas apresentam-se como focos homogêneos, sem realce e com paredes regulares e finas, com conteúdo com características semelhantes ao líquido, contrastando com áreas de necrose, que geralmente são heterogêneas e com margens irregulares (CHOI et al., 2010; CURTIN et al., 2020; LOHLE et al., 2000). Apesar de maior frequência em gliomas de baixo grau e associação com melhor prognóstico, nenhuma característica definitiva de imagem à RM permite diferenciar definitivamente áreas císticas sem realce de necrose tumoral. Sendo assim, optamos por incluir estas regiões na segmentação.

Caramanti *et al.*, por outro lado, encontrou associação entre Ki67 e volumes de edema peritumoral no FLAIR, mas curiosamente, não conforme esperado na área de realce, esperada por haver maior celularidade (CARAMANTI et al., 2022). Este achado pode sugerir que nestes casos, anormalidades de sinal peritumoral estão mais relacionadas à microinvasão que edema vasogênico.

Armocida *et al.* também estudou volumetria de glioblastomas e encontrou uma correlação forte entre o volume das lesões e índices de Ki67 em tumores grandes, com um *cutoff* de 45 cm³. Contudo, o autor não especifica quais parâmetros ou sequências foram usados para a segmentação. Nas imagens nota-se que a segmentação foi feita em sequências 2D ponderadas em T1 sem gadolínio, (ARMOCIDA et al., 2020).

Apesar de se tratarem de amostras semelhantes e homogêneas, os achados que correlacionam Ki67 e volumetria são conflitantes. Henker *et al.* foi o único autor a utilizar técnica semelhante e confiável de segmentação semiautomática baseada em *contour expansion* que leva em conta mínimas variações na intensidade de pixels e os contornos da lesão.

Caramanti *et al.* and Armocida *et al.* utilizaram técnicas com *manual tracing*, o que não promove um contorno delicado da lesão e se baseia em geração automática de ROIs faltantes entre os cortes segmentados, resultando em uma segmentação poligonal e não acurada. Estes autores também não especificam as sequências

utilizadas, nos fazendo questionar se imagens 2D são apropriadas para segmentação semiautomática.

Nós enfatizamos como aspecto forte do nosso estudo o emprego apenas de sequências volumétricas 3D FLAIR e T1 pós-gadolínio, com um método de segmentação validado e confiável que utiliza *region growing* a partir de uma semente sinalizadora. Sendo assim, acreditamos que encontramos os resultados esperados apesar da amostra limitada.

Conforme proposto previamente por Henker *et al*, segmentação volumétrica 3D é a única técnica confiável, postulando que técnicas cuboides, elipsoides ou esferoides não devem ser usadas para este propósito. Ferramentas de medida não acuradas são provavelmente o principal motivo para achados da literatura inconsistentes, reforçando que a segmentação 3D é mais precisa e superior aos outros métodos (HENKER *et al.*, 2017).

Ainda no tocante ao Ki67, em nossa amostra os pacientes com fluorescência residual das margens operatórias possuíam índices significativamente maiores de Ki67 quando comparado com margens 5-ALA livres ($p = 0,01$). O Ki67 é mais elevado na interface entre a porção sólida do tumor e o tecido circunjacente (EIDEL *et al.*, 2017). Este fenômeno pode explicar tais achados, visto que a amostragem do Ki67 foi obtida nas margens mais infiltrativas e com fluorescência mais positiva.

Apesar de partirem de mecanismos bioquímicos completamente diferentes, a fluorescência 5-ALA ocorrendo devido à permeabilidade capilar, *downregulation* da expressão da ferroquelatase e acúmulo citosólico de protoporfirina IX, e o Ki67 refletindo atividade nuclear relacionada a mitose e proliferação celular, existe uma correlação bem estabelecida entre o grau de fluorescência 5-ALA à fotometria quantitativa e os índices Ki67 em astrocitomas (SHIMIZU *et al.*, 2000; VALDÉ S *et al.*, 2011). No nosso conhecimento, este é o primeiro estudo de volumetria tumoral em uma amostra submetida a cirurgia assistida por 5-ALA, e que correlaciona as margens operatórias e segmentação volumétrica com Ki67.

Além da natureza retrospectiva do desenho do estudo, sua principal limitação é o aspecto subjetivo da amostragem e interpretação do Ki67, dada a extensa heterogeneidade intralesional. Variação interobservador também é um problema do Ki67, que não se limita aos gliomas. Limitações semelhantes são válidas para o 5-ALA, visto que não utilizamos análise fotométrica quantitativa.

Dentre os potenciais deste estudo, destacam-se a amostra significativa, semelhante ou maior a maior parte dos demais artigos publicados com tema semelhante, além do emprego de imagens de RM 3D, com um método preciso de segmentação.

Os resultados desta pesquisa colocam em voga o real papel do volume das lesões no prognóstico dos gliomas. Apesar de volumosas, lesões com perfil molecular favorável e adequadamente ressecadas, destacando-se o auxílio do 5-ALA, possuem prognóstico menos sombrio que o esperado.

Acreditamos que os próximos passos incluem a validação e reprodutibilidade desses achados em uma amostra multicêntrica com populações diferentes, imagens adquiridas em equipamentos e com protocolos de aquisição diferentes, bem como com métodos de segmentação mais recentes e totalmente automáticos.

Sendo assim, justificam-se cada vez menos as técnicas inconsistentes e dispendiosas de tempo para segmentação. Os métodos semiautomáticos promoveram avanços significativos à segmentação manual. Similarmente, o crescimento de algoritmos de *machine learning* para segmentação em neuroimagem tem sido exponencial. Apesar da acurácia satisfatória de alguns algoritmos, sua implantação em larga escala ainda é limitada devido aos resultados controversos, heterogeneidade da literatura e diretrizes incertas (VAN KEMPEN et al., 2021).

Visando transpor as barreiras éticas e legais do compartilhamento de imagens, limitando o treinamento de algoritmos com grandes amostras, destacamos também o papel do *federated learning* como perspectiva futura. Trata-se de estratégia de descentralização, onde um algoritmo é treinado por cada centro individualmente, cada um com sua amostra, com suas subseqüentes correções. Desse modo, promete ampliar os horizontes dos algoritmos de *deep learning*, e simplificar o treinamento dos mesmos, permitindo amostras cada vez maiores, sem as limitações ético-legais.

CONCLUSÃO

Não encontramos correlação da volumetria tumoral com os aspectos clínicos prognósticos de sobrevida ou com as mutações IDH, MGMT e ATRX. Existe uma correlação entre Ki67, os volumes tumorais metabolicamente ativos.

Quanto aos aspectos cirúrgicos, não houve associação entre os volumes das subáreas tumorais e o tipo de ressecção ou fluorescência 5-ALA residual no leito operatório. Entretanto, índices maiores de Ki67 estavam relacionados a maior incidência de fluorescência residual ao 5-ALA.

REFERÊNCIAS

AGNIHOTRI, S. et al. Glioblastoma, a brief review of history, molecular genetics, animal models and novel therapeutic strategies. **Archivum Immunologiae et Therapiae Experimentalis**, v. 61, n. 1, p. 25–41, fev. 2013.

ALBRIGHT RE; FRAM EK. Microcomputer-Based Technique for 3-D Reconstruction and Volume Measurement of Computed Tomographic Images - Part 2: Anaplastic Primary Brain Tumors. **Investigative Radiology**, v. 23, n. 12, p. 886–890, 1988a.

ALBRIGHT RE; FRAM EK. Microcomputer-Based Technique for 3-D Reconstruction and Volume Measurement of Computed Tomographic Images - Part 1: Phantom Studies. **Investigative Radiology**, v. 23, n. 12, p. 881–885, 1988b.

ANDREOU, J. et al. CT Prognostic Criteria of Survival after Malignant Glioma Surgery. **AJNR**, v. 4, p. 488–490, 1983b.

APPAY, R. et al. CDKN2A homozygous deletion is a strong adverse prognosis factor in diffuse malignant IDH-mutant gliomas. **Neuro-Oncology**, v. 21, n. 12, p. 1519–1528, 1 dez. 2019.

ARMOCIDA, D. et al. Is Ki-67 index overexpression in IDH wild type glioblastoma a predictor of shorter Progression Free survival? A clinical and Molecular analytic investigation. **Clinical Neurology and Neurosurgery**, v. 198, 1 nov. 2020.

BONDY, M. L. et al. Brain tumor epidemiology: Consensus from the Brain Tumor Epidemiology Consortium. **Cancer**, v. 113, n. 7, p. 1953–1968, 1 out. 2008.

BRITO, C. et al. Clinical insights gained by refining the 2016 WHO classification of diffuse gliomas with: EGFR amplification, TERT mutations, PTEN deletion and MGMT methylation. **BMC Cancer**, v. 19, n. 1, 17 out. 2019.

CARAMANTI, R. et al. Is Edema Zone Volume Associated With Ki-67 Index in Glioblastoma Patients? **Cureus**, v. 14, n. 4, 18 abr. 2022.

CHARTRAND, G. et al. Deep learning: A primer for radiologists. **Radiographics**, v. 37, n. 7, p. 2113–2131, 1 nov. 2017.

CHEN, W. J. et al. Ki-67 is a valuable prognostic factor in gliomas: Evidence from a systematic review and meta-analysis. **Asian Pacific Journal of Cancer Prevention**, v. 16, n. 2, p. 411–420, 2015.

CHOI, S. J. et al. The Radiologic Features of Cystic versus Noncystic Glioblastoma Multiforme as Significant Prognostic Factors. **J Korean Soc Radiol**, v. 63, p. 299–306, 2010.

COBURGER, J.; WIRTZ, C. R. Fluorescence guided surgery by 5-ALA and intraoperative MRI in high grade glioma: a systematic review. **Journal of Neuro-Oncology**, v. 141, n. 3, p. 533–546, 15 fev. 2019.

COSTA, E. et al. Long-term survival after glioblastoma resection: hope despite poor prognosis factors. **Journal of Neurosurgical Sciences**, v. 63, n. 3, p. 251–257, 2019.

CURTIN, L. et al. Assessment of Prognostic Value of Cystic Features in Glioblastoma Relative to Sex and Treatment With Standard-of-Care. **Frontiers in Oncology**, v. 10, 16 nov. 2020.

DAR, A. S.; PADHA, D. Medical Image Segmentation A Review of Recent Techniques, Advancements and a Comprehensive Comparison. **International Journal of Computer Sciences and Engineering**, v. 7, n. 7, p. 114–124, 31 jul. 2019.

DEL MAR INDA, M. et al. Chromosomal abnormalities in human glioblastomas: Gain in chromosome 7p correlating with loss in chromosome 10q. **Molecular Carcinogenesis**, v. 36, n. 1, p. 6–14, 1 jan. 2003.

EIDEL, O. et al. Tumor infiltration in enhancing and non-enhancing parts of glioblastoma: A correlation with histopathology. **PLoS ONE**, v. 12, n. 1, 1 jan. 2017.

ESSIG, M. et al. Perfusion MRI: The five most frequently asked technical questions. **American Journal of Roentgenology**, v. 200, n. 1, p. 24–34, jan. 2013.

FALCÃO, A. X.; BERGO, F. P. G. Interactive volume segmentation with differential image foresting transforms. **IEEE Transactions on Medical Imaging**, v. 23, n. 9, p. 1100–1108, set. 2004.

FONOV, V. et al. Unbiased average age-appropriate atlases for pediatric studies. **NeuroImage**, v. 54, n. 1, p. 313–327, 1 jan. 2011.

GANDHI, S. et al. Survival outcomes among patients with high-grade glioma treated with 5-aminolevulinic acid–guided surgery: A systematic review and meta-analysis. **Frontiers in Oncology**, v. 9, 2019.

GERDES, J. et al. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. **Int J Cancer**, v. 31, p. 13–20, 1983.

GRABOWSKI, M. M. et al. Residual tumor volume versus extent of resection: Predictors of survival after surgery for glioblastoma. **Journal of Neurosurgery**, v. 121, n. 5, p. 1115–1123, 1 nov. 2014.

GUO, C. F. et al. Significance of tumor protein p53 mutation in cellular process and drug selection in brain lower grade (WHO grades II and III) glioma. **Biomarkers in Medicine**, v. 14, n. 12, 1 ago. 2020.

HADJIPANAYIS, C. G.; STUMMER, W. 5-ALA and FDA approval for glioma surgery. **Journal of Neuro-Oncology**, v. 141, n. 3, p. 479–486, 15 fev. 2019.

HAMMOUD, M. A. et al. Prognostic significance of preoperative MRI scans in glioblastoma multiforme. **Journal of Neuro-Oncology**, v. 27, p. 65–73, 1996a.

HAN, S. et al. IDH mutation in glioma: molecular mechanisms and potential therapeutic targets. **British Journal of Cancer**, v. 122, n. 11, p. 1580–1589, 26 maio 2020.

HANIF, F. et al. Glioblastoma multiforme: A review of its epidemiology and pathogenesis through clinical presentation and treatment. **Asian Pacific Journal of Cancer Prevention**, v. 18, n. 1, p. 3–9, 2017.

HEGI, M. E. et al. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. **The New England Journal of Medicine**, v. 352, n. 10, p. 997–1003, 2002.

HENKER, C. et al. Volumetric quantification of glioblastoma: experiences with different measurement techniques and impact on survival. **Journal of Neuro-Oncology**, v. 135, n. 2, p. 391–402, 1 nov. 2017.

HENKER, C. et al. Correlation of Ki-67 Index with Volumetric Segmentation and its Value as a Prognostic Marker in Glioblastoma. **World Neurosurgery**, v. 125, p. e1093–e1103, 1 maio 2019.

HERVEY-JUMPER, S. L.; BERGER, M. S. Maximizing safe resection of low- and high-grade glioma. **Journal of Neuro-Oncology**, v. 130, n. 2, p. 269–282, 1 nov. 2016.

HU, K. et al. Brain Tumor Segmentation Using Multi-Cascaded Convolutional Neural Networks and Conditional Random Field. **IEEE Access**, v. 7, p. 92615–92629, 2019.

HU, X. et al. Expression of p53, epidermal growth factor receptor, Ki-67 and O6-methylguanine-DNA methyltransferase in human gliomas. **Oncology Letters**, v. 6, n. 1, p. 130–134, jul. 2013.

ILIADIS, G. et al. Volumetric and MGMT parameters in glioblastoma patients: Survival analysis. **BMC Cancer**, v. 12, n. 3, 2012.

INSTITUTO NACIONAL DO CÂNCER (INCA). Estimativa de Incidência de Câncer no Brasil 2020. **Instituto Nacional do Câncer (INCA) e Ministério da Saúde**, 2019.

JIN, W. et al. Artificial intelligence in glioma imaging: Challenges and advances. **Journal of Neural Engineering**, v. 17, n. 2, 1 abr. 2020.

KATSIGIANNIS, S. et al. MGMT-Positive vs MGMT-Negative Patients with Glioblastoma: Identification of Prognostic Factors and Resection Threshold. **Neurosurgery**, v. 88, n. 4, p. E323–E329, 1 abr. 2021.

KLEESIEK, J. et al. Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. **NeuroImage**, v. 129, p. 460–469, 1 abr. 2016.

KLEIHUES, P.; BURGER, P. C.; SCHEITHAUER, B. W. The New WHO Classification of Brain Tumours. **Brain Pathology**, v. 3, p. 255–268, 1993.

KLINIK, N. et al. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. **Lancet Oncol**, v. 7, p. 392–401, 2006a.

LACROIX, M. et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. **J Neurosurg**, v. 95, p. 190–198, 2001.

LEE, M. H. et al. Prediction of IDH1 Mutation Status in Glioblastoma Using Machine Learning Technique Based on Quantitative Radiomic Data. **World Neurosurgery**, v. 125, p. e688–e696, 1 maio 2019.

LEVIN, V. A. et al. Prognostic significance of the pretreatment CT scan on time to progression for patients with malignant gliomas. **J Neurosurg**, v. 52, p. 642–647, 1980a.

LINN JOHANNESSEN, A.; TORP, H. The Clinical Value of Ki-67/MIB-1 Labeling Index in Human Astrocytomas. **Pathology Oncology Research**, v. 12, n. 3, 2006.

LOHLE, P. N. M. et al. On the pathogenesis of brain tumor cysts: a volumetric study of tumour, edema and cyst. **Neuroradiology**, v. 42, p. 639–642, 2000.

LOTAN, E. et al. State of the art: Machine learning applications in Glioma Imaging. **American Journal of Roentgenology**, v. 212, n. 1, p. 26–37, 1 jan. 2019.

LOUIS, D. N. et al. The 2007 WHO classification of tumours of the central nervous system. **Acta Neuropathologica**, v. 114, n. 2, p. 97–109, ago. 2007.

LOUIS, D. N. et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. **Acta Neuropathologica**, v. 131, n. 6, p. 803–820, 1 jun. 2016.

LOUIS, D. N. et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. **Neuro-Oncology**, v. 23, n. 8, p. 1231–1251, 2 ago. 2021.

LU, V. M. et al. The prognostic significance of CDKN2A homozygous deletion in IDH-mutant lower-grade glioma and glioblastoma: a systematic review of the contemporary literature. **Journal of Neuro-Oncology**, v. 148, n. 2, p. 221–229, 1 jun. 2020.

MAHMOUDI, K. et al. 5-aminolevulinic acid photodynamic therapy for the treatment of high-grade gliomas. **Journal of Neuro-Oncology**, v. 141, n. 3, p. 595–607, 15 fev. 2019.

MARTINS, S. B. et al. An adaptive probabilistic atlas for anomalous brain segmentation in MR images. **Medical Physics**, v. 46, n. 11, p. 4940–4950, 1 nov. 2019.

NANDAKUMAR, P.; MANSOURI, A.; DAS, S. The role of ATRX in glioma biology. **Frontiers in Oncology**, v. 7, 29 set. 2017.

NOROUZI, A. et al. Medical image segmentation methods, algorithms, and applications. **IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)**, v. 31, n. 3, p. 199–213, 2014.

OMURO, A.; DEANGELIS, L. M. Glioblastoma and other malignant gliomas: A clinical review. **JAMA - Journal of the American Medical Association**, v. 310, n. 17, p. 1842–1850, 6 nov. 2013.

OSBORN, A. G.; HEDLUND, G.; SALZMAN, K. L. **Osborn's Brain**. 1st. ed. [S.l.]: Elsevier, 2012.

ÖZ, G. et al. Clinical proton MR spectroscopy in central nervous system disorders. **Radiology**, v. 270, n. 3, p. 658–679, 5 mar. 2014.

PERRY, A.; WESSELING, P. Histologic classification of gliomas. Em: **Handbook of Clinical Neurology**. [s.l.] Elsevier, 2016. v. 134p. 71–95.

PICART, T. et al. Optimization of high-grade glioma resection using 5-ALA fluorescence-guided surgery: A literature review and practical recommendations from the neuro-oncology club of the French society of neurosurgery. **Neurochirurgie**, v. 65, n. 4, p. 164–177, 1 ago. 2019.

RAMINA, R. et al. 5-Aminolevulinic Acid-Protoporphyrin IX Fluorescence-Guided Surgery for CNS Tumors. First 41 cases in Latin America. **J Bras Neurocirurg**, v. 27, n. 1, p. 13–19, 2016.

REEVES, G. I.; MARKS, J. E. Prognostic Significance of Lesion Size for Glioblastoma Multiforme 1. **Radiology**, v. 132, p. 469–471, 1979.

RONG, Y. et al. Pseudopalisading Necrosis in Glioblastoma: A Familiar Morphologic Feature That Links Vascular Pathology, Hypoxia, and Angiogenesis. **J Neuropathol Exp Neurol**, v. 65, n. 6, 2006.

SANAI, N. et al. An extent of resection threshold for newly diagnosed glioblastomas: Clinical article. **Journal of Neurosurgery**, v. 115, n. 1, p. 3–8, jul. 2011.

SCHEITHAUER, B. W. Development of the WHO classification of tumors of the central nervous system: A historical perspective. **Brain Pathology**, v. 19, n. 4, p. 551–564, out. 2009.

SENDERS, J. T. et al. Agents for fluorescence-guided glioma surgery: a systematic review of preclinical and clinical results. **Acta Neurochirurgica** Springer-Verlag Wien, , 1 jan. 2017.

SHIMIZU, H. et al. Correlation between Choline Level Measured by Proton MR Spectroscopy and Ki-67 Labeling Index in Gliomas. **AJNR Am J Neuroradiol**, v. 21, p. 659–665, 2000.

SHIMIZU, K. et al. Correlation of Intraoperative 5-ALA-Induced Fluorescence Intensity and Preoperative ¹¹C-Methionine PET Uptake in Glioma Surgery. **Cancers**, v. 14, n. 6, 1 mar. 2022.

SHUKLA, G. et al. Advanced magnetic resonance imaging in glioblastoma: A review. **Chinese Clinical Oncology**, v. 6, n. 4, 1 ago. 2017.

SOTOUDEH, H. et al. Artificial intelligence in the management of glioma: Era of personalized medicine. **Frontiers in Oncology**, v. 9, 2019.

STOYANOV, G. S.; DZHENKOV, D. L. On the Concepts and History of Glioblastoma Multiforme - Morphology, Genetics and Epigenetics. **Folia medica**, v. 60, n. 1, p. 48–66, 1 mar. 2018.

STUMMER, W. et al. Technical Principles for Protoporphyrin-IX-Fluorescence Guided Microsurgical Resection of Malignant Glioma Tissue. **Acta Neurochir**, v. 140, p. 95–100, 1998a.

STUMMER, W. et al. Intraoperative Detection of Malignant Gliomas by 5-Aminolevulinic Acid-induced Porphyrin Fluorescence. **Neurosurgery**, v. 42, n. 3, p. 518–526, 1998b.

STUMMER, W. et al. Counterbalancing risks and gains from extended resections in malignant glioma surgery: A supplemental analysis from the randomized 5-aminolevulinic acid glioma resection study: Clinical article. **Journal of Neurosurgery**, v. 114, n. 3, p. 613–623, mar. 2011.

STUPP, R. et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. **New England Journal of Medicine**, v. 352, n. 10, p. 987–996, 2005.

THAKKAR, J. P. et al. Epidemiologic and molecular prognostic review of glioblastoma. **Cancer Epidemiology Biomarkers and Prevention**, v. 23, n. 10, p. 1985–1996, 1 out. 2014.

TONN, J. C.; WESTPHAL, M.; RUTKA, J. T. **Oncology of CNS tumors**. [s.l.] Springer Berlin Heidelberg, 2010.

TUSTISON, N. J. et al. N4ITK: Improved N3 bias correction. **IEEE Transactions on Medical Imaging**, v. 29, n. 6, p. 1310–1320, jun. 2010.

VALDÉ S, P. A. et al. δ -aminolevulinic acid-induced protoporphyrin IX concentration correlates with histopathologic markers of malignancy in human gliomas: The need for quantitative fluorescence-guided resection to identify regions of increasing malignancy. **Neuro-Oncology**, v. 13, n. 8, p. 846–856, ago. 2011.

VAN KEMPEN, E. J. et al. Performance of machine learning algorithms for glioma segmentation of brain MRI: a systematic literature review and meta-analysis. **European Radiology**, v. 31, p. 9638–9653, 2021.

WANGARYATTAWANICH, P. et al. Multicenter imaging outcomes study of The Cancer Genome Atlas glioblastoma patient cohort: Imaging predictors of overall and progression-free survival. **Neuro-Oncology**, v. 17, n. 11, p. 1525–1537, 1 nov. 2015.

WOOD, J. R.; GREEN, S. B.; SHAPIRO, W. R. The Prognostic Importance of Tumor Size in Malignant Gliomas: A Computed Tomographic Scan Study by the Brain Tumor Cooperative Group. **J Clin Oncol**, v. 6, p. 338–343, 1988a.

WOOD, M. D.; HALFPENNY, A. M.; MOORE, S. R. Applications of molecular neuro-oncology - A review of diffuse glioma integrated diagnosis and emerging molecular entities. **Diagnostic Pathology**, v. 14, n. 1, 9 abr. 2019.

XUE, D.; ALBRIGHT, R. E. Preoperative Anaplastic Glioma Tumor Volume Effects on Patient Survival. **J. Surg. Oncol**, v. 72, p. 199–205, 1999a.

YUAN, Y. et al. TERT mutation in glioma: Frequency, prognosis and risk. **Journal of Clinical Neuroscience**, v. 26, p. 57–62, 1 abr. 2016.

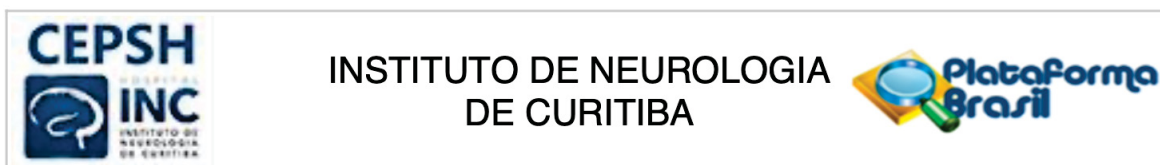
ZHANG, Y. et al. **The p53 pathway in glioblastoma**. **CancersMDPI AG**, , 1 set. 2018.

ZHONG, L. et al. (TS)2WM: Tumor Segmentation and Tract Statistics for Assessing White Matter Integrity with Applications to Glioblastoma Patients. **NeuroImage**, v. 223, 1 dez. 2020.

ZINN, P. O. et al. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. **PLoS ONE**, v. 6, n. 10, 5 out. 2011.

ZÜLCH, K.; KÖLN. Principles of the New World Health Organization (WHO) Classification of Brain Tumors. **Neuroradiology**, v. 19, p. 59–66, 1980.

ANEXO I – APROVAÇÃO DO COMITÊ DE ÉTICA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Inteligência Artificial e Radiogenômica dos Gliomas

Pesquisador: Flávia Sprenger

Área Temática:

Versão: 2

CAAE: 40280320.7.0000.5227

Instituição Proponente: Hospital ECOVILLE/ Instituto de Neurologia de Curitiba S/C Ltda

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.467.490

Apresentação do Projeto:

aprovado

Objetivo da Pesquisa:

aprovado

Avaliação dos Riscos e Benefícios:

aprovado

Comentários e Considerações sobre a Pesquisa:

sem comentários

Considerações sobre os Termos de apresentação obrigatória:

sem comentários

Conclusões ou Pendências e Lista de Inadequações:

projeto aprovado após alterações

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

CURITIBA, 16 de Dezembro de 2020

Assinado por:
Samanta Fabricio Blattes da Rocha
(Coordenador(a))



Federated learning enables big data for rare cancer boundary detection

Received: 7 April 2022

Accepted: 16 September 2022

Published online: 05 December 2022



A list of authors and their affiliations appears at the end of the paper

Although machine learning (ML) has shown promise across disciplines, out-of-sample generalizability is concerning. This is currently addressed by sharing multi-site data, but such centralization is challenging/infeasible to scale due to various limitations. Federated ML (FL) provides an alternative paradigm for accurate and generalizable ML, by only sharing numerical model updates. Here we present the largest FL study to-date, involving data from 71 sites across 6 continents, to generate an automatic tumor boundary detector for the rare disease of glioblastoma, reporting the largest such dataset in the literature ($n = 6,314$). We demonstrate a 33% delineation improvement for the surgically targetable tumor, and 23% for the complete tumor extent, over a publicly trained model. We anticipate our study to: 1) enable more healthcare studies informed by large diverse data, ensuring meaningful results for rare diseases and underrepresented populations, 2) facilitate further analyses for glioblastoma by releasing our consensus model, and 3) demonstrate the FL effectiveness at such scale and task-complexity as a paradigm shift for multi-site collaborations, alleviating the need for data-sharing.

Recent technological advancements in healthcare, coupled with patients' culture shifting from reactive to proactive, have resulted in a radical growth of primary observations generated by health systems. This contributes to the burnout of clinical experts, as such observations require thorough assessment. To alleviate this situation, there have been numerous efforts for the development, evaluation, and eventual clinical translation of machine learning (ML) methods to identify relevant relationships among these observations, thereby reducing the burden on clinical experts. Advances in ML, and particularly deep learning (DL), have shown promise in addressing these complex healthcare problems. However, there are concerns about their generalizability on data from sources that did not participate in model training, i.e., "out-of-sample" data^{1,2}. Literature indicates that training robust and accurate models requires large amounts of data^{3–5}, the diversity of which affects model generalizability to "out-of-sample" cases⁶. To address these concerns, models need to be trained on data originating from numerous sites representing diverse population samples. The current paradigm for such multi-site collaborations is "centralized learning" (CL), in which data from different sites are shared to a centralized location following inter-site agreements^{6–9}.

However, such data centralization is difficult to scale (and might not even be feasible), especially at a global scale, due to concerns^{10,11} relating to privacy, data ownership, intellectual property, technical challenges (e.g., network and storage limitations), as well as compliance with varying regulatory policies (e.g., Health Insurance Portability and Accountability Act (HIPAA) of the United States¹² and the General Data Protection Regulation (GDPR) of the European Union¹³). In contrast to this centralized paradigm, "federated learning" (FL) describes a paradigm where models are trained by only sharing model parameter updates from decentralized data (i.e., each site retains its data locally)^{10,11,14–16}, without sacrificing performance when compared to CL-trained models^{11,15,17–21}. Thus, FL can offer an alternative to CL, potentially creating a paradigm shift that alleviates the need for data sharing, and hence increase access to geographically distinct collaborators, thereby increasing the size and diversity of data used to train ML models.

FL has tremendous potential in healthcare^{22,23}, particularly towards addressing health disparities, under-served populations, and "rare" diseases²⁴, by enabling ML models to gain knowledge from ample and diverse data that would otherwise not be available. With

✉ e-mail: sbakas@upenn.edu

that in mind, here we focus on the “rare” disease of glioblastoma, and particularly on the detection of its extent using multi-parametric magnetic resonance imaging (mpMRI) scans²⁵. While glioblastoma is the most common malignant primary brain tumor^{26–28}, it is still classified as a “rare” disease, as its incidence rate (i.e., 3/100,000 people) is substantially lower than the rare disease definition rate (i.e., <10/100,000 people)²⁴. This means that single sites cannot collect large and diverse datasets to train robust and generalizable ML models, necessitating collaboration between geographically distinct sites. Despite extensive efforts to improve the prognosis of glioblastoma patients with intense multimodal therapy, their median overall survival is only 14.6 months after standard-of-care treatment, and 4 months without treatment²⁹. Although the subtyping of glioblastoma has been improved³⁰ and the standard-of-care treatment options have expanded during the last 20 years, there have been no substantial improvements in overall survival³¹. This reflects the major obstacle in treating these tumors which is their intrinsic heterogeneity^{26,28}, and the need for analyses of larger and more diverse data toward a better understanding of the disease. In terms of radiologic appearance, glioblastomas comprise of three main sub-compartments, defined as (i) the “enhancing tumor” (ET), representing the vascular blood-brain barrier breakdown within the tumor, (ii) the “tumor core” (TC), which includes the ET and the necrotic (NCR) part, and represents the surgically relevant part of the tumor, and (iii) the “whole tumor” (WT), which is defined by the union of the TC and the peritumoral edematous/infiltrated tissue (ED) and represents the complete tumor extent relevant to radiotherapy (Fig. 1b). Detecting these sub-compartment boundaries, therefore, defines a multi-parametric multi-class learning problem and is a critical first step towards further quantifying and assessing this heterogeneous rare disease and ultimately influencing clinical decision-making.

Co-authors in this study have previously introduced FL in healthcare in a simulated setting¹⁵ and further conducted a thorough quantitative performance evaluation of different FL workflows¹¹ (refer to supplementary figures for illustration) for the same use-case as the present study, i.e., detecting the boundaries of glioblastoma sub-compartments. Findings from these studies supported the superiority of the FL workflow used in the present study (i.e., based on an aggregation server^{10,14}), which had almost identical performance to CL, for this use-case. Another study³² has explored the first real-world federation for a breast cancer classification task using 5 sites, and another¹⁶ used electronic medical records along with x-ray images from 20 sites to train a classifier to output a label corresponding to future oxygen requirement for COVID-19 patients.

This study describes the largest to-date global FL effort to develop an accurate and generalizable ML model for detecting glioblastoma sub-compartment boundaries, based on data from 6314 glioblastoma patients from 71 geographically distinct sites, across six continents (Fig. 1a). Notably, this describes the largest and most diverse dataset of glioblastoma patients ever considered in the literature. It was the use of FL that successfully enabled our ML model to gain knowledge from such an unprecedented dataset. The extended global footprint and the task complexity are what sets this study apart from current literature, since it dealt with a multi-parametric multi-class problem with reference standards that require expert clinicians following an involved manual annotation protocol, rather than simply recording a categorical entry from medical records^{16,32}. Moreover, varying characteristics of the mpMRI data due to scanner hardware and acquisition protocol differences^{33,34} were handled at each collaborating site via established harmonized preprocessing pipelines^{35–39}.

The scientific contributions of this manuscript can be summarized by (i) the insights garnered during this work that can pave the way for more successful FL studies of increased scale and task complexity, (ii) making a potential impact for the treatment of the rare disease of glioblastoma by publicly releasing clinically deployable trained

consensus models, and most importantly, iii) demonstrating the effectiveness of FL at such scale and task complexity as a paradigm shift redefining multi-site collaborations, while alleviating the need for data sharing.

Results

The complete federation followed a staged approach, starting from a “public initial model” (trained on data of 231 cases from 16 sites), followed by a “preliminary consensus model” (involving data of 2471 cases from 35 sites), to conclude on the “final consensus model” (developed on data of 6314 cases from 71 sites). To quantitatively evaluate the performance of the trained models, 20% of the total cases contributed by each participating site were excluded from the model training process and used as “local validation data”. To further evaluate the generalizability of the models in unseen data, 6 sites were not involved in any of the training stages to represent an unseen “out-of-sample” data population of 590 cases. To facilitate further evaluation without burdening the collaborating sites, a subset ($n = 332$) of these cases was aggregated to serve as a “centralized out-of-sample” dataset. The training was initiated from a pre-trained model (i.e., our public initial model) rather than a random initialization point, in order to have faster convergence of the model performance^{40,41}. Model performance was quantitatively evaluated here using the Dice similarity coefficient (DSC), which assesses the spatial agreement between the model’s prediction and the reference standard for each of the three tumor sub-compartments (ET, TC, WT).

Increased data can improve performance

When the federation began, the public initial model was evaluated against the local validation data of all sites, resulting in an average (across all cases of all sites) DSC per sub-compartment, of $DSC_{ET} = 0.63$, $DSC_{TC} = 0.62$, $DSC_{WT} = 0.75$. To summarize the model performance with a single collective score, we then calculate the average DSC (across all 3 tumor sub-compartments per case, and then across all cases of all sites) as equal to 0.66. Following model training across all sites, the final consensus model garnered significant performance improvements against the collaborators’ local validation data of 27% ($p_{ET} < 1 \times 10^{-36}$), 33% ($p_{TC} < 1 \times 10^{-59}$), and 16% ($p_{WT} < 1 \times 10^{-21}$), for ET, TC, and WT, respectively (Fig. 1c). To further evaluate the potential generalizability improvements of the final consensus model on unseen data, we compared it with the public initial model against the complete out-of-sample data and noted significant performance improvements of 15% ($p_{ET} < 1 \times 10^{-5}$), 27% ($p_{TC} < 1 \times 10^{-16}$), and 16% ($p_{WT} < 1 \times 10^{-7}$), for ET, TC, and WT, respectively (Fig. 1d). Notably, the only difference between the public initial model and the final consensus model, was that the latter gained knowledge during training from increased datasets contributed by the complete set of collaborators. The conclusion of this finding reinforces the importance of using large and diverse data for generalizable models to ultimately drive patient care.

Data size alone may not predict success

This is initially observed in our federated setting, where the comparative evaluation of the public initial model, the preliminary consensus model, and the final consensus model, against the centralized out-of-sample data, indicated performance improvements not directly related to the amount of data used for training. Specifically, we noted major significant ($p < 7 \times 10^{-18}$, Wilcoxon signed-rank test) performance improvements between the public initial model and the preliminary consensus model, as opposed to the insignificant ($p > 0.067$, Wilcoxon signed-rank test) ones between the preliminary and the final consensus model, as quantified in the centralized out-of-sample data for all sub-compartments and their average (Fig. 2).

We further expanded this analysis to assess this observation in a non-federated configuration, where we selected the largest collaborating sites (comprehensive cancer centers contributing > 200 cases,

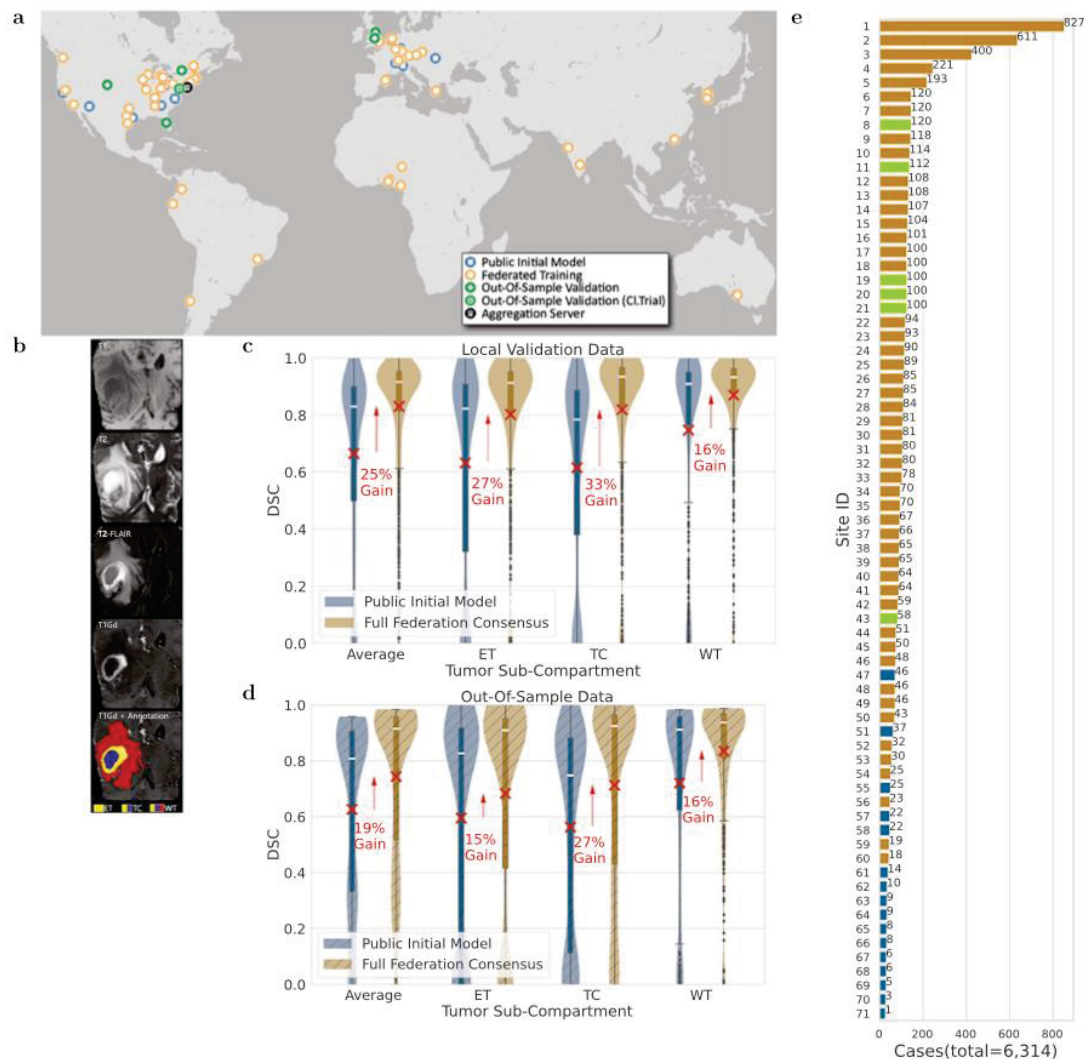


Fig. 1 | Representation of the study's global scale, diversity, and complexity.

a The map of all sites involved in the development of FL consensus model. **b** Example of a glioblastoma mpMRI scan with corresponding reference annotations of the tumor sub-compartments (ET enhancing tumor, TC tumor core, WT whole tumor). **c, d** Comparative Dice similarity coefficient (DSC) performance evaluation of the final consensus model with the public initial model on the collaborators' local validation data (in **c** with $n=1043$ biologically independent cases) and on the complete out-of-sample data (in **d** with $n=518$ biologically independent cases), per tumor sub-compartment (ET enhancing tumor, TC tumor core, WT whole tumor). Note the box and whiskers inside each violin plot represent the true

min and max values. The top and bottom of each "box" depict the 3rd and 1st quartile of each measure. The white line and the red 'x', within each box, indicate the median and mean values, respectively. The fact that these are not necessarily at the center of each box indicates the skewness of the distribution over different cases. The "whiskers" drawn above and below each box depict the extremal observations still within 1.5 times the interquartile range, above the 3rd or below the 1st quartile. Equivalent plots for the Jaccard similarity coefficient (JSC) can be observed in supplementary figures. **e** Number of contributed cases per collaborating site.

and familiar with computational analyses), and coordinated independent model training for each, starting from the public initial model and using only their local training data. The findings of this evaluation indicate that the final consensus model performance is always superior or insignificantly different ($p_{\text{Average}} = 0.1$, $p_{\text{ET}} = 0.5$, $p_{\text{TC}} = 0.2$, $p_{\text{WT}} = 0.06$, Wilcoxon signed-rank test) to the ensemble of the local models of these four largest contributing collaborators, for all tumor sub-compartments (Fig. 2). This finding highlights that even large sites can benefit from collaboration.

FL is robust to data quality issues

Data quality issues relating to erroneous reference annotations (with potential negative downstream effects on output predictions) were identified by monitoring the global consensus model performance during training. However, only data quality issues that largely affected the global validation score could be identified and corrected during training. Those with more subtle effects in the global validation score were only identified after the completion of the model training by looking for relatively low local validation scores of the consensus

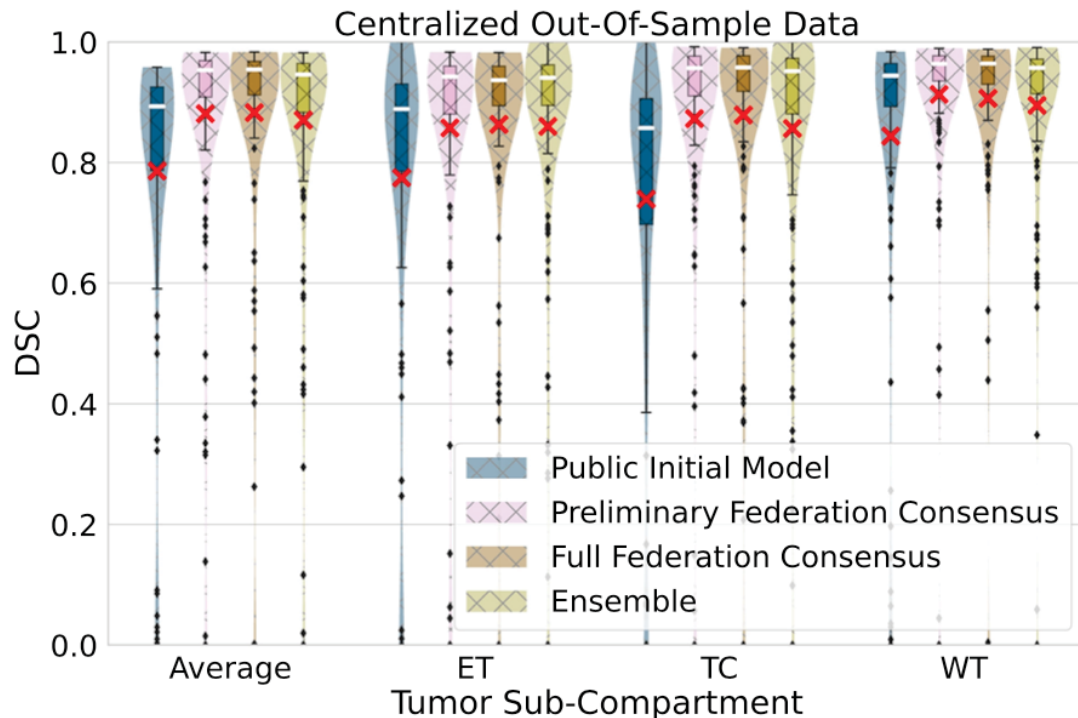


Fig. 2 | Generalizable Dice similarity coefficient (DSC) evaluation on ‘centralized’ out-of-sample data ($n = 154$ biologically independent cases), per tumor sub-compartment (ET enhancing tumor, TC tumor core, WT whole tumor) and averaged across cases. Comparative performance evaluation across the public initial model, the preliminary consensus model, the final consensus model, and an ensemble of single site models from collaborators holding > 200 cases. Note the box and whiskers inside each violin plot, represent the true min and max values. The top and bottom of each “box” depict the 3rd and 1st quartile of

each measure. The white line and the red ‘x’, within each box, indicate the median and mean values, respectively. The fact that these are not necessarily at the center of each box indicates the skewness of the distribution over different cases. The “whiskers” drawn above and below each box depict the extremal observations still within 1.5 times the interquartile range, above the 3rd or below the 1st quartile. Equivalent plots for Jaccard similarity coefficient (JSC) can be observed in supplementary figures.

model across collaborating sites. An example of such a quality issue with erroneous reference labels (from Site 48) is shown in Fig. 3c. Looking closer, local validation scores at Site 48 (Fig. 3b) are significantly different ($p_{ET} < 3 \times 10^{-12}$, $p_{TC} < 3 \times 10^{-12}$, $p_{WT} < 3 \times 10^{-12}$, Wilcoxon signed-rank test) than the average scores across the federation (Fig. 3a). Significant differences were calculated by sample pairs for each federated round, where a sample pair consists of the mean validation score over samples for Site 48 paired with those across all sites. These local validation scores (Fig. 3b) indicate that the model is not gaining knowledge from these local data, and their comparison with the average scores across the federation (Fig. 3a) indicates that the global consensus model performance is not adversely affected. This finding supports the importance of robustness at a global scale.

FL benefits the more challenging tasks

The complexity of boundary detection drops when moving from smaller to larger sub-compartments, i.e., from ET to TC, and then to WT^{35–38}. This is further confirmed here, as evidenced by the model’s relative performance indicated by the local validation curves and their underlying associated areas in Fig. 3a. Since the current clinically actionable sub-compartments are TC (i.e., considered for surgery) and WT (i.e., considered for radiotherapy)⁴², performance improvements of their boundary detection may contribute to the model’s clinical impact and relevance.

Our findings indicate that the benefits of FL are more pronounced for the more challenging sub-compartments, i.e., larger performance improvements for ET and TC compared to WT (Fig. 1c). Notably, the largest and most significant improvement (33%, $p < 7 \times 10^{-60}$) is noted for the TC sub-compartment, which is surgically actionable and not a trivial sub-compartment to delineate accurately^{43,44}. This finding of FL benefiting the more challenging tasks rather than boosting performance on the relatively easier task (e.g., thresholding the abnormal T2-FLAIR signal for the WT sub-compartment) by gaining access to larger amounts of good quality data holds a lot of promise for FL in healthcare.

Optimal model selection is non-trivial

Using the performance of the global consensus model during training across all local validation cases, two distinct model configurations were explored for selecting the final consensus model. Analyzing the sequence of consensus models produced during each federated round, we selected four different models: the *singlet*, for which the average DSC across all sub-compartments scored high, and three independent models, each of which yielded high DSC scores for each tumor sub-compartment, i.e., ET, TC, WT. We defined the collection of these three independent consensus models as a *triplet*.

To identify the best model, 5 *singlets* and 5 *triplets* were selected based on their relative performance on all local validation cases and

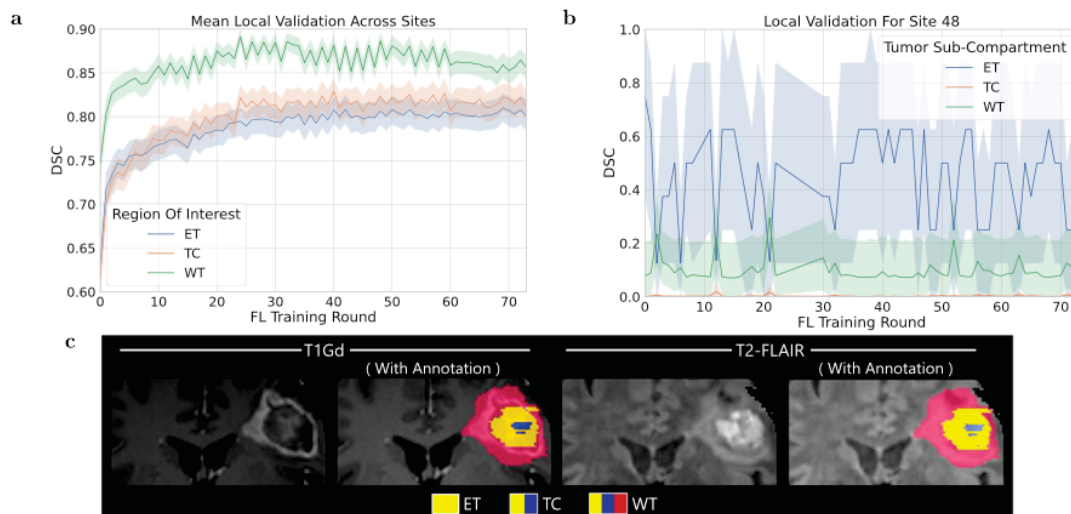


Fig. 3 | Per-tumor region (ET enhancing tumor, TC tumor core, WT whole tumor) mean Dice similarity coefficient (DSC) over validation samples (with shading indicating 95% confidence intervals again over samples). a At all participating sites across training rounds showing that the score is greater for sub-compartments with larger volumes. **b** For a site with problematic annotations (Site

48). The instability in these curves could be caused by errors in annotation for the local validation data (similar to errors that were observed for a small shared sample of data from this site). **c** Provides an example of a case with erroneous annotations in the data used by Site 48. Equivalent plots for Jaccard similarity coefficient (JSC) can be observed in supplementary figures.

evaluated against the centralized out-of-sample data. Only small differences are observed between the *singlet* and *triplet* models, and these differences diminish as the sub-compartment size increases. Comparing the means of *singlet* and *triplet*, the larger (and only significant) performance improvement difference compared to the public initial model is noted for the ET sub-compartment (improved by < 3%, $p_{ET} = 0.02$), followed by TC (improved by < 1.4%, $p_{TC} = 0.09$), and then lastly WT (improved by < 1.1%, $p_{WT} = 0.2$) (Tables S1 and S2). However, the decision of using a *singlet* or a *triplet* model should also rely on computational cost considerations, as *triplets* will be three times more expensive than *singlets* during model inference.

Discussion

In this study, we have described the largest real-world FL effort to-date utilizing data of 6314 glioblastoma patients from 71 geographically unique sites spread across 6 continents, to develop an accurate and generalizable ML model for detecting glioblastoma sub-compartment boundaries. Notably, this extensive global footprint of the collaborating sites in this study also yields the largest dataset ever reported in the literature assessing this rare disease. It is the use of FL that successfully enabled (i) access to such an unprecedented dataset of the most common and fatal adult brain tumor, and (ii) meaningful ML training to ensure the generalizability of models across out-of-sample data. In comparison with the limited existing real-world FL studies^{16,32}, our use-case is larger in scale and substantially more complex, since it (1) addresses a multi-parametric multi-class problem, with reference standards that require expert collaborating clinicians to follow an involved manual annotation protocol, rather than simply recording a categorical entry from medical records, and (2) requires the data to be preprocessed in a harmonized manner to account for differences in MRI acquisition. Since glioblastoma boundary detection is critical for treatment planning and the requisite first step for further quantitative analyses, the models generated during this study have the potential to make a far-reaching clinical impact.

The large and diverse data that FL enabled, led to the final consensus model garnering significant performance improvements over

the public initial model against both the collaborators' local validation data and the complete out-of-sample data. The improved result is a clear indication of the benefit that can be afforded through access to more data. However, increasing the data size for model training without considerations relating to data quality, reference labels, and potential site bias (e.g., scanner acquisition protocols, demographics, or sociocultural considerations, such as more advanced presentation of disease at diagnosis in low-income regions⁴³) might not always improve results. Literature also indicates an ML performance stagnation effect, where each added case contributes less to the model performance as the number of cases increase⁴⁶. This is in line with our finding in the federated setting (Fig. 2), where performance improvements across the public initial model, the preliminary consensus model, and the final consensus model, were not directly/linearly related to the amount of data used for training. This happened even though the final consensus model was trained on over twice the number of cases (and included 2 of the largest contributing sites—Sites 1 and 4) when compared to the preliminary consensus model. Further noting that the preliminary federation model was already within the intra- and inter-rater variability range for this use-case (20% and 28%, respectively)⁴⁷, any further improvements for the full federation consensus model would be expected to be minimal^{35–38}.

To further assess these considerations, we coordinated independent model training for the four largest collaborating sites (i.e., >200 cases) by starting from the same public initial model and using only their local training data. The ensemble of these four largest site local models did not show significant performance differences to the final consensus model for any tumor sub-compartment, yet the final consensus model showed superior performance indicating that even sites with large datasets can benefit from collaboration. The underlying assumption for these results is that since each of these collaborators initiated their training from the public initial model (which included diverse data from 16 sites), their independent models and their ensemble could have inherited some of the initial model's data diversity, which could justify the observed insignificant differences (Fig. 2 and Supplementary Fig. 3). Though these findings are an indication

that the inclusion of more data alone may not lead to better performance, it is worth noting that these four largest sites used for the independent model training represent comprehensive cancer centers (compared to hospitals in community settings) with affiliated sophisticated labs focusing on brain tumor research, and hence were familiar with the intricacies of computational analyses. Further considering the aforementioned ML performance stagnation effect, we note the need for generalizable solutions to quantify the contribution of collaborating sites to the final consensus model performance, such that future FL studies are able to formally assess both the quantity and the quality of the contributed data needed by the collaborating sites and decide on their potential inclusion on use-inspired studies.

As noted in our results, due to the lack of such generalizable solutions, we were only able to identify quality issues after the model training. Specifically, we hypothesize that although Site 48 had data quality issues, its effect on the consensus model performance was not significant due to its relatively small dataset ($n = 46$) when compared to the other collaborating sites. The curves of Fig. 3a indicate that the global consensus model continues to consistently gain knowledge from the federation as a whole during training, highlighting robustness to such data quality issues. It remains unknown, however, how much better the consensus model would have performed if sites with problematic data were excluded or if these specific problematic data at Site 48 were excluded or corrected. These findings are aligned with literature observations (on the same use-case)⁴⁸, where a DL model⁴⁹ trained on 641 glioblastoma cases from 8 sites produced higher quality predictions on average than those created as reference standard labels by radiology expert operators. Quality was judged by 20 board-certified neuroradiologists, in a blinded side-by-side comparison of 100 sequestered unseen cases, and concluded that perfect or near-perfect reference labels may not be required to produce high-quality prediction systems. In other words, DL models may learn to see past imperfect reference training labels. These findings provide the impetus for further experimentation as they have implications for future FL studies. Future research is needed to automatically detect anomalies in the consensus model performance during training, particularly associated with contributions from individual sites.

There are a number of practical considerations that need to be taken into account to set up a multi-national real-world federation, starting with a substantial amount of coordination between each participating site. As this study is the first at this scale and task complexity, we have compiled a set of governance insights from our experience that can serve as considerations for future successful FL studies. These insights differ from previous literature that describes studies that were smaller in scale and involved simpler tasks^{16,32}. By “governance” of the federation we refer both to the accurate definition of the problem statement (including reference labels and harmonization considerations accounting for inter-site variability), and the coordination with the collaborating sites for eligibility and compliance with the problem statement definition, as well as security and technical considerations. For future efforts aiming to conduct studies of a similar global scale, it would be beneficial to identify a solution for governance prior to initiating the study itself.

The coordination began with engaging the security teams of collaborating sites and providing them access to the source code of the platform developed to facilitate this study. These security discussions highlighted the benefit of the platform being open-source, making security code reviews easier. Resource gathering was then carried out by identifying technical leads and assessing computational resources at each site. With the technical leads, we then proceeded to test the complete workflow to further identify gaps in the requirements, such as network configurations and hardware requirements. We then proceeded with data curation and preprocessing, and finally connected individual sites to the aggregation server to initiate their participation.

Following the precise definition of our problem statement^{35–38}, ensuring strict compliance with the preprocessing and annotation protocol for the generation of reference standards was vital for the model to learn correct information during training. To this end, we instituted an extensively and comprehensively documented annotation protocol with visual example representations and common expected errors (as observed in the literature^{38,50}) to all collaborators. We have further circulated an end-to-end platform³⁹ developed to facilitate this federation, providing to each collaborating site all the necessary functionalities to (i) uniformly curate their data and account for inter-site acquisition variability, (ii) generate the reference standard labels, and (iii) participate in the federated training process. Finally, we held interactive sessions to complement the theoretical definition of the reference standards, and further guide collaborating sites. Particular pain points regarding these administrative tasks included managing the large volume of communication (i.e., emails and conference calls) needed to address questions and issues that arose, as well as the downtime incurred in FL training due to issues that had not yet been identified and were adversely affecting the global model. Though we developed many ad-hoc tools for this workflow ourselves (particularly for the data processing and orchestration steps), many issues we encountered were common enough in retrospect (for example common Transport Layer Security (TLS) errors) that mature automated solutions will address them. Many of these automations will be use-case dependent, such as the MRI data corruption checks we used from the FeTS tool³⁹. For these use-case-dependent automation, more associated tools are expected to become available as various domain experts enter into the FL community, while some will be more general purpose. As our inspection of both local and global model validation scores was manual during our deployment, we in retrospect see great value in automated notifications (performed at the collaborator infrastructure to help minimize data information leakage) to alert a collaborator (or the governor) when their local or global model validation is significantly low. Such an alert can indicate the potential need to visually inspect example failure cases in their data for potential issues. With continued efforts towards developing automated administration tools around FL deployments, we expect the coordination for large FL deployments to become easier.

In general, debugging issues with the inputted local data and annotations is more difficult during FL due to the level of coordination and/or privacy issues involved, since the data are always retained at the collaborating site. We gained substantial experience during this effort that went into further development of use-inspired but generalizable data sanity-checking functionality in the tools we developed, towards facilitating further multi-site collaborations.

Upon conclusion of the study, sites participating in the model training process were given a survey to fill in regarding various aspects of their experience. According to the provided feedback, 96% of the sites found the comprehensive documentation on preprocessing and data curation essential and thought that lack of such documentation could have resulted in inconsistent annotations. Additionally, 92% found the documentation relating to establishing secure connectivity to the aggregation server easy to follow and essential to expedite reviews by the related groups. Furthermore, 84% of the sites appreciated the user-friendly interface of the provided tool and its associated complete functionality (beyond its FL backend), and indicated their intention to use it and recommend it for projects and data analysis pipelines beyond the scope of this study. To generate the reference standard labels for their local data, 86% of the collaborating sites indicated that they used either the FeTS Tool³⁹ (i.e., the tool developed for this study), CaPTk⁵¹, or ITK-SNAP⁵², whereas the remaining 14% used either 3D-Slicer⁵³, the BraTS toolkit⁵⁴, or something else. In terms of hardware requirements at each site, 88% used a dedicated workstation for their local workload, and the remaining 12% used either a containerized form of the FeTS tool or a virtual machine.

Although data are always retained within the acquiring site during FL (and hence FL is defined as private-by-design), different security and privacy threats remain^{55–57}. These threats include attempted extraction of training data information from intermediate and final models, model theft, and submission of poison model updates with the goal of introducing unwanted model behavior (including incentivizing the model to memorize more information about the training data in support of subsequent extraction, i.e., leakage). A number of technologies can be used to mitigate security and privacy concerns during FL^{55–57}. Homomorphic encryption⁵⁸, secure multiparty compute⁵⁹, and trusted execution environments (TEEs)^{60,61} allow for collaborative computations to be performed with untrusted parties while maintaining confidentiality of the inputs to the computation. Differentially private training algorithms^{62–64} allow for mitigation of information leakage from both the collaborator model updates and the global consensus aggregated models. Finally, assurance that remote computations are executed with integrity can be designed for with the use of hardware-based trust provided by TEEs, as well as with some software-based integrity checking⁶⁵. Each of these technologies comes with its own benefits in terms of security and/or privacy, as well as costs and limitations, such as increased computational complexity, associated hardware requirements and/or reduced quality of computational output (such as the reduction of model utility that can be associated with differentially private model training). Further experimentation needs to be done in order to best inform prospective federations as to which technologies to use towards addressing their specific concerns within the context of the collaborator infrastructure and trust levels, depending on the use-case, the extent of the collaborating network, and the level of trust within the involved parties. Our study was based on a collaborative network of trusted sites, where authentication was based on personal communication across collaborating sites and the combination of TLS and TEEs were considered sufficient.

Although our study has the potential to become the baseline upon which future ML research studies will be done, there is no automated mechanism to assess inputted data quality from collaborators, which could result in models trained using sub-optimal data. Additionally, we used a single off-the-shelf neural network architecture for training, but it has been shown that model ensembles perform better for the task at hand^{35–38}, and it remains to be explored how such a strategy could be explored in a federated study. Moreover, the instantiation of the federation involved a significant amount of coordination between each site and considering the limited real-world FL studies at the time, there were no tools available to automate such coordination and orchestration. These involved (i) getting interviewed by information security officers of collaborating sites, (ii) ensuring that the harmonized pre-processing pipeline was used effectively, (iii) clear communication of the annotation protocol, and (iv) testing the network communication between the aggregator and each site. This amount of effort, if not aided by automated tools, will continue to be a huge roadblock for FL studies, and dedicated coordination and orchestration resources are required to conduct this in a reproducible and scalable manner.

We have demonstrated the utility of an FL workflow to develop an accurate and generalizable ML model for detecting glioblastoma sub-compartment boundaries, a finding which is of particular relevance for neurosurgical and radiotherapy planning in patients with this disease. This study is meant to be used as an example for future FL studies between collaborators with an inherent amount of trust that can result in clinically deployable ML models. Further research is required to assess privacy concerns in a detailed manner^{63,64} and to apply FL to different tasks and data types^{66–69}. Building on this study, a continuous FL consortium would enable downstream quantitative analyses with implications for both routine practice and clinical trials, and most importantly, increase access to high-quality precision care worldwide. Furthermore, the lessons learned from this study with such a global footprint are invaluable and can be applied to a broad array of clinical

scenarios with the potential for great impact on rare diseases and underrepresented populations.

Methods

The study and results presented in this manuscript comply with all relevant ethical regulations and follow appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding the treatment of human subjects. Use of the private retrospective data collection of each collaborating site has been approved by their respective institutional review board, where informed consent from all participants was also obtained and stored.

Data

The data considered in this study described patient populations with adult-type diffuse glioma³⁰, and specifically displaying the radiological features of glioblastoma, scanned with mpMRI to characterize the anatomical tissue structure²⁵. Each case is specifically described by (i) native T1-weighted (T1), (ii) Gadolinium-enhanced T1-weighted (T1Gd), (iii) T2-weighted (T2), and (iv) T2-weighted-Fluid-Attenuated-Inversion-Recovery (T2-FLAIR) MRI scans. Cases with any of these sequences missing were not included in the study. Note that no inclusion/exclusion criterion applied relating to the type of acquisition (i.e., both 2D axial and 3D acquisitions were included, with a preference for 3D if available), or the exact type of sequence (e.g., MP-RAGE vs. SPGR). The only exclusion criterion was for T1-FLAIR scans that were intentionally excluded to avoid mixing varying tissue appearance due to the type of sequence, across native T1-weighted scans.

The publicly available data from the International Brain Tumor Segmentation (BraTS) 2020 challenge^{35–37}, was used to train the public initial model of this study. The BraTS challenge^{35–38}, seeking methodological advancements in the domain of neuro-oncology, has been providing the community with (i) the largest publicly available and manually-curated mpMRI dataset of diffuse glioma patients (an example of which is illustrated in Fig. 1b), and (ii) a harmonized pre-processing pipeline^{51,70,71} to handle differences in inter-site acquisition protocols. The public initial model was used to initialize the FL training, instead of a randomly generated initialization, as starting from a pre-trained model leads to faster convergence⁴¹. The complete BraTS 2020 dataset originally included cases from sites that also participated in this study as independent collaborators. To avoid any potential data leakage, we reduced the size of the complete BraTS dataset by removing cases acquired by these specific sites, resulting in a dataset of 231 cases from 16 international sites, with varying contributing cases across sites (Fig. 1e). The exact site IDs that construct the data of the public initial model are: 47, 51, 55, 57, 58, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, and 71. Subsequently, the resulting dataset was split at a 4:1 ratio between cases for training ($n = 185$) and validation ($n = 46$).

The eligibility of collaborating sites to participate in the federation was determined based on data availability, and approval by their respective institutional review board. 55 sites participated as independent collaborators in the study defining a dataset of 6083 cases. The MRI scanners used for data acquisition were from multiple vendors (i.e., Siemens, GE, Philips, Hitachi, Toshiba), with magnetic field strength ranging from 1T to 3T. The data from all 55 collaborating sites followed a male:female ratio of 1.47:1 with ages ranging between 7 and 94 years.

From all 55 collaborating sites, 49 were chosen to be part of the training phase, and 6 sites were categorized as “out-of-sample”, i.e., none of these were part of the training stage. These specific 6 out-of-sample sites (Site IDs: 8, 11, 19, 20, 21, 43) were allocated based on their availability, i.e., they have indicated expected delayed participation rendering them optimal for model generalizability validation. One of these 6 out-of-sample sites (Site 11) contributed aggregated a priori data from a multi-site randomized clinical trial for newly diagnosed

glioblastoma (ClinicalTrials.gov Identifier: NCT00884741, RTOG0825^{72,73}, ACIN6686^{74,75}), with inherent diversity benefiting the intended generalizability validation purpose. The American College of Radiology (ACR - Site 11) serves as the custodian of this trial's imaging data on behalf of ECOG-ACRIN, which made the data available for this study. Following screening for the availability of the four required mpMRI scans with sufficient signal-to-noise ratio judged by visual observation, a subset of 362 cases from the original trial data were included in this study. The out-of-sample data totaled 590 cases intentionally held out of the federation, with the intention of validating the consensus model in completely unseen cases. To facilitate further such generalizability evaluation without burdening the collaborating sites, a subset consisting of 332 cases (including the multi-site clinical data provided by ACR) from this out-of-sample data was aggregated, to serve as the "centralized out-of-sample" dataset. Furthermore, the 49 sites participating in the training phase define a collective dataset of 5493 cases. The exact 49 site IDs are: 1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 13, 14, 15, 16, 17, 18, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 48, 49, 50, 52, 53, 54, 56, 59, 60. These cases were automatically split at each site following a 4:1 ratio between cases for training and local validation. During the federated training phase, the data used for the public initial model were also included as a dataset from a separate node, such that the contribution of sites providing the publicly available data is not forgotten within the global consensus model. This results in the final consensus model being developed based on data from 71 sites over a total dataset of 6314 cases. Collective demographic information of the included population is provided in Table S3.

Harmonized data preprocessing

Once each collaborating site identified its local data, they were asked to use the preprocessing functionality of the software platform we provided. This functionality follows the harmonized data preprocessing protocol defined by the BraTS challenge^{35–38}, as described below. This would allow accounting for inter-site acquisition protocol variations, e.g., 3D vs. 2D axial plane acquisitions.

File-type conversion/patient de-identification. The respective mpMRI scans (i.e., T1, T1Gd, T2, T2-FLAIR) of every case are downloaded onto a local machine in the Digital Imaging and Communications in Medicine (DICOM) format^{6–78} and converted to the Neuroimaging Informatics Technology Initiative (NIFTI) file format⁷⁹ to ensure easier parsing of the volumetric scans during the computational process. The conversion of DICOM to NIFTI files has the benefit of eliminating all patient-identifiable metadata from the header portion of the DICOM format^{80,81}.

Rigid registration. Once the scans are converted to the NIFTI format, each volume is registered to a common anatomical space, namely the SRI24 atlas⁸², to ensure a cohesive data shape ([240, 240, 155]) and an isotropic voxel resolution (1 mm³), thereby facilitating in the tandem analysis of the mpMRI scans. One of the most common types of MRI noise is based on the inhomogeneity of the magnetic field⁸³. It has been previously³⁶ shown that the use of non-parametric, non-uniform intensity normalization to correct for these bias fields^{84,85} obliterates the MRI signal relating to the regions of abnormal T2-FLAIR signal. Here, we have taken advantage of this adverse effect and used the bias field-corrected scans to generate a more optimal rigid registration solution across the mpMRI sequences. The bias field-corrected images are registered to the T1Gd image, and the T1Gd image is rigidly registered to the SRI24 atlas, resulting in two sets of transformation matrices per MRI sequence. These matrices are then aggregated into a single matrix defining the transformation of each MRI sequence from its original space to the atlas. We then apply this single aggregated matrix to

the NIFTI scans prior to the application of the bias field correction to maximize the fidelity of the finally registered images.

Brain extraction. This process focuses on generating a brain mask to remove all non-brain tissue from the image (including neck, fat, eye-balls, and skull), to enable further computational analyses while avoiding any potential face reconstruction/recognition⁸⁶. For this step we utilized the Brain Mask Generator (BrainMaGe)⁸⁷, which has been explicitly developed to address brain scans in presence of diffuse glioma and considers brain shape as a prior, hence being agnostic to the sequence/modality input.

Generation of automated baseline delineations of tumor sub-compartment boundaries. We provided the ability to the collaborating sites to generate automated delineations of the tumor sub-compartments from three popular methods from the BraTS challenge, using models trained using the challenge's training data: (i) DeepMedic⁴⁹, (ii) DeepScan⁸⁸, and (iii) nnU-Net⁸⁹. Along with segmentations from each method, label fusion strategies were also employed to provide a reasonable approximation to the reference labels that should be manually refined and approved by expert neuroradiologists to create the final reference labels. The label fusion approaches considered were i) standard voting⁹⁰, (ii) Simultaneous Truth And Performance Level Estimation (STAPLE)^{91,92}, iii) majority voting⁹³, and iv) Selective and Iterative Method for Performance Level Estimation (SIMPLE)⁹⁴.

Manual refinements towards reference standard labels. It was communicated to all participating sites to leverage the annotations generated using the automated mechanism as a baseline on which manual refinements were needed by neuroradiology experts, following a consistently communicated annotation protocol. The reference annotations comprised the Gd-enhancing tumor (ET—label '4'), the peritumoral edematous/invaded tissue (ED—label '2'), and the necrotic tumor core (NCR—label '1'). ET is generally considered the most active portion of the tumor, described by areas with both visually avid, as well as faintly avid, enhancement on the T1Gd scan. NCR is the necrotic part of the tumor, the appearance of which is hypointense on the T1Gd scan. ED is the peritumoral edematous and infiltrated tissue, defined by the abnormal hyperintense signal envelope on the T2-FLAIR scans, which includes the infiltrative non-enhancing tumor, as well as vasogenic edema in the peritumoral region^{35–38} (an illustration can be seen in Fig. 1b).

Data splits. Once the data were preprocessed, training and validation cohorts were created randomly in a 4:1 ratio, and the splits were preserved during the entire duration of the FL training to prevent data leakage. The performance of every model was compared against the local validation data cohort on every federated round.

Data loading and processing

We leveraged the data loading and processing pipeline from the Generally Nuanced Deep Learning Framework (GaNDLF)⁹⁵, to enable experimentation with various data augmentation techniques. Immediately after data loading, we removed the all-zero axial, coronal, and sagittal planes from the image, and performed a z-score normalization of the non-zero image intensities⁹⁶. Each tumor sub-compartment of the reference label is first split into an individual channel and then passed to the neural network for processing. We extracted a single random patch per mpMRI volume set during every federated round. The patch size was kept constant at [128, 128, 128] to ensure that the trained model can fit the memory of the baseline hardware requirement of each collaborator, i.e., a discrete graphics processing unit with a minimum of 11 GB dedicated memory. For data augmentation, we added random noise augmentation ($\mu = 0.0$, $\sigma = 0.1$) with a probability

of $p = 0.2$, random rotations (90° and 180° , with the axis of rotation being uniformly selected in each case from the set of coronal, sagittal, and axial planes) each with a probability of $p = 0.5$, and a random flip augmentation with a probability of $p = 1.0$ with equal likelihood of flips across the sagittal, coronal, and axial planes.

The neural network architecture

The trained model to delineate the different tumor sub-compartments was based on the popular 3D U-Net with residual connections (3D-ResUNet)^{97–101}, an illustration of which can be seen in the Supplementary Fig. 1. The network had 30 base filters, with a learning rate of $lr = 5 \times 10^{-3}$ optimized using the Adam optimizer¹⁰². For the loss function used in training, we used the generalized DSC score^{103,104} (represented mathematically in Eq. (1)) on the absolute complement of each tumor sub-compartment independently. Such mirrored DSC loss has been shown to capture variations in smaller regions better⁸⁹. No penalties were used in the loss function, due to our use of ‘mirrored’ DSC loss^{105–107}. The final layer of the model was a sigmoid layer, providing three channel outputs for each voxel in the input volume, one output channel per tumor sub-compartment. While the generalized DSC score was calculated using a binarized version of the output (check sigmoid value against the threshold 0.5) for the final prediction, we used the floating point DSC¹⁰⁸ during the training process.

$$DSC = \frac{2|RL \odot PM|_1}{|RL|_1 + |PM|_1} \quad (1)$$

where RL serves as the reference label, PM is the predicted mask, $\odot$ is the Hadamard product¹⁰⁹ (i.e., component-wise multiplication), and $|x|_1$ is the L1-norm¹¹⁰, i.e., the sum of the absolute values of all components).

The Federation

The collaborative network of the present study spans 6 continents (Fig. 1), with data from 71 geographically distinct sites. The training process was initiated when each collaborator securely connected to a central aggregation server, which resided behind a firewall at the University of Pennsylvania. We have identified this FL workflow (based on a central aggregation server) as the optimal for this use-case, following a performance evaluation¹¹ for this very same task, i.e., detecting glioblastoma sub-compartment boundaries. As soon as the secure connection was established, the public initial model was passed to the collaborating site. Using FL based on an aggregation server (refer to supplementary figures for illustration), collaborating sites then trained the same network architecture on their local data for one epoch, and shared model updates with the central aggregation server. The central aggregation server received model updates from all collaborators, combined them (by averaging model parameters) and sent the consensus model back to each collaborator to continue their local training. Each such iteration is called a “federated round”. Based on our previously conducted performance evaluation for this use-case¹¹, we chose to perform aggregation of all collaborator updates in the present study, using the federated averaging (FedAvg) approach¹⁴, i.e., average of collaborator’s model updates weighted according to collaborator’s contributing data. We expect these aggregation strategy choices to be use-case dependent, by providing due consideration to the collaborators’ associated compute and network infrastructure. In this study, all the network communications during the FL model training process were based on TLS¹¹¹, to mitigate potential exposure of information during transit. Additionally, we demonstrated the feasibility of TEEs^{60,61} for federated training by running the aggregator workload on the secure enclaves of Intel’s Secure Guard Extensions (SGX) hardware (Intel® Xeon® E-2286M vPro 8-Core 2.4–5.0GHz Turbo), which ensured the confidentiality of the updates being aggregated and the integrity of the consensus model. TLS and TEEs can

help mitigate some of the security and privacy concerns that remain for FL⁵⁵. After not observing any meaningful changes since round 42, we stopped the training after a total of 73 federated rounds. Additionally, we performed all operations on the aggregator on secure hardware (TEE¹¹²), in order to increase the trust by all parties in the confidentiality of the model updates being computed and shared, as well as to increase the confidence in the integrity of the computations being performed¹¹³.

We followed a staged approach for the training of the global consensus model, starting from a preliminary smaller federation across a subset ($n = 35$) of the participating sites to evaluate the complete process and resolve any initial network issues. Note that 16 of these 35 sites were used to train the public initial model, and used in the preliminary federation as an aggregated dataset. The exact 19 site IDs that participated in the training phase of the preliminary federation, as independent sites are: 2, 3, 9, 14, 22, 23, 24, 27, 28, 29, 31, 33, 36, 37, 41, 46, 53, 54, and 59. The total data held by this smaller federation represented approximately 42% ($n = 2471$) of the data used in the full federation. We also trained individual models (initialized using the public initial model) using centralized training at all sites holding >200 training cases, and performed a comparative evaluation of the consensus model with an ensemble of these “single site models”. The per voxel sigmoid outputs of the ensemble were computed as the average of such outputs over the individual single-site models. As with all other models in this study, binary predictions were computed by comparing these sigmoid outputs to a threshold value of 0.5. The single-site model ensemble utilized (via the data at the single site) approximately 33% of the total data across the federation.

Model runtime in low-resource settings

Clinical environments typically have constrained computational resources, such as the availability of specialized hardware (e.g., DL acceleration cards) and increased memory, which affect the runtime performance of DL inference workloads. Thus, taking into consideration the potential deployment of the final consensus model in such low-resource settings, we decided to proceed with a single 3D-ResUNet, rather than an ensemble of multiple models. This decision ensured a reduced computational burden when compared with running multiple models, which is typically done in academic research projects^{35–38}.

To further facilitate use in low-resource environments, we have provided a post-training run-time optimized¹¹⁴ version of the final consensus model. Graph level optimizations (i.e., operators fusion) were initially applied, followed by optimizations for low precision inference, i.e., converting the floating point single precision model to a fixed precision 8-bit integer model (a process known as “quantization”¹¹⁵). In particular, we used accuracy-aware quantization¹¹⁶, where model layers were iteratively scaled to a lower precision format. These optimizations yielded run-time performance benefits, such as lower inference latency (a platform-dependent $4.48 \times$ average speedup and $2.29 \times$ reduced memory requirement when compared with the original consensus model) and higher throughput (equal to the $4.48 \times$ speedup improvement since the batch size used is equal to 1), while the trade-off was an insignificant ($p_{\text{Average}} < 7 \times 10^{-5}$) drop in the average DSC.

Clinically-deployable consensus models. To further encourage the reproducibility of our study, and considering enhancing the potential impact for the study of the rare disease of glioblastoma, we publicly released the trained models of this study. We specifically released the final *singlet* and *triplet* consensus models, including the complete source code used in the project. Taking into consideration the potential deployment of these models in clinical settings, we refrained from training an ensemble of models (as typically done in academic

Article

<https://doi.org/10.1038/s41467-022-33407-5>

research projects^{35–38}), due to the additional computational burden of running multiple models. Furthermore, to facilitate use in low-resource environments, we also provide a post-training run-time optimized¹¹⁴ version of the final consensus model that obviates the need for any specialized hardware (such as DL acceleration cards) and performs insignificantly different from the final consensus model when evaluated against the centralized out-of-sample data.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets used in this study, from the 71 participating sites, are not made publicly available as a collective data collection due to restrictions imposed by acquiring sites. The public initial model data from 16 sites are publicly available through the BraTS challenge^{35–38} and are available from <https://www.med.upenn.edu/cbica/brats2020>. The data from each of the 55 collaborating sites were neither publicly available during the execution of the study, nor shared among collaborating sites or with the aggregator. They were instead used locally, within each of the acquiring sites, for the training and validation of the global consensus model at each federated round. The anatomical template used for co-registration during preprocessing is the SRI24 atlas⁸² and is available from <https://www.nitrc.org/projects/sri24>.

Source data are provided with this paper. Specifically, we provide the raw data, the associated python scripts, and specific instructions to reproduce the plots of this study in a GitHub repository, at: github.com/FETS-AI/2022_Manuscript_Supplement. The file ‘SourceData.tgz’, in the top directory holds an archive of csv files representing the source data. The python scripts are provided in the ‘scripts’ folder which utilize these source data and save ‘.png’ images to disc and/or print latex code (for tables) to stdout. Furthermore, we have provided three sample validation cases, from the publicly available BraTS dataset, to qualitatively showcase the segmentation differences (small, moderate, and large) across the final global consensus model, the public initial model, and the ground truth annotations in the same GitHub repository.

Code availability

Motivated by findability, accessibility, interoperability, and reusability (FAIR) criteria in scientific research¹¹⁷, all the code used to design the Federated Tumor Segmentation (FeTS) platform¹¹⁸ for this study is available through the FeTS Tool³⁹ and it is available at github.com/FETS-AI/Front-End. The functionality related to preprocessing (i.e., DICOM to NIfTI conversion, population-based harmonized preprocessing, co-registration) and manual refinements of annotation is derived from the open-source Cancer Imaging Phenomics Toolkit (CaPTk, github.com/CBICA/CaPTk)^{51,70,71}. The co-registration is performed using the Greedy framework¹¹⁹, available via CaPTk^{51,70,71}, ITK-SNAP⁵², and the FeTS Tool³⁹. The brain extraction is done using the BrainMaGe method⁸⁷, and is available at github.com/CBICA/BrainMaGe, and via GaNDLF⁹⁵ at github.com/mlcommons/GaNDLF. To generate automated annotations, DeepMedic’s⁴⁹ integration with CaPTk was used, and we used the model weights and inference mechanism provided by the other algorithm developers (DeepScan⁸⁸ and nnU-Net⁸⁹ (github.com/MIC-DKFZ/nnunet)). DeepMedic’s original implementation is available in github.com/deepmedic/deepmedic, whereas the one we used in this study can be found at github.com/CBICA/deepmedic. The fusion of the labels was done using the Label Fusion tool¹²⁰ available at github.com/FETS-AI/LabelFusion. The data loading pipeline and network architecture were developed using the GaNDLF framework⁹⁵ by using PyTorch¹²¹. The data augmentation was done via GaNDLF by leveraging TorchIO¹²². The FL backend developed for this project has been open-sourced as a separate software library,

to encourage further research on FL¹²³ and is available at github.com/intel/openfl. The optimization of the consensus model inference workload was performed via OpenVINO¹²⁴ (github.com/openvinotoolkit/openvino/tree/2021.4.1), which is an open-source toolkit enabling acceleration of neural network models through various optimization techniques. The optimizations were evaluated on an Intel Core® i7-1185G7E CPU @ 2.80 GHz with 2 × 8 GB DDR4 3200 MHz memory on Ubuntu 18.04.6 OS and Linux kernel version 5.9.0-050900-generic.

References

- Mårtensson, G. et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. *Med. Image Anal.* **66**, 101714 (2020).
- Zech, J. R. et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med.* **15**, e1002683 (2018).
- Obermeyer, Z. & Emanuel, E. J. Predicting the future-big data, machine learning, and clinical medicine. *New Engl. J. Med.* **375**, 1216 (2016).
- Marcus, G. Deep learning: a critical appraisal. arXiv preprint arXiv:1801.00631 (2018).
- Aggarwal, C. C. et al. *Neural Networks and Deep Learning* Vol. 10, 978–983 (Springer, 2018).
- Thompson, P. M. et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
- Consortium, T. G. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro-Oncology* **20**, 873–884 (2018).
- Davatzikos, C. et al. AI-based prognostic imaging biomarkers for precision neuro-oncology: the RESPOND Consortium. *Neuro-oncology* **22**, 886–888 (2020).
- Bakas, S. et al. iglass: imaging integration into the glioma longitudinal analysis consortium. *Neuro-oncology* **22**, 1545–1546 (2020).
- Rieke, N. et al. The future of digital health with federated learning. *NPJ Digit. Med.* **3**, 1–7 (2020).
- Sheller, M. J. et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci. Rep.* **10**, 1–12 (2020).
- Annas, G. J. et al. HIPAA regulations—a new era of medical-record privacy? *New Engl. J. Med.* **348**, 1486–1490 (2003).
- Voigt, P. & Von dem Bussche, A. The EU General Data Protection Regulation (GDPR). In *A Practical Guide* 1st edition, Vol. 10(3152676), 10–5555 (Springer, 2017).
- McMahan, B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics* (eds Singh, A. & Zhu, J.) 1273–1282 (PMLR, 2017).
- Sheller, M. J., Reina, G. A., Edwards, B., Martin, J. & Bakas, S. Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. et al.) 92–104 (Springer, 2018).
- Dayan, I. et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat. Med.* **27**, 1735–1743 (2021).
- Chang, K. et al. Distributed deep learning networks among institutions for medical imaging. *J. Am. Med. Inform. Assoc.* **25**, 945–954 (2018).
- Nilsson, A., Smith, S., Ulm, G., Gustavsson, E. & Jirstrand, M. A performance evaluation of federated learning algorithms. In *Proceedings of the Second Workshop on Distributed Infrastructures for Deep Learning*, 1–8 (Association for Computing Machinery, New York, 2018).

19. Sarma, K. V. et al. Federated learning improves site performance in multicenter deep learning without data sharing. *J. Am. Med. Inform. Assoc.* **28**, 1259–1264 (2021).
20. Shen, C. et al. Multi-task federated learning for heterogeneous pancreas segmentation. In *Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning* (eds Laura, C. O. et al.) 101–110 (Springer, 2021).
21. Yang, D. et al. Federated semi-supervised learning for covid region segmentation in chest ct using multi-national data from China, Italy, Japan. *Med. Image Anal.* **70**, 101992 (2021).
22. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
23. Hannun, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **25**, 65–69 (2019).
24. Griggs, R. C. et al. Clinical research for rare disease: opportunities, challenges, and solutions. *Mol. Genet. Metab.* **96**, 20–26 (2009).
25. Shukla, G. et al. Advanced magnetic resonance imaging in glioblastoma: a review. *Chin. Clin. Oncol.* **6**, 40 (2017).
26. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
27. Verhaak, R. G. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *pdgfra*, *idh1*, *egfr*, and *nfi*. *Cancer Cell* **17**, 98–110 (2010).
28. Sottoriva, A. et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl Acad. Sci. USA* **110**, 4009–4014 (2013).
29. Ostrom, Q. T. et al. Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. *Neuro-oncology* **21**, v1–v100 (2019).
30. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-oncology* **23**, 1231–1251 (2021).
31. Han, W. et al. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *Am. J. Neuroradiol.* **41**, 40–48 (2020).
32. Roth, H. R. et al. Federated learning for breast density classification: a real-world implementation. In *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning* (eds Albarqouni, S. et al.) 181–191 (Springer, 2020).
33. Chaichana, K. L. et al. Multi-institutional validation of a pre-operative scoring system which predicts survival for patients with glioblastoma. *J. Clin. Neurosci.* **20**, 1422–1426 (2013).
34. Fathi Kazerooni, A. et al. Cancer imaging phenomics via captk: multi-institutional prediction of progression-free survival and pattern of recurrence in glioblastoma. *JCO Clin. Cancer Inform.* **4**, 234–244 (2020).
35. Menze, B. H. et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2014).
36. Bakas, S. et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. data* **4**, 1–13 (2017).
37. Bakas, S. et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629 (2018).
38. Baid, U. et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021).
39. Pati, S. et al. The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research. *Phys Med Biol.* **67**, 204002 (2022).
40. Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. Transfusion: Understanding transfer learning for medical imaging. *Proceedings of the 33rd International Conference on Neural Information Processing Systems* **32**, 3347–3357 (Association for Computing Machinery, 2019).
41. Young, J. C. & Suryadibrata, A. Applicability of various pre-trained deep convolutional neural networks for pneumonia classification based on x-ray images. *Int. J. Adv. Trends Comput. Sci. Eng.* **9**, 2649–2654 (2020).
42. Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New Engl. J. Med.* **352**, 987–996 (2005).
43. Beiko, J. et al. *Idh1* mutant malignant astrocytomas are more amenable to surgical resection and have a survival benefit associated with maximal surgical resection. *Neuro-oncology* **16**, 81–91 (2014).
44. Olson, J. J. Congress of neurological surgeons systematic review and evidence-based guidelines for the treatment of adults with progressive glioblastoma update: introduction and methods. *J. Neuro-oncol* **158**, 133–137 (2022).
45. Curry, W. T. & Barker, F. G. Racial, ethnic and socioeconomic disparities in the treatment of brain tumors. *J. Neuro-oncol.* **93**, 25–39 (2009).
46. Marsland, S. Novelty detection in learning systems. *Neural Comput. Surv.* **3**, 157–195 (2003).
47. Mazzara, G. P., Velthuisen, R. P., Pearlman, J. L., Greenberg, H. M. & Wagner, H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *Int. J. Radiat. Oncol. * Biol. * Phys.* **59**, 300–312 (2004).
48. Mitchell, J. R. et al. Deep neural network to locate and segment brain tumors outperformed the expert technicians who created the training data. *J. Med. Imaging* **7**, 055501 (2020).
49. Kamnitsas, K. et al. Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017).
50. Rudie, J. D. et al. Multi-disease segmentation of gliomas and white matter hyperintensities in the brats data using a 3d convolutional neural network. *Front. Comput. Neurosci.* **13**, 84 (2019).
51. Davatzikos, C. et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. *J. Med. Imaging* **5**, 011018 (2018).
52. Yushkevich, P. A. et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
53. Kikinis, R., Pieper, S. D. & Vosburgh, K. G. 3d slicer: a platform for subject-specific image analysis, visualization, and clinical support. In *Intraoperative imaging and Image-guided Therapy* (ed. Jolesz, F. A.) 277–289 (Springer, 2014).
54. Kofler, F. et al. Brats toolkit: translating brats brain tumor segmentation algorithms into clinical and scientific practice. *Front. Neurosci.* **125**, 125–125 (2020).
55. Kairouz, P. et al. Advances and open problems in federated learning. *Found. Trends® in Mach. Learn.* **14**, 1–210 (2021).
56. Nasr, M., Shokri, R. & Houmansadr, A. Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, 739–753 (IEEE, 2019).
57. Lam, M., Wei, G.-Y., Brooks, D., Reddi, V. J. & Mitzenmacher, M. Gradient disaggregation: breaking privacy in federated learning by reconstructing the user participant matrix. In *International Conference on Machine Learning*, 5959–5968 (PMLR, 2021).
58. Gentry, C. Fully homomorphic encryption using ideal lattices. In *Proc. 41st Annual ACM Symposium on Theory of Computing*, 169–178 (Association for Computing Machinery, New York, 2009).

59. Yao, A. C. Protocols for secure computations. In *23rd Annual Symposium on Foundations of Computer Science (SFCS 1982)*, 160–164 (IEEE, 1982).
60. Sabt, M., Achemlal, M. & Bouabdallah, A. Trusted execution environment: what it is, and what it is not. In *2015 IEEE Trustcom/BigDataSE/ISPA Vol. 1*, 57–64 (IEEE, 2015).
61. Schneider, M., Masti, R. J., Shinde, S., Capkun, S. & Perez, R. Sok: Hardware-supported trusted execution environments. arXiv preprint arXiv:2205.12742 (2022).
62. Dwork, C. Differential privacy: a survey of results. In *International Conference on Theory and Applications of Models of Computation* 1–19 (Springer, 2008).
63. Wei, K. et al. Federated learning with differential privacy: algorithms and performance analysis. *IEEE Trans. Inf. Forensics Secur.* **15**, 3454–3469 (2020).
64. Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W. & Tizhoosh, H. R. Federated learning and differential privacy for medical image analysis. *Sci. Rep.* **12**, 1–10 (2022).
65. Tramer, F. & Boneh, D. Slalom: fast, verifiable and private execution of neural networks in trusted hardware. arXiv preprint arXiv:1806.03287 (2018).
66. Kalra, S., Wen, J., Cresswell, J. C., Volkovs, M. & Tizhoosh, H. R. Proxyfl: decentralized federated learning through proxy model sharing. arXiv preprint arXiv:2111.11343 (2021).
67. Lu, M. Y. et al. Federated learning for computational pathology on gigapixel whole slide images. *Med. Image Anal.* **76**, 102298 (2022).
68. Baid, U. et al. Federated learning for the classification of tumor infiltrating lymphocytes. arXiv preprint arXiv:2203.16622 (2022).
69. Linardos, A., Kushibar, K., Walsh, S., Gkontra, P. & Lekadir, K. Federated learning for multi-center imaging diagnostics: a simulation study in cardiovascular disease. *Sci. Rep.* **12**, 1–12 (2022).
70. Rathore, S. et al. Brain cancer imaging phenomics toolkit (brain-captk): an interactive platform for quantitative analysis of glioblastoma. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. et al.) 133–145 (Springer, 2017).
71. Pati, S. et al. The cancer imaging phenomics toolkit (captk): technical overview. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. & Bakas, S.) 380–394 (Springer, 2019).
72. Gilbert, M. R. et al. Rtoq 0825: Phase iii double-blind placebo-controlled trial evaluating bevacizumab (bev) in patients (pts) with newly diagnosed glioblastoma (gbm). *J. Clin. Oncol.* **31**(18_suppl18), 1–1 (2013).
73. Gilbert, M. R. et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *New Engl. J. Med.* **370**, 699–708 (2014).
74. Boxerman, J. L. et al. Prognostic value of contrast enhancement and flair for survival in newly diagnosed glioblastoma treated with and without bevacizumab: results from acrin 6686. *Neuro-oncology* **20**, 1400–1410 (2018).
75. Schmainda, K. M. et al. Value of dynamic contrast perfusion mri to predict early response to bevacizumab in newly diagnosed glioblastoma: results from acrin 6686 multicenter trial. *Neuro-oncology* **23**, 314–323 (2021).
76. Pianykh, O. S. *Digital Imaging and Communications in Medicine (DICOM): a Practical Introduction and Survival Guide* (Springer, 2012).
77. Kahn, C. E., Carrino, J. A., Flynn, M. J., Peck, D. J. & Horii, S. C. Dicom and radiology: past, present, and future. *J. Am. College Radiol.* **4**, 652–657 (2007).
78. Mustra, M., Delac, K. & Grgic, M. Overview of the dicom standard. In *2008 50th International Symposium ELMAR Vol. 1*, 39–44 (IEEE, 2008).
79. Cox, R. et al. A (sort of) new image data format standard: Nifti-1. In *Proc. 10th Annual Meeting of the Organization for Human Brain Mapping* 22 (Wiley, 2004).
80. Li, X., Morgan, P. S., Ashburner, J., Smith, J. & Rorden, C. The first step for neuroimaging data analysis: Dicom to nifti conversion. *J. Neurosci. Methods* **264**, 47–56 (2016).
81. White, T., Blok, E. & Calhoun, V. D. Data sharing and privacy issues in neuroimaging research: opportunities, obstacles, challenges, and monsters under the bed. *Hum. Brain Mapp.* **43**, 278–291 (2020).
82. Rohlfing, T., Zahr, N. M., Sullivan, E. V. & Pfefferbaum, A. The sri24 multichannel atlas of normal adult human brain structure. *Hum. Brain Mapp.* **31**, 798–819 (2010).
83. Song, S., Zheng, Y. & He, Y. A review of methods for bias correction in medical images. *Biomed. Eng. Rev.* **1**, 2375–9151 (2017).
84. Sled, J. G., Zijdenbos, A. P. & Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans. Med. Imaging* **17**, 87–97 (1998).
85. Tustison, N. J. et al. N4itk: improved n3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
86. Schwarz, C. G. et al. Identification of anonymous mri research participants with face-recognition software. *New Engl. J. Med.* **381**, 1684–1686 (2019).
87. Thakur, S. et al. Brain extraction on MRI scans in presence of diffuse glioma: Multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training. *NeuroImage* **220**, 117081 (2020).
88. McKinley, R., Meier, R. & Wiest, R. Ensembles of densely-connected cnns with label-uncertainty for brain tumor segmentation. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. et al.) 456–465 (Springer, 2018).
89. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
90. Rohlfing, T., Russakoff, D. B. & Maurer, C. R. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans. Med. Imaging* **23**, 983–994 (2004).
91. Warfield, S. K., Zou, K. H. & Wells, W. M. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* **23**, 903–921 (2004).
92. Rohlfing, T. & Maurer Jr, C. R. Multi-classifier framework for atlas-based image segmentation. *Pattern Recognit. Lett.* **26**, 2070–2079 (2005).
93. Huo, J., Wang, G., Wu, Q. J. & Thangarajah, A. Label fusion for multi-atlas segmentation based on majority voting. In *International Conference Image Analysis and Recognition* (eds Kamel, M. & Campilho, A.) 100–106 (Springer, 2015).
94. Langerak, T. R. et al. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple). *IEEE Trans. Med. Imaging* **29**, 2000–2008 (2010).
95. Pati, S. et al. Gandlf: a generally nuanced deep learning framework for scalable end-to-end clinical workflows in medical imaging. arXiv preprint arXiv:2103.01006 (2021).
96. Reinhold, J. C., Dewey, B. E., Carass, A. & Prince, J. L. Evaluating the impact of intensity normalization on MR image synthesis. In *Medical Imaging 2019: Image Processing*, Vol. 10949 (eds Angelini, E. D. & Landman, B. A.) 109493H (International Society for Optics and Photonics, 2019).
97. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A.) 234–241 (Springer, 2015).

98. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (eds Ourselin, S. et al.) 424–432 (Springer, 2016).
99. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (IEEE, 2016).
100. Drozdal, M., Vorontsov, E., Chartrand, G., Kadoury, S. & Pal, C. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications* (eds Carneiro, G. et al.) 179–187 (Springer, 2016).
101. Bhalerao, M. & Thakur, S. Brain tumor segmentation based on 3d residual u-net. In *International MICCAI Brainlesion Workshop* (eds Crimi, A. & Bakas, S.) 218–225 (Springer, 2019).
102. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
103. 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 *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (eds Cardoso, M. J. et al.) 240–248 (Springer, 2017).
104. Zijdenbos, A. P., Dawant, B. M., Margolin, R. A. & Palmer, A. C. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans. Med. Imaging* **13**, 716–724 (1994).
105. Chen, L., Qu, H., Zhao, J., Chen, B. & Principe, J. C. Efficient and robust deep learning with correntropy-induced loss function. *Neural Comput. Appl.* **27**, 1019–1031 (2016).
106. Salehi, S. S. M., Erdogmus, D. & Gholipour, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International Workshop on Machine Learning in Medical Imaging* (eds Wang, Q., Shi, Y., Suk, H. & Suzuki, K.) 379–387 (Springer, 2017).
107. Caliva, F., Iriando, C., Martinez, A. M., Majumdar, S. & Pedoia, V. Distance map loss penalty term for semantic segmentation. arXiv preprint arXiv:1908.03679 (2019).
108. Shamir, R. R., Duchin, Y., Kim, J., Sapiro, G. & Harel, N. Continuous dice coefficient: a method for evaluating probabilistic segmentations. arXiv preprint arXiv:1906.11031 (2019).
109. Horn, R. A. The hadamard product. In *Proc. Symposium on Applied Mathematics*, Vol. 40 (eds Berghel, H. & Talburt, J.) 87–169 (American Mathematical Society, 1990).
110. Barrodale, I. L1 approximation and the analysis of data. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)* **17**, 51–57 (1968).
111. Knauth, T. et al. Integrating remote attestation with transport layer security. arXiv preprint arXiv:1801.05863 (2018).
112. Kaissis, G. A., Makowski, M. R., Rückert, D. & Braren, R. F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat. Mach. Intell.* **2**, 305–311 (2020).
113. Ekberg, J.-E., Kostiaainen, K. & Asokan, N. The untapped potential of trusted execution environments on mobile devices. *IEEE Secur. Priv.* **12**, 29–37 (2014).
114. Rodriguez, A. et al. Lower numerical precision deep learning inference and training. *Intel White Paper* **3**, 1–19 (2018).
115. Lin, D., Talathi, S. & Annapureddy, S. Fixed point quantization of deep convolutional networks. In *International Conference on Machine Learning* (eds Balcan, M. F. & Weinberger, K. Q.) 2849–2858 (PMLR, 2016).
116. Vakili, S., Langlois, J. P. & Bois, G. Enhanced precision analysis for accuracy-aware bit-width optimization using affine arithmetic. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.* **32**, 1853–1865 (2013).
117. Wilkinson, M. D. et al. The fair guiding principles for scientific data management and stewardship. *Sci. data* **3**, 1–9 (2016).
118. Pati, S. & Bakas, S. S. Fets-ai/front-end: release for zenodo <https://doi.org/10.5281/zenodo.7036038> (2022)
119. Yushkevich, P. A. et al. Fast automatic segmentation of hippocampal subfields and medial temporal lobe subregions in 3 tesla and 7 tesla t2-weighted MRI. *Alzheimer's Dement.* **12**, P126–P127 (2016).
120. Pati, S. & Bakas, S. LabelFusion: medical Image label fusion of segmentations <https://doi.org/10.5281/zenodo.4633206> (2021)
121. Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* vol. 32 (eds Wallach, H. M. et al.) 8026–8037 (Neural Information Processing Systems Foundation, Inc., 2019).
122. Pérez-García, F., Sparks, R. & Ourselin, S. Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput. Methods Programs Biomed.* **208**, 106236 (2021).
123. Foley, P. et al. OpenFL: the open federated learning library. *Physics in Medicine & Biology* (2022). Online ahead of print.
124. Gorbachev, Y. et al. Openvino deep learning workbook: Comprehensive analysis and tuning of neural networks inference. In *Proc. IEEE/CVF International Conference on Computer Vision Workshops*, 783–787 (IEEE, 2019).

Acknowledgements

Research and main methodological developments reported in this publication were partly supported by the National Institutes of Health (NIH) under award numbers NIH/NCI:U01CA242871 (S. Bakas), NIH/NINDS:R01NS042645 (C. Davatzikos), NIH/NCI:U24CA189523 (C. Davatzikos), NIH/NCI:U24CA215109 (J. Saltz), NIH/NCI:U01CA248226 (P. Tiwari), NIH/NCI:P30CA51008 (Y. Gusev), NIH:R50CA211270 (M. Muzi), NIH/NCATS:UL1TR001433 (Y. Yuan), NIH/NIBIB:R21EB030209 (Y. Yuan), NIH/NCI:R37CA214955 (A. Rao), and NIH:R01CA233888 (A.L. Simpson). The authors would also like to acknowledge the following NIH funded awards for the multi-site clinical trial (NCT00884741, RTOG0825/ACRIN6686): U10CA21661, U10CA37422, U10CA180820, U10CA180794, U01CA176110, R01CA082500, CA079778, CA080098, CA180794, CA180820, CA180822, CA180868. Research reported in this publication was also partly supported by the National Science Foundation, under award numbers 2040532 (S. Baek), and 2040462 (B. Landman). Research reported in this publication was also supported by i) a research grant from Varian Medical Systems (Palo Alto, CA, USA) (Y. Yuan), ii) the Ministry of Health of the Czech Republic (Grant Nr. NU21-08-00359) (M. Kerkovský and M. Kozubek), iii) Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 404521405, SFB 1389, Work Package CO2, and Priority Program 2177 “Radiomics: Next Generation of Biomedical Imaging” (KI 2410/1-1 | MA 6340/18-1) (P. Vollmuth), iv) DFG Project-ID B12, SFB 824 (B. Wiestler), v) the Helmholtz Association (funding number ZT-I-0014) (K. Maier-Hein), vi) the Dutch Cancer Society (KWF project number EMCR 2015-7859) (S.R. van der Voort), vii) the Chilean National Agency for Research and Development (ANID-Basal FB0008 (AC3E) and FB210017 (CENIA)) (P. Guevara), viii) the Canada CIFAR AI Chairs Program (M. Vallières), ix) Leeds Hospital Charity (Ref: 9R01/1403) (S. Currie), x) the Cancer Research UK funding for the Leeds Radiotherapy Research Centre of Excellence (RadNet) and the grant number C19942/A28832 (S. Currie), xi) Medical Research Council (MRC) Doctoral Training Program in Precision Medicine (Award Reference No. 2096671) (J. Bernal), xii) The European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant Agreement No. 757173) (B. Glocker), xiii) The UKRI London Medical Imaging & Artificial Intelligence Centre for Value-Based Healthcare (K. Kamnitsas), xiv) Wellcome/Engineering and Physical Sciences Research Council (EPSRC) Center for Medical Engineering (WT 203148/Z/16/Z) (T.C. Booth), xv) American Cancer Society Research Scholar Grant RSG-16-005-01 (A. Rao), xvi) the Department of Defense (DOD)

Peer Reviewed Cancer Research Program (PRCRP) W81XWH-18-1-0404, Dana Foundation David Mahoney Neuroimaging Program, the V Foundation Translational Research Award, Johnson & Johnson WiSTEM2D Award (P. Tiwari), (xvii) RSNA Research & Education Foundation under grant number RR2011 (E. Calabrese), (xviii) the National Research Fund of Luxembourg (FNR) (grant number: C20/BM/14646004/GLASS-LUX/Niclou) (S.P. Niclou), (xix) EU Marie Curie FP7-PEOPLE-2012-ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958) (J. Slotboom), and (xx) CNPq 303808/2018-7 and FAPESP 2014/12236-1 (A. Xavier Falcão). The content of this publication is solely the responsibility of the authors and does not represent the official views of the NIH, the NSF, the RSNA R&E Foundation, or any of the additional funding bodies.

Author contributions

Study conception: S. Pati, U. Baid, B. Edwards, M. Sheller, G.A. Reina, J. Martin, S. Bakas. Development of software used in the study: S. Pati, B. Edwards, M. Sheller, S. Wang, G.A. Reina, P. Foley, A. Gruzdev, D. Karkada, S. Bakas. Data acquisition: M. Bilello, S. Mohan, E. Calabrese, J. Rudie, J. Saini, R.Y. Huang, K. Chang, T. So, P. Heng, T.F. Cloughesy, C. Raymond, T. Oughourlian, A. Hagiwara, C. Wang, M. To, M. Kerkovský, T. Koprivová, M. Dostál, V. Vybíhal, J.A. Maldjian, M.C. Pinho, D. Reddy, J. Holcomb, B. Wiestler, M. Metz, R. Jain, M. Lee, P. Tiwari, R. Verma, Y. Gusev, K. Bhuvaneshwar, C. Bencheqroun, A. Belouali, A. Abayazeed, A. Abbassy, S. Gamal, M. Qayati, M. Mekhaïmar, M. Reyes, R.R. Colen, M. Ak, P. Vollmuth, G. Brugnara, F. Sahm, M. Bendszus, W. Wick, A. Mahajan, C. Balaña Quintero, J. Capellades, J. Puig, Y. Choi, M. Muzi, H.F. Shaykh, A. Herrera-Trujillo, W. Escobar, A. Abello, P. LaMontagne, B. Landman, K. Ramadass, K. Xu, S. Chotai, L.B. Chambless, A. Mistry, R.C. Thompson, J. Bapuraj, N. Wang, S.R. van der Voort, F. Incekara, M.M.J. Wijnenga, R. Gahrman, J.W. Schouten, H.J. Dubbink, A.J.P.E. Vincent, M.J. van den Bent, H.I. Sair, C.K. Jones, A. Venkataraman, J. Garrett, M. Larson, B. Menze, T. Weiss, M. Weller, A. Bink, Y. Yuan, S. Sharma, T. Tseng, B.C.A. Teixeira, F. Sprenger, S.P. Niclou, O. Keunen, L.V.M. Dixon, M. Williams, R.G.H. Beets-Tan, H. Franco-Maldonado, F. Loayza, J. Slotboom, P. Radojewski, R. Meier, R. Wiest, J. Trenkler, J. Pichler, G. Necker, S. Meckel, E. Torche, F. Vera, E. Lóópez, Y. Kim, H. Ismael, B. Allen, J.M. Buatti, J. Park, P. Zampakis, V. Panagiotopoulos, P. Tsiganos, E. Chaliasos, D.M. Kardamakis, P. Prasanna, K.M. Mani, D. Payne, T. Kurc, L. Poisson, M. Vallières, D. Fortin, M. Lepage, F. Morón, J. Mandel, C. Badve, A.E. Sloan, J.S. Barnholtz-Sloan, K. Waite, G. Shukla, S. Liem, G.S. Alexandre, J. Lombardo, J.D. Palmer, A.E. Flanders, A.P. Dicker, G. Ogbale, D. Oyekunle, O. Odafe-Oyibotha, B. Osobu, M. Shu'aibu, F. Dako, A. Dorcas, D. Murcia, R. Haas, J. Thompson, D.R. Ormond, S. Currie, K. Fatania, R. Frood, J. Mitchell, J. Farinhas, A.L. Simpson, J.J. Peoples, R. Hu, D. Cutler, F.Y. Moraes, A. Tran, M. Hamghalam, M.A. Boss, J. Gimpel, B. Bialecki, A. Chelliah. Data processing: C. Sako, S. Ghodasara, E. Calabrese, J. Rudie, M. Jadhav, U. Pandey, R.Y. Huang, M. Jiang, C. Chen, C. Raymond, S. Bhardwaj, C. Chong, M. Agzarian, M. Kozubek, F. Lux, J. Michálek, P. Matula, C. Bangalore Yogananda, D. Reddy, B.C. Wagner, I. Ezhov, M. Lee, Y.W. Lui, R. Verma, R. Bareja, I. Yadav, J. Chen, N. Kumar, K. Bhuvaneshwar, A. Sayah, C. Bencheqroun, K. Kolodziej, M. Hill, M. Reyes, L. Pei, M. Ak, A. Kotrotsou, P. Vollmuth, G. Brugnara, C.J. Preetha, M. Zenk, J. Puig, M. Muzi, H.F. Shaykh, A. Abello, J. Bernal, J. Gómez, P. LaMontagne, K. Ramadass, S. Chotai, N. Wang, M. Smits, S.R. van der Voort, A. Alafandi, F. Incekara, M.M.J. Wijnenga, G. Kapsas, R. Gahrman, A.J.P.E. Vincent, P.J. French, S. Klein, H.I. Sair, C.K. Jones, J. Garrett, H. Li, F. Kofler, Y. Yuan, S. Adabi, A. Xavier Falcão, S.B. Martins, D. Menotti, D.R. Lucio, O. Keunen, A. Hau, K. Kamnitsas, L. Dixon, S. Benson, E. Pelaez, H. Franco-Maldonado, F. Loayza, S. Quevedo, R. McKinley, J. Trenkler, A. Haunschild, C. Mendoza, E. Ríos, J. Choi, S. Baek, J. Yun, P. Zampakis,

V. Panagiotopoulos, P. Tsiganos, E.I. Zacharakis, C. Kalogeropoulos, P. Prasanna, S. Shreshtha, T. Kurc, B. Luo, N. Wen, M. Vallières, D. Fortin, F. Morón, C. Badve, V. Vadmal, G. Shukla, G. Ogbale, D. Oyekunle, F. Dako, D. Murcia, E. Fu, S. Currie, R. Frood, M.A. Vogelbaum, J. Mitchell, J. Farinhas, J.J. Peoples, M. Hamghalam, D. Kattil Veetil, K. Schmidt, B. Bialecki, S. Marella, T.C. Booth, A. Chelliah, M. Modat, C. Dragos, H. Shuaib. Data analysis & interpretation: S. Pati, U. Baid, B. Edwards, M. Sheller, S. Bakas. Site PI/Senior member (of each collaborating group): C. Davatzikos, J. Villanueva-Meyer, M. Ingallhalikar, R.Y. Huang, Q. Dou, B.M. Ellingson, M. To, M. Kozubek, J.A. Maldjian, B. Wiestler, R. Jain, P. Tiwari, Y. Gusev, A. Abayazeed, R.R. Colen, P. Vollmuth, A. Mahajan, C. Balaña Quintero, S. Lee, M. Muzi, H.F. Shaykh, M. Trujillo, D. Marcus, B. Landman, A. Rao, M. Smits, H.I. Sair, R. Jeraj, B. Menze, Y. Yuan, A. Xavier Falcão, S.P. Niclou, B. Glocker, J. Teuwen, E. Pelaez, R. Wiest, S. Meckel, P. Guevara, S. Baek, H. Kim, D.M. Kardamakis, J. Saltz, L. Poisson, M. Vallières, F. Morón, A.E. Sloan, A.E. Flanders, G. Ogbale, D.R. Ormond, S. Currie, J. Farinhas, A.L. Simpson, C. Apgar, T.C. Booth. Writing the original manuscript: S. Pati, U. Baid, B. Edwards, M. Sheller, S. Bakas. Review, edit, & approval of the final manuscript: All authors.

Competing interests

The Intel-affiliated authors (B. Edwards, M. Sheller, S. Wang, G.A. Reina, P. Foley, A. Gruzdev, D. Karkada, P. Shah, J. Martin) would like to disclose the following (potential) competing interests as Intel employees. Intel may develop proprietary software that is related in reputation to the OpenFL open source project highlighted in this work. In addition, the work demonstrates feasibility of federated learning for brain tumor boundary detection models. Intel may benefit by selling products to support an increase in demand for this use-case. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-33407-5>.

Correspondence and requests for materials should be addressed to Spyridon Bakas.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Sarthak Pati^{1,2,3,4,154}, Ujjwal Baid^{1,2,3,154}, Brandon Edwards^{5,154}, Micah Sheller⁵, Shih-Han Wang⁵, G. Anthony Reina⁵, Patrick Foley⁵, Alexey Gruzdev⁵, Deepthi Karkada⁵, Christos Davatzikos^{1,2}, Chiharu Sako^{1,2}, Satyam Ghodasara², Michel Bilello^{1,2}, Suyash Mohan^{1,2}, Philipp Vollmuth⁶, Gianluca Brugnara⁶, Chandrakanth J. Preetha⁶, Felix Sahm^{7,8}, Klaus Maier-Hein^{9,10}, Maximilian Zenk⁹, Martin Bendszus⁶, Wolfgang Wick^{7,11}, Evan Calabrese¹², Jeffrey Rudie¹², Javier Villanueva-Meyer¹², Soonmee Cha¹², Madhura Ingalthalikar¹³, Manali Jadhav¹³, Umang Pandey¹³, Jitender Saini¹⁴, John Garrett^{15,16}, Matthew Larson¹⁵, Robert Jeraj^{15,16}, Stuart Currie¹⁷, Russell Froid¹⁷, Kavi Fatania¹⁷, Raymond Y. Huang¹⁸, Ken Chang¹⁹, Carmen Balaña Quintero²⁰, Jaume Capellades²¹, Josep Puig²², Johannes Trenkler²³, Josef Pichler²⁴, Georg Necker²³, Andreas Haunschmidt²³, Stephan Meckel^{23,25}, Gaurav Shukla^{1,26}, Spencer Liem²⁷, Gregory S. Alexander²⁸, Joseph Lombardo^{27,29}, Joshua D. Palmer³⁰, Adam E. Flanders³¹, Adam P. Dicker²⁹, Haris I. Sair^{32,33}, Craig K. Jones³³, Archana Venkataraman³⁴, Meirui Jiang³⁵, Tiffany Y. So³⁵, Cheng Chen³⁵, Pheng Ann Heng³⁵, Qi Dou³⁵, Michal Kozubek³⁶, Filip Lux³⁶, Jan Michálek³⁶, Petr Matula³⁶, Miloš Keřkovský³⁷, Tereza Kopřivová³⁷, Marek Dostál^{37,38}, Václav Vybihal³⁹, Michael A. Vogelbaum⁴⁰, J. Ross Mitchell^{41,42}, Joaquim Farinhas⁴³, Joseph A. Maldjian⁴⁴, Chandan Ganesh Bangalore Yogananda⁴⁴, Marco C. Pinho⁴⁴, Divya Reddy⁴⁴, James Holcomb⁴⁴, Benjamin C. Wagner⁴⁴, Benjamin M. Ellingson^{45,46}, Timothy F. Cloughesy⁴⁶, Catalina Raymond⁴⁵, Talia Oughourlian^{45,47}, Akifumi Hagiwara⁴⁷, Chencai Wang⁴⁷, Minh-Son To^{48,49}, Sargam Bhardwaj⁴⁸, Chee Chong⁵⁰, Marc Agzarian^{50,51}, Alexandre Xavier Falcão⁵², Samuel B. Martins⁵³, Bernardo C. A. Teixeira^{54,55}, Flávia Sprenger⁵⁵, David Menotti⁵⁶, Diego R. Lucio⁵⁶, Pamela LaMontagne⁵⁷, Daniel Marcus⁵⁷, Benedikt Wiestler^{58,59}, Florian Kofler^{58,59,60}, Ivan Ezhov^{4,59,60}, Marie Metz⁵⁸, Rajan Jain^{61,62}, Matthew Lee⁶¹, Yvonne W. Lui⁶¹, Richard McKinley⁶³, Johannes Slotboom⁶³, Piotr Radojewski⁶³, Raphael Meier⁶³, Roland Wiest⁶³, Derrick Murcia⁶⁴, Eric Fu⁶⁴, Rourke Haas⁶⁴, John Thompson⁶⁴, David Ryan Ormond⁶⁴, Chaitra Badve⁶⁵, Andrew E. Sloan^{66,67,68}, Vachan Vadmal⁶⁸, Kristin Waite⁶⁹, Rivka R. Colen^{70,71}, Linmin Pei⁷², Murat Ak⁷⁰, Ashok Srinivasan⁷³, J. Rajiv Bapuraj⁷³, Arvind Rao⁷⁴, Nicholas Wang⁷⁴, Ota Yoshiaki⁷³, Toshio Moritani⁷³, Sevcen Turk⁷³, Joonsang Lee⁷⁴, Snehal Prabhudesai⁷⁴, Fanny Morón⁷⁵, Jacob Mandel⁵¹, Konstantinos Kamnitsas^{76,77}, Ben Glocker⁷⁶, Luke V. M. Dixon⁷⁸, Matthew Williams⁷⁹, Peter Zampakis⁸⁰, Vasileios Panagiotopoulos⁸¹, Panagiotis Tsiganos⁸², Sotiris Alexiou⁸³, Ilias Haliassos⁸⁴, Evangelia I. Zacharaki⁸³, Konstantinos Moustakas⁸³, Christina Kalogeropoulou⁸⁰, Dimitrios M. Kardamakis⁸⁵, Yoon Seong Choi⁸⁶, Seung-Koo Lee⁸⁶, Jong Hee Chang⁸⁶, Sung Soo Ahn⁸⁶, Bing Luo⁸⁷, Laila Poisson⁸⁸, Ning Wen^{87,89}, Pallavi Tiwari⁹⁰, Ruchika Verma^{42,90}, Rohan Bareja⁹⁰, Ipsa Yadav⁹⁰, Jonathan Chen⁹⁰, Neeraj Kumar^{41,42}, Marion Smits⁹¹, Sebastian R. van der Voort⁹¹, Ahmed Alafandi⁹¹, Fatih Incekara^{91,92}, Maarten M. J. Wijnenga⁹³, Georgios Kapsas⁹¹, Renske Gahrman⁹¹, Joost W. Schouten⁹², Hendrikus J. Dubbink⁹⁴, Arnaud J. P. E. Vincent⁹², Martin J. van den Bent⁹³, Pim J. French⁹³, Stefan Klein⁹⁵, Yading Yuan⁹⁶, Sonam Sharma⁹⁶, Tzu-Chi Tseng⁹⁶, Saba Adabi⁹⁶, Simone P. Niclou⁹⁷, Olivier Keunen⁹⁸, Ann-Christin Hau^{97,99}, Martin Vallières^{100,101}, David Fortin^{101,102}, Martin Lepage^{101,103}, Bennett Landman¹⁰⁴, Karthik Ramadas¹⁰⁴, Kaiwen Xu¹⁰⁵, Silky Chotali¹⁰⁶, Lola B. Chambless¹⁰⁶, Akshikumar Mistry¹⁰⁶, Reid C. Thompson¹⁰⁶, Yuriy Gusev¹⁰⁷, Krithika Bhuvaneshwar¹⁰⁷, Anousheh Sayah¹⁰⁸, Camelia Bencheqroun¹⁰⁷, Anas Belouali¹⁰⁷, Subha Madhavan¹⁰⁷, Thomas C. Booth^{109,110}, Alysha Chelliah¹⁰⁹, Marc Modat¹⁰⁹, Haris Shuaib^{111,112}, Carmen Dragos¹¹¹, Aly Abayazeed¹¹³, Kenneth Kolodziej¹¹³, Michael Hill¹¹³, Ahmed Abbassy¹¹⁴, Shady Gamal¹¹⁴, Mahmoud Mekhaimar¹¹⁴, Mohamed Qayati¹¹⁴, Mauricio Reyes¹¹⁵, Ji Eun Park¹¹⁶, Jihye Yun¹¹⁶, Ho Sung Kim¹¹⁶, Abhishek Mahajan¹¹⁷, Mark Muzi¹¹⁸, Sean Benson¹¹⁹, Regina G. H. Beets-Tan^{120,121}, Jonas Teuwen¹¹⁹, Alejandro Herrera-Trujillo^{122,123}, Maria Trujillo¹²³, William Escobar^{122,123}, Ana Abello¹²³, Jose Bernal^{123,124}, Jhon Gómez¹²³, Joseph Choi¹²⁵, Stephen Baek¹²⁶, Yusung Kim¹²⁷, Heba Ismael¹²⁷, Bryan Allen¹²⁷, John M. Buatti¹²⁷, Aikaterini Kotrotsou¹²⁸, Hongwei Li¹²⁹, Tobias Weiss¹³⁰, Michael Weller¹³⁰, Andrea Bink¹³¹, Bertrand Pouymayou¹³¹, Hassan F. Shaykh¹³², Joel Saltz¹³³, Prateek Prasanna¹³³, Sampurna Shrestha¹³³, Kartik M. Mani^{133,134}, David Payne¹³⁵, Tahsin Kurc^{133,136}, Enrique Pelaez¹³⁷, Heydy Franco-Maldonado¹³⁸, Francis Loayza¹³⁷, Sebastian Quevedo¹³⁹, Pamela Guevara¹⁴⁰, Esteban Torche¹⁴⁰, Cristobal Mendoza¹⁴⁰, Franco Vera¹⁴⁰, Elvis Rios¹⁴⁰, Eduardo López¹⁴⁰, Sergio A. Velastin¹⁴¹, Godwin Ogbale¹⁴², Mayowa Soneye¹⁴², Dotun Oyekunle¹⁴², Olubunmi Odafe-Oyibo¹⁴³, Babatunde Osobu¹⁴², Mustapha Shu'aibu¹⁴⁴, Adeleye Dorcas¹⁴⁵, Farouk Dako^{2,146}, Amber L. Simpson^{112,147}, Mohammad Hamghalam^{147,148}, Jacob J. Peoples¹⁴⁷, Ricky Hu¹⁴⁷, Anh Tran¹⁴⁷, Danielle Cutler¹⁴⁹, Fabio Y. Moraes¹⁵⁰, Michael A. Boss¹⁵¹, James Gimpel¹⁵¹, Deepak Kattil Veetil¹⁵¹, Kendall Schmidt¹⁵², Brian Bialecki¹⁵², Sailaja Marella¹⁵¹, Cynthia Price¹⁵¹, Lisa Cimino¹⁵¹, Charles Apgar¹⁵¹, Prashant Shah⁵, Bjoern Menze^{4,129}, Jill S. Barnholtz-Sloan^{69,153}, Jason Martin⁵ & Spyridon Bakas^{1,2,3}✉

¹Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, PA, USA. ²Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁴Department of Informatics, Technical University of Munich, Munich, Bavaria, Germany. ⁵Intel Corporation, Santa Clara, CA, USA. ⁶Department of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany. ⁷Clinical Cooperation Unit Neuropathology, German Cancer Consortium (DKTK) within the German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸Department of Neuropathology, Heidelberg University Hospital, Heidelberg, Germany. ⁹Division of Medical Image Computing, German Cancer Research Center, Heidelberg, Germany. ¹⁰Pattern Analysis and Learning Group, Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany. ¹¹Neurology Clinic, Heidelberg University Hospital, Heidelberg, Germany. ¹²Department of Radiology & Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA. ¹³Symbiosis Center for Medical Image Analysis, Symbiosis International University, Pune, Maharashtra, India. ¹⁴Department of Neuroimaging and Interventional Radiology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India. ¹⁵Department of Radiology, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA. ¹⁶Department of Medical Physics, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA. ¹⁷Leeds Teaching Hospitals Trust, Department of Radiology, Leeds, UK. ¹⁸Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ¹⁹Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA. ²⁰Catalan Institute of Oncology, Badalona, Spain. ²¹Consorci MAR Parc de Salut de Barcelona, Catalonia, Spain. ²²Department of Radiology (IDI), Girona Biomedical Research Institute (IdIBGi), Josep Trueta University Hospital, Girona, Spain. ²³Institute of Neuroradiology, Neuromed Campus (NMC), Kepler University Hospital Linz, Linz, Austria. ²⁴Department of Neurooncology, Neuromed Campus (NMC), Kepler University Hospital Linz, Linz, Austria. ²⁵Institute of Diagnostic and Interventional Neuroradiology, RKH Klinikum Ludwigsburg, Ludwigsburg, Germany. ²⁶Department of Radiation Oncology, Christiana Care Health System, Philadelphia, PA, USA. ²⁷Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA. ²⁸Department of Radiation Oncology, University of Maryland, Baltimore, MD, USA. ²⁹Department of Radiation Oncology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA. ³⁰Department of Radiation Oncology, The James Cancer Hospital and Solove Research Institute, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. ³¹Department of Radiology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA. ³²The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ³³The Malone Center for Engineering in Healthcare, The Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA. ³⁴Department of Electrical and Computer Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA. ³⁵The Chinese University of Hong Kong, Hong Kong, China. ³⁶Centre for Biomedical Image Analysis, Faculty of Informatics, Masaryk University, Brno, Czech Republic. ³⁷Department of Radiology and Nuclear Medicine, Faculty of Medicine, Masaryk University, Brno and University Hospital Brno, Brno, Czech Republic. ³⁸Department of Biophysics, Faculty of Medicine, Masaryk University, Brno, Czech Republic. ³⁹Department of Neurosurgery, Faculty of Medicine, Masaryk University, Brno, and University Hospital and Czech Republic, Brno, Czech Republic. ⁴⁰Department of Neuro Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. ⁴¹University of Alberta, Edmonton, AB, Canada. ⁴²Alberta Machine Intelligence Institute, Edmonton, AB, Canada. ⁴³Department of Radiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. ⁴⁴University of Texas Southwestern Medical Center, Dallas, TX, USA. ⁴⁵UCLA Brain Tumor Imaging Laboratory (BTIL), Center for Computer Vision and Imaging Biomarkers, Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ⁴⁶UCLA Neuro-Oncology Program, Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ⁴⁷Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ⁴⁸College of Medicine and Public Health, Flinders University, Bedford Park, SA, Australia. ⁴⁹Division of Surgery and Perioperative Medicine, Flinders Medical Centre, Bedford Park, SA, Australia. ⁵⁰South Australia Medical Imaging, Flinders Medical Centre, Bedford Park, SA, Australia. ⁵¹Department of Neurology, Baylor College of Medicine, Houston, TX, USA. ⁵²Institute of Computing, University of Campinas, Campinas, São Paulo, Brazil. ⁵³Federal Institute of São Paulo, Campinas, São Paulo, Brazil. ⁵⁴Instituto de Neurologia de Curitiba, Curitiba, Paraná, Brazil. ⁵⁵Department of Radiology, Hospital de Clínicas da Universidade Federal do Paraná, Curitiba, Paraná, Brazil. ⁵⁶Department of Informatics, Universidade Federal do Paraná, Curitiba, Paraná, Brazil. ⁵⁷Department of Radiology, Washington University in St. Louis, St. Louis, MO, USA. ⁵⁸Department of Diagnostic and Interventional Neuroradiology, School of Medicine, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany. ⁵⁹TranslaTUM (Zentralinstitut für translationale Krebsforschung der Technischen Universität München), Klinikum rechts der Isar, Munich, Germany. ⁶⁰Image-Based Biomedical Modeling, Department of Informatics, Technical University of Munich, Munich, Germany. ⁶¹Department of Radiology, NYU Grossman School of Medicine, New York, NY, USA. ⁶²Department of Neurosurgery, NYU Grossman School of Medicine, New York, NY, USA. ⁶³Support Center for Advanced Neuroimaging, University Institute of Diagnostic and Interventional Neuroradiology, University Hospital Bern, Inselspital, University of Bern, Bern, Switzerland. ⁶⁴Department of Neurosurgery, Anschutz Medical Campus, University of Colorado, Aurora, CO, USA. ⁶⁵Department of Radiology, University Hospitals Cleveland, Cleveland, OH, USA. ⁶⁶Department of Neurological Surgery, University Hospitals-Seidman Cancer Center, Cleveland, OH, USA. ⁶⁷Case Comprehensive Cancer Center, Cleveland, OH, USA. ⁶⁸Department of Neurosurgery, Case Western Reserve University School of Medicine, Cleveland, OH, USA. ⁶⁹National Cancer Institute, National Institute of Health, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA. ⁷⁰Department of Radiology, Neuroradiology Division, University of Pittsburgh, Pittsburgh, PA, USA. ⁷¹Department of Diagnostic Radiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁷²University of Pittsburgh Medical Center, Pittsburgh, PA, USA. ⁷³Department of Neuroradiology, University of Michigan, Ann Arbor, MI, USA. ⁷⁴Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. ⁷⁵Department of Radiology, Baylor College of Medicine, Houston, TX, USA. ⁷⁶Department of Computing, Imperial College London, London, UK. ⁷⁷Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK. ⁷⁸Department of Radiology, Imperial College NHS Healthcare Trust, London, UK. ⁷⁹Computational Oncology Group, Institute for Global Health Innovation, Imperial College London, London, UK. ⁸⁰Department of NeuroRadiology, University of Patras, Patras, Greece. ⁸¹Department of Neurosurgery, University of Patras, Patras, Greece. ⁸²Clinical Radiology Laboratory, Department of Medicine, University of Patras, Patras, Greece. ⁸³Department of Electrical and Computer Engineering, University of Patras, Patras, Greece. ⁸⁴Department of Neuro-Oncology, University of Patras, Patras, Greece. ⁸⁵Department of Radiation Oncology, University of Patras, Patras, Greece. ⁸⁶Yonsei University College of Medicine, Seoul, Korea. ⁸⁷Department of Radiation Oncology, Henry Ford Health System, Detroit, MI, USA. ⁸⁸Public Health Sciences, Henry Ford Health System, Detroit, MI, USA. ⁸⁹SJTU-Ruijin-UIH Institute for Medical Imaging Technology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, 200025 Shanghai, China. ⁹⁰Case Western Reserve University, Cleveland, OH, USA. ⁹¹Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Centre Rotterdam, Rotterdam, Netherlands. ⁹²Department of Neurosurgery, Brain Tumor Center, Erasmus MC University Medical Centre Rotterdam, Rotterdam, Netherlands. ⁹³Department of Neurology, Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, Netherlands. ⁹⁴Department of Pathology, Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, Netherlands. ⁹⁵Biomedical Imaging Group Rotterdam, Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Centre Rotterdam, Rotterdam, Netherlands. ⁹⁶Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁹⁷NORLUX Neuro-

Oncology Laboratory, Department of Cancer Research, Luxembourg Institute of Health, Luxembourg, Luxembourg. ⁹⁸Translation Radiomics, Department of Cancer Research, Luxembourg Institute of Health, Luxembourg, Luxembourg. ⁹⁹Luxembourg Center of Neuropathology, Laboratoire National De Santé, Luxembourg, Luxembourg. ¹⁰⁰Department of Computer Science, Université de Sherbrooke, Sherbrooke, QC, Canada. ¹⁰¹Centre de Recherche du Centre Hospitalière Universitaire de Sherbrooke, Sherbrooke, QC, Canada. ¹⁰²Division of Neurosurgery and Neuro-Oncology, Faculty of Medicine and Health Science, Université de Sherbrooke, Sherbrooke, QC, Canada. ¹⁰³Department of Nuclear Medicine and Radiobiology, Sherbrooke Molecular Imaging Centre, Université de Sherbrooke, Sherbrooke, QC, Canada. ¹⁰⁴Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA. ¹⁰⁵Department of Computer Science, Vanderbilt University, Nashville, TN, USA. ¹⁰⁶Department of Neurosurgery, Vanderbilt University Medical Center, Nashville, TN, USA. ¹⁰⁷Innovation Center for Biomedical Informatics (ICBI), Georgetown University, Washington, DC, USA. ¹⁰⁸Division of Neuroradiology & Neurointerventional Radiology, Department of Radiology, MedStar Georgetown University Hospital, Washington, DC, USA. ¹⁰⁹School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK. ¹¹⁰Department of Neuroradiology, Ruskin Wing, King's College Hospital NHS Foundation Trust, London, UK. ¹¹¹Stoke Mandeville Hospital, Mandeville Road, Aylesbury, UK. ¹¹²Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada. ¹¹³Neosoma Inc., Groton, MA, USA. ¹¹⁴University of Cairo School of Medicine, Giza, Egypt. ¹¹⁵University of Bern, Bern, Switzerland. ¹¹⁶Department of Radiology, Asan Medical Center, Seoul, South Korea. ¹¹⁷The Clatterbridge Cancer Centre NHS Foundation Trust Pembroke Place, Liverpool, UK. ¹¹⁸Department of Radiology, University of Washington, Seattle, WA, USA. ¹¹⁹Netherlands Cancer Institute, Amsterdam, Netherlands. ¹²⁰Department of Radiology, Netherlands Cancer Institute, Amsterdam, Netherlands. ¹²¹GROW School of Oncology and Developmental Biology, Maastricht, Netherlands. ¹²²Clínica Imbanaco Grupo Quirón Salud, Cali, Colombia. ¹²³Universidad del Valle, Cali, Colombia. ¹²⁴The University of Edinburgh, Edinburgh, UK. ¹²⁵Department of Industrial and Systems Engineering, University of Iowa, Iowa, USA. ¹²⁶Department of Industrial and Systems Engineering, Department of Radiation Oncology, University of Iowa, Iowa City, IA, USA. ¹²⁷Department of Radiation Oncology, University of Iowa, Iowa City, IA, USA. ¹²⁸MD Anderson Cancer Center, University of Texas, Houston, TX, USA. ¹²⁹Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland. ¹³⁰Department of Neurology, Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich, Switzerland. ¹³¹Department of Neuroradiology, Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich, Switzerland. ¹³²University of Alabama in Birmingham, Birmingham, AL, USA. ¹³³Department of Biomedical Informatics, Stony Brook University, Stony Brook, New York, USA. ¹³⁴Department of Radiation Oncology, Stony Brook University, Stony Brook, NY, USA. ¹³⁵Department of Radiology, Stony Brook University, Stony Brook, NY, USA. ¹³⁶Scientific Data Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA. ¹³⁷Escuela Superior Politécnica del Litoral, Guayaquil, Guayas, Ecuador. ¹³⁸Sociedad de Lucha Contra el Cáncer - SOLCA, Guayaquil Ecuador, Guayaquil, Ecuador. ¹³⁹Universidad Católica de Cuenca, Cuenca, Ecuador. ¹⁴⁰Universidad de Concepción, Concepción, Biobío, Chile. ¹⁴¹School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK. ¹⁴²Department of Radiology, University College Hospital Ibadan, Oyo, Nigeria. ¹⁴³Clinix Healthcare, Lagos, Lagos, Nigeria. ¹⁴⁴Department of Radiology, Muhammad Abdullahi Wase Teaching Hospital, Kano, Nigeria. ¹⁴⁵Department of Radiology, Obafemi Awolowo University Ile-Ife, Ile-Ife, Osun, Nigeria. ¹⁴⁶Center for Global Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ¹⁴⁷School of Computing, Queen's University, Kingston, ON, Canada. ¹⁴⁸Department of Electrical Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran. ¹⁴⁹The Faculty of Arts & Sciences, Queen's University, Kingston, ON, Canada. ¹⁵⁰Department of Oncology, Queen's University, Kingston, ON, Canada. ¹⁵¹Center for Research and Innovation, American College of Radiology, Philadelphia, PA, USA. ¹⁵²Data Science Institute, American College of Radiology, Reston, VA, USA. ¹⁵³Center for Biomedical Informatics and Information Technology, National Cancer Institute (NCI), National Institute of Health, Bethesda, MD, USA. ¹⁵⁴These authors contributed equally: Sarthak Pati, Ujjwal Baid, Brandon Edwards.

✉ e-mail: sbakas@upenn.edu

ANEXO III – APRESENTAÇÃO DOS RESULTADOS PARCIAIS

ECR 2022: Notification of Abstract Decision

1 mensagem

ECR Scientific Programme Team <abstracts@myesr.org>

15 de dezembro de 2021 às 10:48

Responder a: abstracts@myesr.org

Para: flaviasprenger@gmail.com

Dear Mrs. Sprenger,

Thank you very much for submitting your abstract(s) for ECR 2022 and for your support during these difficult times!

As you might have already heard, it has been decided to hold two ECR events next year - ECR 2022 and the ECR 2022 Overture: <https://www.myesr.org/congress/announcement>

The Programme Planning Committee has carefully reviewed all submissions, and the accepted abstracts have been allocated to the most suitable presentation format.

We would like to wholeheartedly congratulate you on the acceptance of your abstract **#15065** "Semi-automatic segmentation and prognostic features in high-grade gliomas", which has been accepted as **EPOS Radiologist (scientific)**.

You can also review the results in your ESR account. To do so, simply log in to your [MyUserArea > Abstract Submission](#) with your **e-mail** and **password**.

Further steps for oral presentations

Next week, you will receive a separate e-mail with further details and instructions for the poster upload.

Invest in the Youth

We are happy to inform you that your application for Invest in the Youth programme was **accepted**. You are entitled to **free registration** for ECR 2022 and receive a travel support up to 100€ for one night.

To confirm your free registration for ECR 2022, simply go to:

[ESR MyUserArea > Registration](#) and log in with your e-mail and password. If you need assistance regarding the log-in or your registration, please contact registration@myesr.org. Please note that there will be no refund for fees already paid.

The accommodation needs to be booked on the following platform: [ZENhotels](#)

A booking made on a different platform is not eligible for reimbursement. The reimbursement form will be sent to you after your active participation at ECR 2022 in July.

Should you have any questions regarding Invest in the Youth please contact invest@myesr.org.

If you have any further questions, please feel free to contact us at: abstracts@myesr.org

With kind regards,
ECR Scientific Programme Team

--

ANEXO IV – CONFIRMAÇÃO DE SUBMISSÃO

WNS-22-5661 - Confirming your submission to World Neurosurgery

1 mensagem

World Neurosurgery <em@editorialmanager.com>
 Responder a: World Neurosurgery <support@elsevier.com>
 Para: Flavia Sprenger <flaviasprenger@gmail.com>

28 de dezembro de 2022 às 10:45



Dear Dr. Flavia Sprenger,

Your submission entitled "Ki67 Index Correlates with Tumoral Volumetry, and Residual 5-ALA Tumor Fluorescence in Glioblastoma" has been received by World Neurosurgery. It has been assigned the following manuscript number: WNS-22-5661.

You may check on the progress of your paper by logging in, as an author, to the Editorial Manager. The URL is <https://www.editorialmanager.com/worldneurosurgery/>.

Your username is: flaviasprenger@gmail.com
 If you need to retrieve password details,
 please go to: [click here to reset your password](#)

You'll receive your official manuscript number by email shortly. Please use this manuscript number in all future correspondence regarding this manuscript.

Sincerely,

Edward C. Benzel, M.D.
 Editor-in-Chief
 World Neurosurgery

This journal uses the Elsevier Article Transfer Service. This means that if an editor feels your manuscript is more suitable for an alternative journal, then you might be asked to consider transferring the manuscript to such a journal. The recommendation might be provided by a Journal Editor, a dedicated Scientific Managing Editor, a tool assisted recommendation, or a combination. For more details see the journal guide for authors.

#AU_WORLDNEUROSURGERY#

To ensure this email reaches the intended recipient, please do not delete the above code



ANEXO V – ARTIGO SUBMETIDO (TEXTO NA ÍNTEGRA)

Title: Ki67 Index Correlates with Tumoral Volumetry, and Residual 5-ALA Tumor Fluorescence in Glioblastoma

ABSTRACT

Background: Malignant gliomas are the most prevalent primary cerebral tumor. Preoperative imaging plays an important role, and the prognosis is closely related to surgical resection and histomolecular aspects. Our goal was to correlate Ki67 indexes with tumoral volumetry in semiautomatic segmentation on preoperative magnetic resonance images in a 5-ALA assisted resection cohort.

Methods: We included initially seventy-eight IDH-wildtype glioblastoma patients with complete preoperative imaging submitted to 5-ALA assisted resections. Clinical, surgical, and histomolecular findings were also obtained. Preoperative magnetic resonance studies were preprocessed and segmented semiautomatically on VISVA for whole tumor (WT) on 3D FLAIR, enhancing tumor (ET), and necrotic core (NC) on 3D post-gadolinium T1. Edema volumes (ED) were obtained from the subtraction of ET and NC from WT. According to previous studies, we used a 20% cutoff for Ki67 for statistical analysis.

Results: Final sample was of 48 patients. Higher Ki-67 indexes correlated positively with higher WT ($p = 0,01$), ET ($p = 0,03$) and ED volumes ($p = 0,04$), with no correlation with NC volumes ($p = 0,18$). Ki67 indexes were also higher in 5-ALA compromised margins ($p = 0,01$).

Conclusions: There is a correlation between Ki67, the metabolically active tumoral volumes, and 5-ALA compromised margins. Tumoral volumetry aspects (WT, ET, and ED), except for NC, correlated directly with Ki67 indexes. Methodological inconsistencies are probably responsible for contradictory literature findings, and further studies are needed to validate and reproduce these findings.

Keywords: Ki67, segmentation, volumetry, glioblastoma, 5-ALA

ARTICLE TEXT

Introduction

Malignant gliomas are the most prevalent primary cerebral tumor, with an estimated incidence of two to four cases per 100.000 people. Despite recent advances and a relentless search for targeted therapies, high-grade gliomas still have a dismal prognosis, with a mean survival of 35% after one year (BONDY et al., 2008; OMURO; DEANGELIS, 2013; THAKKAR et al., 2014; TONN; WESTPHAL; RUTKA, 2010).

Imaging plays a crucial role in glioblastoma diagnosis and management. MR features have a predictive value, but the most crucial prognosis-changing factor is still a maximal surgical resection with functional preservation (HANIF et al., 2017; OMURO; DEANGELIS, 2013; THAKKAR et al., 2014). 5-ALA fluorescence-guided surgery has been a feasible intraoperative tool to achieve total gross removal of high-grade gliomas, facilitating the posterior adjuvant treatment consisting of irradiation plus temozolomide (KLINIK et al., 2006a; STUMMER et al., 2011).

Prognosis is also closely related to clinical aspects like KPS scores, age and histomolecular features, like Ki67 (HANIF et al., 2017; OMURO; DEANGELIS, 2013; THAKKAR et al., 2014).

Many studies aimed to correlate tumoral volumetry with survival rates and histomolecular findings (ALBRIGHT RE; FRAM EK, 1988a; ANDREOU et al., 1983a; GRABOWSKI et al., 2014; HAMMOUD et al., 1996a; ILIADIS et al., 2012; LACROIX et al., 2001; LEVIN et al., 1980a; WANGARYATTAWANICH et al., 2015; WOOD; GREEN; SHAPIRO, 1988b; XUE; ALBRIGHT, 1999a; ZINN et al., 2011). However, given its well-established relation to cellular proliferation and the scarce and controversial exploit of this theme, this study's goal is to correlate Ki67 indexes with tumoral volumetry in semiautomatic segmentation on preoperative MR studies, as well as clinical, surgical, and histopathological aspects in a 5-ALA assisted resection cohort.

Methods

Study design

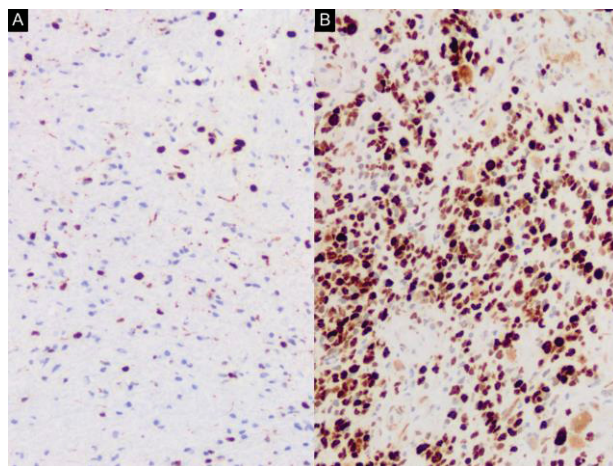
This study was approved by the institutional ethical committee. In this observational retrospective study, we initially included 78 patients with new IDH-wildtype glioblastoma who underwent surgical manipulation (resection and biopsy) in a tertiary neurosurgery facility, between 2015 and 2021, with available preoperative MR images including T1 and T2-weighted sequences, FLAIR and 3D post-gadolinium T1. Preoperative scans were obtained the earliest at 25 days before surgery.

Patients with unavailable preoperative MR images or inadequate protocol, artifacts that harmed adequate analysis, previous neurosurgery, unavailable medical charts, or histopathological reports were excluded.

Clinical and histopathological data, such as 5-ALA margins status, intraoperative and definite histological diagnosis, IDH status, MGMT, ATRX, TP53 mutations, and Ki67 indexes were collected.

MIB-I, a mouse-derived anti-Ki67 antibody, was used for Ki67 immunohistochemistry. Slices were analyzed on a light microscope. The labeling index consisted of the percentage of immunoreactive cells among all cells, with a pool of at least 1000 examined cells, in tumoral areas with the highest immunostaining activity (Figure 1).

Figure 1 – Low and high Ki67 indexes examples

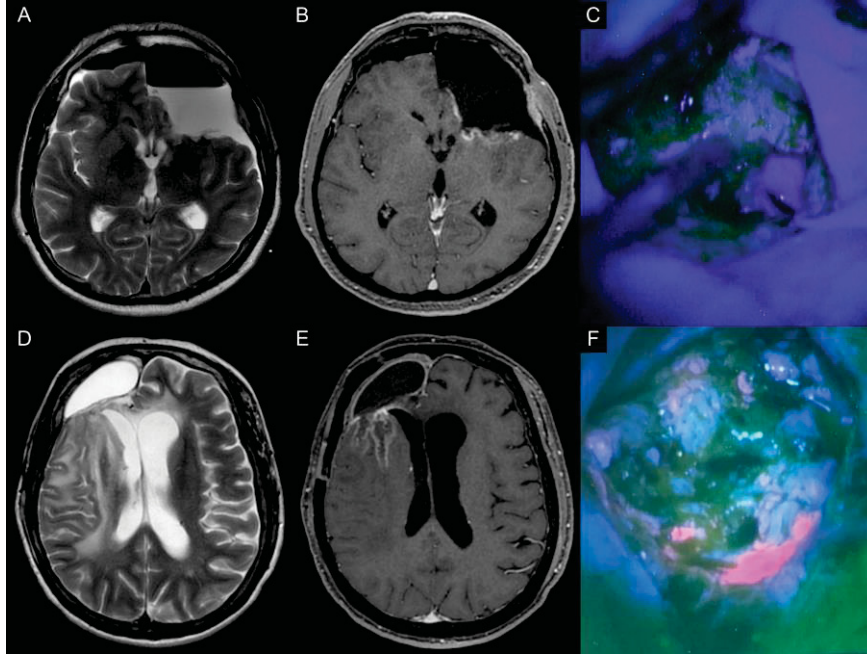


Legend: A: Ki67 labelling of approximately 5-10% of the tumor cells (Ki67 immunohistochemistry, 200x). B: Ki67 labelling of approximately 80% of the tumor cells (Ki67 immunohistochemistry, 200x).

The same neurosurgery team performed the surgical approach, and 5-ALA was administered orally on a 20mg/kg dose diluted in 50 mL of water three hours before anesthesia. 5-ALA free margins were considered in the absence of residual intraoperative visible fluorescence. Complete resection was considered in the absence of nodular

enhancements on immediate postoperative MRI (Figure 2). All patients underwent the same adjuvant treatment following the Stupp protocol (STUPP et al., 2005).

Figure 2 – Examples of surgical margins in immediate postoperative MRI and 5-ALA fluorescence



Legend: A-B: Immediate postoperative MRI with an axial T2 (A) and post-gadolinium T1 (B) showing a left frontal lobe resection, with linear enhancement around the surgical cavity on B. C: Intraoperative image shows no residual 5-ALA fluorescence. This was considered a complete resection with 5-ALA free margins. D-E: Immediate postoperative MRI with an axial T2 (D) and post-gadolinium T1 (E) showing a right frontal lobe surgical cavity, with residual irregular and solid enhancement on the posterior margin. F: Intraoperative image shows residual red 5-ALA fluorescence in pink. This was considered an incomplete resection with 5-ALA compromised margins.

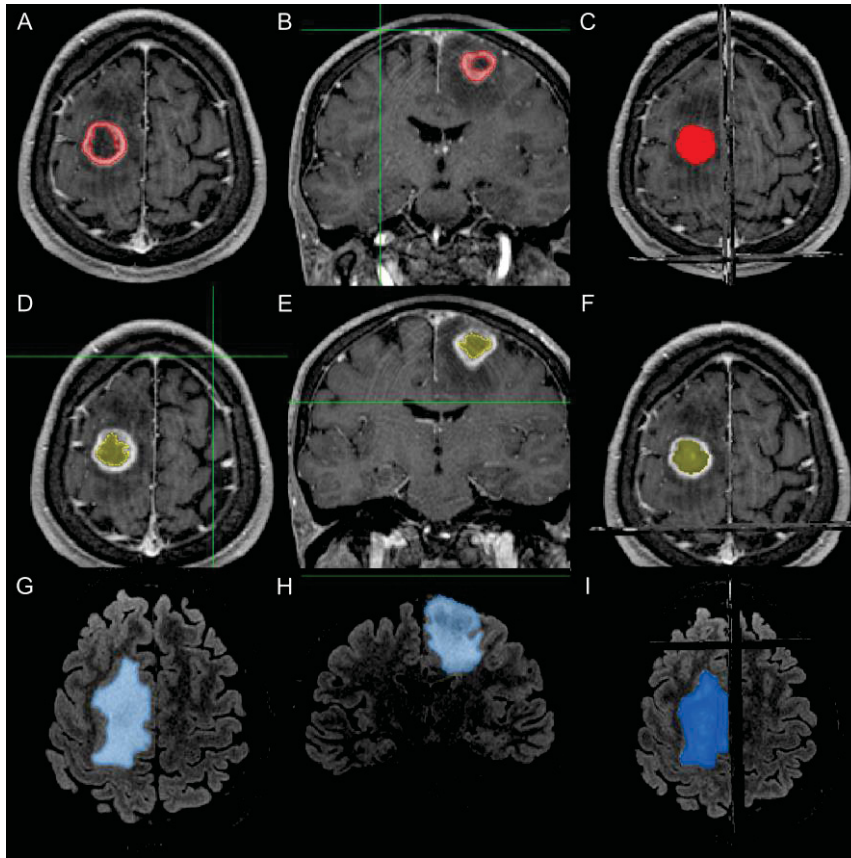
Images preprocessing and segmentation

Preoperative MR images were acquired from General Electric (GE) Healthcare Signa HDxt 1.5T and Discovery MR750W 3.0T scans with a 28-channel head and neck coil. Images were downloaded in DICOM format from the Advantage Workstation 4.7 (GE Medical System, Milwaukee, WI).

DICOM images were interpolated to isometric conversion to 1 x 1 x 1 mm voxels and converted to NIFTI format. They were registered on the same coordinate axis and processed for noise removal, inhomogeneity correction, and histogram matching.

Preprocessed images segmented semiautomatically on Volumetric Image Segmentation for Visualization and Analysis (VISVA), developed by the Laboratory of Image Data Science of the University of Campinas, with a previously described technique (FALCÃO; BERGO, 2004) for whole tumor (WT) on 3D FLAIR and enhancing tumor (ET) and necrotic core (NC) on 3D post-gadolinium T1 (Figure 3). Edema (ED) volumes were obtained from the subtraction of ET and NC from WT (Figure 3). Segmentations were performed by a third-year radiology resident (FS) and confirmed by a neuroradiologist (BCAT) with seven years of experience. Volumes from all generated labels were exported on an electronic spreadsheet.

Figure 3 – Semiautomatic segmentation on VISVA



Legend: Example of semiautomatic segmentation on VISVA. A-C: Segmentation for enhancing tumor (ET, red label) on 3D T1Gd+ axial and coronal planes, and volumetric rendering, respectively. D-F: Segmentation for necrotic core (NC, yellow label) on 3D T1Gd+ axial and coronal planes, and volumetric rendering, respectively. Segmentation for whole tumor (WT, blue label) on 3D FLAIR on axial and coronal planes, and volumetric rendering, respectively.

Statistical analysis

Statistical analysis was performed using a Microsoft Excel spreadsheet and SPSS20 software (IBM Corp., New York, USA), establishing a significance level of 0.05 (5%). The normality of the distribution of variables was tested using the Kolmogorov-Smirnov test, allowing the use of parametric tests. Student T-tests were used to analyze continuous quantitative variables, chi-square, and Fisher's test for comparing two groups with discrete variables. Pearson's correlation was used for the relationship between continuous quantitative variables. According to previous studies, we used a 20% cutoff for Ki67 for statistical analysis (ARMOCIDA et al., 2020; HENKER et al., 2019). A 30% cutoff for Ki67 was also tested (sample's median).

Results

Clinical, surgical, and segmentation aspects

Thirty patients were excluded. Ki67 indexes were unavailable for 22 patients, and eight patients did not have preoperative images. The final sample included 48 patients, which were all semiautomatically segmented.

The sample was predominantly composed of males ($n = 29$, 60%), with a mean age of 56,92 years old, ranging from 35 and 81 years old. Demographic and histomolecular findings are summarized in table 1 (Table 1).

Table 1 – Demographic findings summary

	Ki67 < 20% (n = 11)	Ki67 ≥ 20% (n = 37)	p-value
Sex	Male N= 6	Male N = 23	0,61
	Female N = 5	Female N =14	
Age (Years)	56,64 ± 8,49	59,19 ± 11,40	0,50
KPS	80 ± 0	74 ± 12,31	0,25
MGMT (n = 29)	Negative N = 3	Negative N = 9	0,36
	Positive N = 2	Positive N = 15	
ATRX (n = 40)	Negative N = 0	Negative N = 2	0,10
	Positive N = 11	Positive N = 27	
TP53 (n = 39)	Negative N = 1	Negative N = 1	0,57
	Positive N = 10	Positive N = 27	
Lobe	Frontal N = 6	Frontal N = 17	
	Parietal N = 3	Parietal N = 9	
	Temporal N = 1	Temporal N = 7	
	Occipital N = 0	Occipital N = 4	
	Cerebellum N = 1	Cerebellum N = 0	
Side	Left N = 3	Left N = 20	
	Right N = 8	Right N = 14	
	Bilateral N = 0	Bilateral N = 3	
5-ALA margins	Free margins N = 9	Free margins N = 17	0,01
	Compromised margins N = 2	Compromised margins N = 20	
WT (mm ³)	Minimum = 7003	Minimum = 18624	0,01
	Maximum = 101035	Maximum = 241853	
	Mean = 38684 ± 34004	Mean = 113547 ± 65575	
ET (mm ³)	Minimum = 0	Minimum = 0	0,03
	Maximum = 15772	Maximum = 80419	
	Mean = 7157 ± 5729	Mean = 28870 ± 26756	
NC (mm ³)	Minimum = 0	Minimum = 0	0,18
	Maximum = 3094	Maximum = 73487	
	Mean = 1120 ± 1279	Mean = 12764 ± 20209	
ED (mm ³)	Minimum = 0	Minimum = 6898	0,04
	Maximum = 92086	Maximum = 176300	
	Mean = 32383 ± 32526	Mean = 71912 ± 44994	

Legend: KPS = Karnofsky Performance Score; WT = whole tumor; ET = enhancing tumor; NC = necrotic core; ED = edema.

Mean initial KPS was of 75,89, with no significant difference according to 5-ALA margins ($p = 0,18$) or Ki67 indexes ($p = 0,25$). Volumes of WT ($p = 0,12$), ET ($p = 0,37$), NC ($p = 0,49$) or ED ($p = 0,15$) did not correlate with KPS scores.

Eighteen patients (38%) were submitted to complete resection, 25 patients underwent incomplete resections (52%), and the remaining five patients were submitted to surgical biopsy. Volumetry did not correlate with type of

resection. 5-ALA margins were free of fluorescence on 26 patients (54%) and compromised in 22 (46%). WT ($p = 0,61$), ET ($p = 0,99$), NC ($p = 0,79$) and ED ($p = 0,82$) volumes did not correlate with 5-ALA margins. There was a positive and significant correlation between WT's volume and ET's ($p < 0,001$), NC's ($p < 0,001$) and ED's volumes ($p < 0,001$). Therefore, higher volumes of one of the tumoral components were linked to higher volumes of the remaining components.

Histopathological aspects

Ki-67 index had a median of 30%, with a minimum of 5% and a maximum of 80%, showing no correlation with age ($p = 0,50$) or KPS scores ($p = 0,25$).

Higher Ki-67 indexes correlated positively with higher WT ($p = 0,01$), ET ($p = 0,03$) and ED volumes ($p = 0,04$), with no correlation with NC volumes ($p = 0,18$). Ki67 indexes were also higher in 5-ALA compromised margins, with a mean of 43% *versus* 30% in 5-ALA free margins group ($p = 0,01$).

A 30% Ki67 cutoff was also tested, with no correlation to age ($p = 0,56$), sex ($p = 0,61$), KPS ($p = 0,52$) or NC ($p = 0,11$), but correlations with WT ($p = 0,01$), WT ($p = 0,03$) and ED ($p = 0,05$) were still present.

MGMT, ATRX, or TP53 mutations did not correlate with Ki67, 5-ALA margins, or tumoral volumetry.

Discussion

In the genetic era of glioma investigation, and considering the dismal prognosis of glioblastoma patients, our goal was to revisit the role of Ki67, a highly accessible tool, and its correlation with semiautomatic segmentation, clinical, surgical, and histopathological aspects in a 5-ALA fluorescence assisted resections cohort. We found a strong correlation between Ki67 indexes, tumoral volumetry, and 5-ALA margins.

Ki67 received its name after its first description by Gerdes *et al.* in the city of Kiel in Germany, and its 67 copies were found to be related to cell proliferation in mice with Hodgkin's lymphoma. It consists of a monoclonal IgG antibody that operates as a marker of cell proliferation. In the context of astrocytomas, its fraction is related to increasing grade of malignancy and more severe disease (GERDES *et al.*, 1983; HU *et al.*, 2013; LINN JOHANNESSEN; TORP, 2006).

In our sample, Ki67 indexes did not correlate with age or initial KPS scores, in consonance with similar studies' previous findings (ARMOCIDA *et al.*, 2020; HENKER *et al.*, 2017).

Glioma's most important prognosis-changing factor remains surgery for maximal cytoreduction and functional preservation, but gross resection and surgical identification of tumoral margins can be challenging. Therefore, the introduction of 5-ALA has revolutionized and improved gross total resection rates (HANIF *et al.*, 2017; OMURO; DEANGELIS, 2013; SENDERS *et al.*, 2017; STUMMER *et al.*, 1998a, 1998b; THAKKAR *et al.*, 2014).

5-ALA (5-aminolevulinic acid) is a natural metabolite of the hemoglobin pathway that can accumulate in increased vascular permeability tissues, like gliomas and their surroundings. Usually, it is converted to protoporphyrin IX. However, the decreased expression of the ferrochelatase enzyme in neoplastic cells promotes its accumulation, which can be detected after blue-light excitation, enhancing the intraoperative guidance for tumor removal. The intensity of the fluorescence may also predict the degree of tumor cellularity. 5-ALA fluorescence-guided resections aid more complete excisions, improving survival rates compared to the conventional white light microscopic approach (COBURGER; WIRTZ, 2019; HADJIPANAYIS; STUMMER, 2019; HERVEY-JUMPER; BERGER, 2016; KLINIK *et al.*, 2006b, 2006a; MAHMOUDI *et al.*, 2019; PICART *et al.*, 2019).

Our sample was entirely submitted to 5-ALA fluorescence-guided resection, and 5-ALA compromised margins patients had significantly higher Ki67 indexes when compared with fluorescence-free margins. Ki67 index is the highest on the interface between the solid portion of the tumor and the surrounding tissue (EIDEL et al., 2017). This phenomenon may explain these findings since sampling for Ki67 was possibly performed close to the lesion's infiltrating and fluorescence positive margins.

Tumoral volumetry aspects (WT, ET, and ED), except for NC, correlated directly with Ki67 indexes. As expected, ET reflects the cellular and proliferating component of the tumor and NC the acellular portion of the lesion. However, literature is still scarce on Ki67 impact on preoperative quantitative MR features (Table 2).

Table 2 – Summary of literature findings on Ki67 and volumetric segmentation

Author	Year	N	Sample	Technique	Segmentation	Findings
Henker <i>et al</i>	2019	152	IDH wild-type GBM	Semiautomatic (contour expansion - SmartBrush by Brainlabs)	Edema on 2D FLAIR; Enhancing tumor and necrosis on 3D T1Gd+; subtracts cystic areas	No correlation between Ki67 and volumetry
Armocida <i>et al</i>	2020	127	IDH wild-type GBM	Manual tracer (Osirix)	Does not specify (area of altered signal on T1, not specified if 2D/3D)	Correlation between Ki67 and tumoral volume on > 45 cm3 lesions
Caramanti <i>et al</i>	2022	70	Does not specify IDH status	Manual (Horos)	Edema on FLAIR; enhancing tumor on T1Gd+ (does not specify 2D/3D)	Correlation between edema ($p < 0,004$) and Ki67. No correlation was found for enhancing tumor or whole tumor
Our study	2022	48	IDH wild-type GBM	Semiautomatic (VISVA by LIDS University of Campinas)	WT on 3D FLAIR, ET, and NC on 3D T1Gd+	Ki67 correlates with WT, ET, and ED volumes, and 5-ALA compromised margins

Legend: GBM – glioblastoma.

Henker *et al.* studied the solid enhancement, necrosis, and peritumoral edema volumes and found no correlation with Ki67 indexes. The author proposed that the lack of correlation between Ki67 and volumetry on his sample

was due to the imaging timing, probably performed on initial clinical manifestation, not thoroughly assessing its progression and multiplication over time (HENKER *et al.*, 2019). We have addressed this limitation in our study, and all preoperative imaging was performed within three weeks of surgical manipulation, reflecting, therefore, closely the surgical specimen characteristics.

A remarkable difference between Henker's study and ours is that the former subtracted cystic areas from the total tumoral volume and analysis. Pathophysiology of cyst formation in gliomas is still questionable, with most theories establishing a disruption of the blood-brain barrier and hyperproteic effusion. Cystic areas are homogeneous, non-enhancing, and smooth-walled areas, with signal intensity similar to CSF, contrasting with generally central, heterogeneous, and irregular marginated necrosis areas (CHOI *et al.*, 2010; CURTIN *et al.*, 2020; LOHLE *et al.*, 2000). Despite its relation to its higher frequency in lower-grade gliomas and better prognosis, no definite MR feature can differentiate unenhancing tumor cystic areas from tumor necrosis. Therefore, we opted to include these unenhancing cystic areas in the segmentation.

Caramanti *et al.*, on the other hand, found a positive correlation between Ki67 indexes and peritumoral edema on FLAIR, but curiously, not as expected on the solid enhancing component, hypothesized to have the highest cellularity (CARAMANTI *et al.*, 2022). This finding may suggest that in these cases, peritumoral signal abnormalities may be more related to microinvasion than vasogenic edema.

Armocida *et al.* also studied tumor volume and found a strong correlation between higher volume lesions and higher Ki67 indexes in more extensive lesions, with a cutoff of 45 cm³. However, the author does not specify which parameters or sequences were used for preoperative segmentation (ARMOCIDA *et al.*, 2020).

Despite similar and homogeneous samples, Ki67 indexes and volumetry are conflicting. Henker *et al.* was the only author to use a similar and reliable semiautomatic segmentation technique based on contour expansion, which considers minimal variations in pixel intensity and lesion contours.

Caramanti *et al.* and Armocida *et al.* used manual tracing segmentation, which does not promote a delicate lesion's contouring, and relies on automatically generating missing ROIs between segmented slices, resulting in inaccurate and polygonal segmentation. These authors also do not specify the type of sequences used for segmentation, questioning whether 2D sequences are adequate or reliable for volumetric segmentation.

We emphasize as a strength of our study that only volumetric sequences (3D FLAIR and post-gadolinium T1) were used and segmented with a validated and reliable method that utilizes region growth from a starting signalized seed. Therefore, we believe that we could find the expected pathophysiological result even though we had the smallest sample.

As proposed previously by Henker *et al.*, 3D volumetric assessment is the only reliable technique, suggesting that cuboid, ellipsoid, or spheroid techniques should not be used for these purposes. Inaccurate measuring tools are probably the main reason for inconsistent literature findings, and 3D segmentation proves to be more accurate and superior to other methods (HENKER *et al.*, 2017).

Despite entirely different biochemical mechanisms, 5-ALA fluorescence occurring secondary to blood-brain barrier permeability, ferrochelatase downregulation and cytosolic accumulation of protoporphyrin IX, and Ki67 indexes reflecting nuclear activity related to mitosis and cell proliferation, there is a well-established correlation between 5-ALA degree of fluorescence through quantitative fluorometry and Ki67 indexes in astrocytomas (SHIMIZU *et al.*, 2022; VALDÉ S *et al.*, 2011). To our knowledge, this is the first study of tumoral volumetry in

a sample that underwent 5-ALA assisted surgery and to correlate fluorescence margins and segmentation volumetry with Ki67 indexes.

Besides the retrospective study design, this study's main limitation is the subjective aspect of Ki67 sampling and labeling, given the extensive intralesional heterogeneity. Interobserver variation is a known problem of Ki67 indexes, not limited to glioma examination, and within this sample, labeling was analyzed retrospectively and not prospectively under a study's established conditions. Similar limitations are also valid for 5-ALA resections status since we did not perform a quantitative photometric analysis.

Finally, semiautomatic segmentation methods promoted significant advances from manual segmentation, a laborious and time-consuming modality. Similarly, the growth of machine learning algorithms for automatic segmentation in neuroimaging has been exponential. Despite satisfactory accuracy of algorithms, large-scale implementation is still limited secondary to controversial results, literature heterogeneity, and unclear guidelines (VAN KEMPEN et al., 2021). Therefore, semi-supervised methods are still considered the gold standard for glioma quantitative volumetric analysis, considering the value of the expert signalization of the area of interest and real-time over or under inclusion correction.

Finally, considering the dismal prognosis in glioblastoma, this field of research's ultimate goal is to improve these patients' quality of life and survival rates. It is known that Ki67 is an independent prognosis factor in gliomas, and so are fluorescence-compromised margins in 5-ALA microsurgery (CHEN et al., 2015; KLINIK et al., 2006a). However, due to this study's transversal and retrospective nature, we did not perform prognostic analysis in our sample.

Conclusions

There is a correlation between Ki67, the metabolically active tumoral volumes and 5-ALA compromised margins. Tumoral volumetry aspects (WT, ET, and ED), except for NC, correlated directly with Ki67 indexes. We believe the next steps include the validation and reproducibility of these findings in a multicentric and multivendor sample, as well as with more recent and fully automatic segmentation methods.

References

1. Bondy ML, Scheurer ME, Malmer B, Barnholtz-Sloan JS, Davis FG, Il'yasova D, et al. Brain tumor epidemiology: Consensus from the Brain Tumor Epidemiology Consortium. Vol. 113, *Cancer*. 2008. p. 1953–68.
2. Tonn JC, Westphal M, Rutka JT. *Oncology of CNS tumors*. Oncology of CNS Tumors. Springer Berlin Heidelberg; 2010. 1–793 p.
3. Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: A clinical review. Vol. 310, *JAMA - Journal of the American Medical Association*. American Medical Association; 2013. p. 1842–50.
4. Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, et al. Epidemiologic and molecular prognostic review of glioblastoma. Vol. 23, *Cancer Epidemiology Biomarkers and Prevention*. American Association for Cancer Research Inc.; 2014. p. 1985–96.
5. Hanif F, Muzaffar K, Perveen K, Malhi SM, Simjee SU. Glioblastoma multiforme: A review of its epidemiology and pathogenesis through clinical presentation and treatment. Vol. 18, *Asian Pacific Journal of Cancer Prevention*. Asian Pacific Organization for Cancer Prevention; 2017. p. 3–9.

6. Klinik N, Stummer W, Pichlmeier U, Meinel T, Wiestler D, Zanella F, et al. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol* [Internet]. 2006;7:392–401. Available from: <http://oncology.thelancet.com>
7. Stummer W, Tonn JC, Mehdorn HM, Nestler U, Franz K, Goetz C, et al. Counterbalancing risks and gains from extended resections in malignant glioma surgery: A supplemental analysis from the randomized 5-aminolevulinic acid glioma resection study: Clinical article. *Journal of Neurosurgery*. 2011 Mar;114(3):613–23.
8. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg*. 2001;95:190–8.
9. Zinn PO, Majadan B, Sathyan P, Singh SK, Majumder S, Jolesz FA, et al. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. *PLoS ONE*. 2011 Oct 5;6(10).
10. Iliadis G, Kotoula V, Chatzistiriou A, Televantou D, Eleftheraki AG, Lambaki S, et al. Volumetric and MGMT parameters in glioblastoma patients: Survival analysis [Internet]. 2011. Available from: <http://www.biomedcentral.com/1471-2407/12/3>
11. Grabowski MM, Recinos PF, Nowacki AS, Schroeder JL, Angelov L, Barnett GH, et al. Residual tumor volume versus extent of resection: Predictors of survival after surgery for glioblastoma. *Journal of Neurosurgery*. 2014 Nov 1;121(5):1115–23.
12. Wangaryattawanich P, Hatami M, Wang J, Thomas G, Flanders A, Kirby J, et al. Multicenter imaging outcomes study of The Cancer Genome Atlas glioblastoma patient cohort: Imaging predictors of overall and progression-free survival. *Neuro-Oncology*. 2015 Nov 1;17(11):1525–37.
13. Xue D, Albright RE. Preoperative Anaplastic Glioma Tumor Volume Effects on Patient Survival. Vol. 72, *J. Surg. Oncol*. 1999.
14. Andreou J, George AE, Wise A, de Leon M, Kricheff, II, Ransohoff J, et al. CT Prognostic Criteria of Survival after Malignant Glioma Surgery.
15. Hammoud MA, Sawaya R, Shi W, Thall PE, Leeds NE. Prognostic significance of preoperative MRI scans in glioblastoma multiforme. Vol. 27, *Journal of Neuro-Oncology*. 1996.
16. Levin VA, Hoffman WF, Heilbron DC, Norman D. Prognostic significance of the pretreatment CT scan on time to progression for patients with malignant gliomas. Vol. 52, *J Neurosurg*. 1980.
17. Albright RE, Fram EK. Microcomputer-Based Technique for 3-D Reconstruction and Volume Measurement of Computed Tomographic Images - Part 2: Anaplastic Primary Brain Tumors. *Investigative Radiology*. 1988;23(12):886–90.
18. Wood JR, Green SB, Shapiro WR. The Prognostic Importance of Tumor Size in Malignant Gliomas: A Computed Tomographic Scan Study by the Brain Tumor Cooperative Group. Vol. 6, *J Clin Oncol*. 1988.
19. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *New England Journal of Medicine* [Internet]. 2005;352(10):987–96. Available from: www.nejm.org
20. Falcão AX, Bergo FPG. Interactive volume segmentation with differential image foresting transforms. *IEEE Transactions on Medical Imaging*. 2004 Sep;23(9):1100–8.

21. Henker C, Kriesen T, Schneider B, Glass Ä, Scherer M, Langner S, et al. Correlation of Ki-67 Index with Volumetric Segmentation and its Value as a Prognostic Marker in Glioblastoma. *World Neurosurgery*. 2019 May 1;125:e1093–103.
22. Armocida D, Frati A, Salvati M, Santoro A, Pesce A. Is Ki-67 index overexpression in IDH wild type glioblastoma a predictor of shorter Progression Free survival? A clinical and Molecular analytic investigation. *Clinical Neurology and Neurosurgery*. 2020 Nov 1;198.
23. Hu X, Miao W, Zou Y, Zhang W, Zhang Y, Liu H. Expression of p53, epidermal growth factor receptor, Ki-67 and O6-methylguanine-DNA methyltransferase in human gliomas. *Oncology Letters*. 2013 Jul;6(1):130–4.
24. Linn Johannessen A, Torp H. The Clinical Value of Ki-67/MIB-1 Labeling Index in Human Astrocytomas [Internet]. Vol. 12, *PATHOLOGY ONCOLOGY RESEARCH*. 2006. Available from: <http://www.webio.hu/por/2006/12/3/0143>
25. Gerdes J, Schwab U, Lemke H, Stein H. PRODUCTION OF A MOUSE MONOCLONAL ANTIBODY REACTIVE WITH A HUMAN NUCLEAR ANTIGEN ASSOCIATED WITH CELL PROLIFERATION and 2Znstitute of Bioche. 1983.
26. Henker C, Kriesen T, Glass Ä, Schneider B, Piek J. Volumetric quantification of glioblastoma: experiences with different measurement techniques and impact on survival. *Journal of Neuro-Oncology*. 2017 Nov 1;135(2):391–402.
27. Stummer W, Stepp H, Mo Èller G, Ehrhardt A, Leonhard M, Reulen HJ. Technical Principles for Protoporphyrin-IX-Fluorescence Guided Microsurgical Resection of Malignant Glioma Tissue. *Acta Neurochir*. 1998;140:95–100.
28. Stummer W, Stocker S, Wagner S, Stepp H, Fritsch C, Goetz C, et al. Intraoperative Detection of Malignant Gliomas by 5-Aminolevulinic Acid-induced Porphyrin Fluorescence. *Neurosurgery*. 1998;42(3):518–26.
29. Senders JT, Muskens IS, Schnoor R, Karhade A v., Cote DJ, Smith TR, et al. Agents for fluorescence-guided glioma surgery: a systematic review of preclinical and clinical results. Vol. 159, *Acta Neurochirurgica*. Springer-Verlag Wien; 2017. p. 151–67.
30. Coburger J, Wirtz CR. Fluorescence guided surgery by 5-ALA and intraoperative MRI in high grade glioma: a systematic review. Vol. 141, *Journal of Neuro-Oncology*. Springer New York LLC; 2019. p. 533–46.
31. Hadjipanayis CG, Stummer W. 5-ALA and FDA approval for glioma surgery. Vol. 141, *Journal of Neuro-Oncology*. Springer New York LLC; 2019. p. 479–86.
32. Hervey-Jumper SL, Berger MS. Maximizing safe resection of low- and high-grade glioma. Vol. 130, *Journal of Neuro-Oncology*. Springer New York LLC; 2016. p. 269–82.
33. Klinik N, Stummer W, Pichlmeier U, Meinel T, Wiestler D, Zanella F, et al. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol* [Internet]. 2006;7:392–401. Available from: <http://oncology.thelancet.com>
34. Mahmoudi K, Garvey KL, Bouras A, Cramer G, Stepp H, Jesu Raj JG, et al. 5-aminolevulinic acid photodynamic therapy for the treatment of high-grade gliomas. Vol. 141, *Journal of Neuro-Oncology*. Springer New York LLC; 2019. p. 595–607.
35. Picart T, Berhouma M, Dumot C, Pallud J, Metellus P, Armoiry X, et al. Optimization of high-grade glioma resection using 5-ALA fluorescence-guided surgery: A literature review and practical recommendations

from the neuro-oncology club of the French society of neurosurgery. Vol. 65, Neurochirurgie. Elsevier Masson SAS; 2019. p. 164–77.

36. Eidel O, Burth S, Neumann JO, Kieslich PJ, Sahm F, Jungk C, et al. Tumor infiltration in enhancing and non-enhancing parts of glioblastoma: A correlation with histopathology. *PLoS ONE*. 2017 Jan 1;12(1).
37. Curtin L, Whitmire P, Rickertsen CR, Mazza GL, Canoll P, Johnston SK, et al. Assessment of Prognostic Value of Cystic Features in Glioblastoma Relative to Sex and Treatment With Standard-of-Care. *Frontiers in Oncology*. 2020 Nov 16;10.
38. Choi SJ, Hwang HY, Kim NR, Lee SW, Kim JH, Choi HY, et al. The Radiologic Features of Cystic versus Noncystic Glioblastoma Multiforme as Significant Prognostic Factors. *J Korean Soc Radiol*. 2010;63:299–306.
39. Lohle PNM, van Mameren H, Zwinderman KH, Teepen HLJM, Go KG, Wilmink JT. On the pathogenesis of brain tumor cysts: a volumetric study of tumour, edema and cyst. *Neuroradiology*. 2000;42:639–42.
40. Caramanti R, Aprígio RM, D'Aglío Rocha CE, Morais DF, Góes MJ, Chaddad-Neto F, et al. Is Edema Zone Volume Associated With Ki-67 Index in Glioblastoma Patients? *Cureus* [Internet]. 2022 Apr 18; Available from: <https://www.cureus.com/articles/85919-is-edema-zone-volume-associated-with-ki-67-index-in-glioblastoma-patients>
41. Shimizu K, Tamura K, Hara S, Inaji M, Tanaka Y, Kobayashi D, et al. Correlation of Intraoperative 5-ALA-Induced Fluorescence Intensity and Preoperative 11C-Methionine PET Uptake in Glioma Surgery. *Cancers (Basel)*. 2022 Mar 1;14(6).
42. Vald  S PA, Kim A, Brantsch M, Niu C, Moses ZB, Tosteson TD, et al. δ -aminolevulinic acid-induced protoporphyrin IX concentration correlates with histopathologic markers of malignancy in human gliomas: The need for quantitative fluorescence-guided resection to identify regions of increasing malignancy. *Neuro-Oncology*. 2011 Aug;13(8):846–56.
43. van Kempen EJ, Post M, Mannil M, Witkam RL, ter Laan M, Patel A, et al. Performance of machine learning algorithms for glioma segmentation of brain MRI: a systematic literature review and meta-analysis. *European Radiology* [Internet]. 2021;31:9638–53. Available from: www.r-project.org;
44. Chen WJ, He DS, Tang RX, Ren FH, Chen G. Ki-67 is a valuable prognostic factor in gliomas: Evidence from a systematic review and meta-analysis. *Asian Pacific Journal of Cancer Prevention*. 2015;16(2):411–20.

Funding statement

This study did not have any kind of funding.

Conflicts of interest statement

The authors declare no conflicts of interest.

Author contributions

FS and BCAT conceived the study idea and design. SBM, MAC and AXF were responsible for software development and computational resources. FS performed the segmentations. All authors performed data analysis and interpretation. FS was the primary author, and all authors provided revisions to the scientific and grammatical content of the manuscript. BCAT was the head of the study.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics statement

Ethical approval for this study was obtained from the Ethics and Research Committee at Instituto de Neurologia de Curitiba with a waiver of the informed consent form on December 16th 2020 (CAAE 40280320.7.0000.5227).