Exploring adult glioma through MRI: A review of publicly available datasets to guide efficient image analysis

Meryem Abbad Andaloussi[®]**, Raphael Maser, Frank Hertel, François Lamoline**[‡]**, and Andreas Dominik Husch**[‡]

All author affiliations are listed at the end of the article
[‡]These authors contributed equally to this work.

Corresponding Author: Andreas Dominik Husch, PhD, Imaging AI Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg (andreas.husch@uni.lu).

Background. Publicly available data are essential for the progress of medical image analysis, in particular for crafting machine learning models. Glioma is the most common group of primary brain tumors, and magnetic resonance imaging (MRI) is a widely used modality in their diagnosis and treatment. However, the availability and quality of public datasets for glioma MRI are not well known.

Methods. In this review, we searched for public datasets of glioma MRI using Google Dataset Search, The Cancer Imaging Archive, and Synapse.

Results. A total of 28 datasets published between 2005 and May 2024 were found, containing 62 019 images from 5515 patients. We analyzed the characteristics of these datasets, such as the origin, size, format, annotation, and accessibility. Additionally, we examined the distribution of tumor types, grades, and stages among the datasets. The implications of the evolution of the World Health Organization (WHO) classification on tumors of the brain are discussed, in particular the 2021 update that significantly changed the definition of glioblastoma.

Conclusions. Potential research questions that could be explored using these datasets were highlighted, such as tumor evolution through malignant transformation, MRI normalization, and tumor segmentation. Interestingly, only 2 datasets among the 28 studied reflect the current WHO classification. This review provides a comprehensive overview of the publicly available datasets for glioma MRI currently at our disposal, providing aid to medical image analysis researchers in their decision-making on efficient dataset choice.

Key Points

- Twenty-eight different adult glioma datasets were evaluated.
- Only 2 datasets adhere to the newest WHO 2021 tumor classification.
- WHO versions of the datasets are rarely stated: potential tumor misclassification.
- BraTS has complex dataset inclusions: potential bias in machine learning model training and testing.

The study of glioma, some of the most prevalent brain tumor types, has been gaining interest due to advancements in imaging and modeling techniques. Despite their prevalence within the group of intracranial processes, they are overall still a rare disease, with an incidence rate of approximately 3 per 100 000,[1] posing challenges in gathering extensive datasets for training reliable AI models. Various research groups have joined efforts to release a range of publicly available brain tumor datasets, each focusing on a distinct tumor type, study goal, and clinical setup. However, data scarcity remains a

Importance of the Study

Advances in medical imaging and modeling methods have led to the release of various public datasets, which are underused due to the lack of comprehensive information such as histopathological confirmation, which is essential for accurate labeling and compliance with personalized medicine approaches. This work addresses this gap by clarifying the potential applications and limitations of publicly available datasets, thereby guiding researchers in selecting appropriate datasets. We present the first comprehensive and comparative list of public adult glioma magnetic resonance imaging datasets from 2005 to 2024. This work is significant as it bridges the gap between dataset availability and usability for training AI models, potentially accelerating research in glioma imaging research.

major challenge, particularly when considering the very diverse acquisition domains in clinical medical imaging, with different scanners, protocols, patient characteristics, and budget constraints in globally diverse healthcare systems.

Motivation

The regular BraTS (Brain Tumor Segmentation) challenges[2] and their respective datasets, have played a pivotal role in driving the development of brain tumor segmentation algorithms and have boosted the development of the field of medical image analysis overall. Over the last decade, it has helped ML researchers to develop, train, validate, and refine their algorithms. While benchmark-based competitions have made significant contributions, there is a need to expand beyond these curated and annotated datasets from a single source: Such highly preprocessed data—in the example of BraTS even resampled to the same voxel size and aligned to a common space—may not adequately generalize to clinical magnetic resonance imaging (MRI) scans from different institutions and healthcare systems. This strong focus on highly preprocessed datasets with low variance could potentially introduce bias in models and result in poor out-of-domain generalizability.

More specifically for glioma tumors, the currently available public datasets mostly provide MR imaging information. However, many of them lack complementary information such as histopathological confirmation of tumor type following the last World Health Organization (WHO) classification or medical reports, which would lead to some errors in labeling or classification and be therefore not compliant with the current personalized medicine approach.

Scope and Organization of the Article

The purpose of this study is to provide a comprehensive overview of publicly available adult glioma MRI datasets and their different features to medical image analysis researchers, aiding them in more efficient method development. We evaluate 28 different adult glioma datasets between 2005 and May 2024, presenting their properties and application scopes. Among the datasets, we show the complex BraTS inclusions. We present the main features of each dataset such as patients and image number, MRI modalities, tumor types, grades, and corresponding WHO classification.

To the best of our knowledge, this is the first attempt to provide a comprehensive and comparative list of public adult glioma MRI datasets. In 2022, Yearley et al.[3] provided a more general overview of various glioma data registries, including clinical and molecular data resources available for glioma research. However, the authors did not delve into the specific details of MRI imaging studies related to adult glioma and consequently, the implications of MRI features in the diagnosis, monitoring, and treatment of adult gliomas are not discussed. In contrast, the present work has the needs of medical image analysis in focus.

The article is organized as follows. First, the search methodology leading to the selection of datasets is presented in Search Methodology. The search results and the datasets' characteristics are presented in Results. Additionally, in the Discussion, we delve into the practical applications and challenges of these datasets in addressing various potential research questions. Finally, we conclude with perspectives for future works.

Search Methodology

We followed a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses[4]) workflow detailed in the flowchart in Figure 1, which was slightly modified to fit our unique situation where datasets, not studies, are the final object of interest.

To reveal publicly accessible glioma MRI datasets, we searched for datasets using 2 different search methodologies, a direct and an indirect search. For the direct search a dataset search engine, namely the *Google Dataset search* (https://datasetsearch.research.google.com) was queried with the terms "Glioma" and "MRI." The indirect search was performed on 2 common dataset archives, *The Cancer Imaging Archive* (TCIA; https://www.cancerimagingarchive.net) and *Synapse* (https://www.synapse.org/) using the terms "Brain" and "Glioma" respectively. There were no date restrictions in these searches. During the PRISMA screening phase, datasets from the Google dataset search were excluded if they pertained to animals, were pediatric, consisted of 2D images, lacked MRI images, or were not publicly accessible. Duplicates were also removed.

To verify that this approach covered relevant datasets, we conducted a validation experiment by looking for dataset references in published articles obtained via *PubMed*, *medarXiv,* and *arXiv* using the keywords "Glioma," "MRI,"
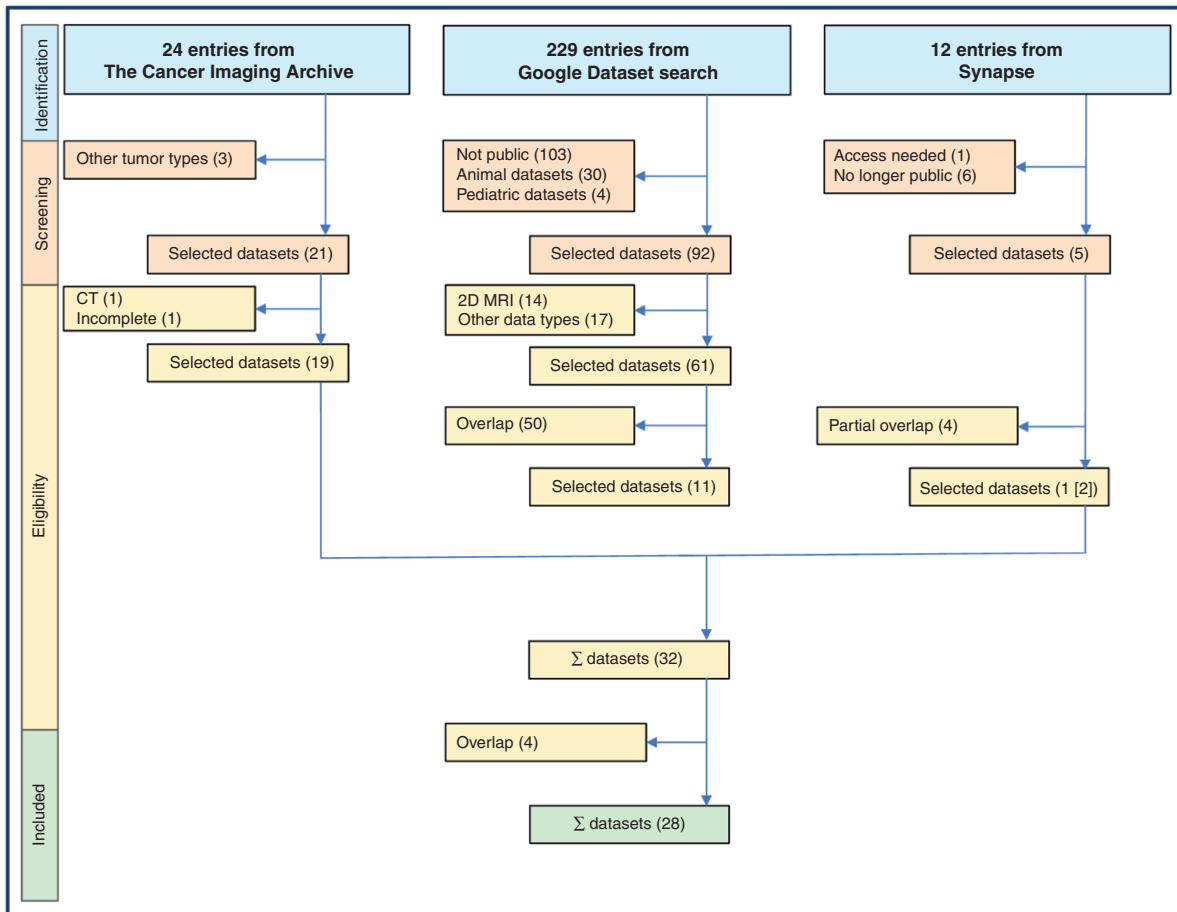
Figure 1. PRISMA workflow diagram for the systematic analysis of public glioma MR Imaging datasets search.

and "Dataset." For PubMed, we limited the verification analysis to the first 50 papers returned. This search yielded no additional datasets, indicating that our search strategy achieved thorough coverage.

Consequently, 28 datasets remained and were included in the study. The BraTS 2023, which is 2 subsets: *BraTS Adult Glioma* and *BraTS Africa*, is the only one of the BraTS challenge datasets further considered in this review due to BraTS dataset inclusion relations discussed in Dataset Overlaps. Note that the field of medical imaging is evolving fast, and thus this review provides a snapshot of the adult glioma MRI dataset state until May 2024.

Results

Four-letter acronyms for each dataset are introduced in Table 1 to enhance the accessibility and readability of our paper.

General Overview of Glioma MRI Public Datasets

The greater goal of all these datasets is to help advance the medical cancer research field through medical image analysis. However, they were acquired and released in different contexts and for different primary purposes. Table 1 introduces the collected datasets, their respective focuses, and the journals that published their related works. Figure S1 in the supplementary material shows the volume in gigabytes and the patient number of the datasets as an overview of their size.

The ADMB and AFMB were integral components of separate *American College of Radiology Imaging Network* protocols, studying the roles of perfusion MRI, MR spectroscopy, and FMISO PET in the treatment response and survival of GBM patients. QGTR and QBDM are affiliated with the *Quantitative Imaging Network* initiative, while the images from TGBM and TLGG are part of more extensive projects of *The Cancer Genome Atlas* focused on understanding genomics.

Among the 28 selected datasets, 21 were associated with published papers, and 7 datasets lack additional publications but are available directly on the TCIA website. Several datasets, including RMND, LUMI, EGD, BITE, UPDG, RGBM, and UGBM, have dedicated papers explaining their contents, see Table 1. The gathered datasets were released for various primary purposes. For example, BTC1, BTC2, and TRTR were involved in comparative studies between glioblastoma patients and healthy controls. Some datasets, such as EGD, DLGG, IRAD, and GLRT, were mainly

Table 1. Overview of Publicly Available Glioma Magnetic Resonance Imaging Datasets

Dataset	Acronym	Focus	Journal/Conference	WHO revision
ACRIN-DSC-MR-Brain [5]	ADMB	Role of perfusion MRI and MR spectroscopy in early treatment response in patients receiving bevacizumab	*Neuro-Oncology*	(2007)
ACRIN-FMISO-Brain [6]	AFMB	FMISO PET and perfusion imaging (Ktrans, CBV) as predictors of survival in GBM	*Clinical Cancer Research*	2007
Brain Images of Tumors for Evaluation database [7]	BITE	Development and validation of new image processing algorithms*	*Medical Physics*	(2007)
Brain-Tumor-Progression [8]	BTUP	Deep learning for tumor progression prediction	*Journal of Digital Imaging*	-
BraTS 2023 Adult Glioma [9] - BraTS 2023 Sub Saharan Africa [10]	BRAG, BRSA	Brain Tumor Segmentation Challenge	—	(2007)
Brain Tumor Connectomics Data Preoperative data [11]	BTC1	Variability of brain activity model parameters between brain tumor patients and healthy controls	*NeuroImage*	(2007)
Brain Tumor Connectomics Data Postoperative data [12]	BTC2	Changes in model parameters from pre-to postoperative assessment	*eNeuro*	(2007)
Burdenko-GBM-Progression	BGBM	Systematic data collection*	—	2016
CPTAC-GBM [13]	CGBM	Cancer phenotypes correlation with proteomic, genomic and clinical data	—	-
Diffuse Low-grade Glioma Database [14]	DLGG	Tumor segmentation methods and preferential localizations	*PLoS One*	(2007)
Erasmus Glioma Database [15]	EGD	Tumor grading and classification*	*Data in Brief*	2016
GLIS-RT [16]	GLRT	Cross-Modality Brain Structures Image Segmentation	*MICCAI 2020*	(2016)
IvyGAP [17]	IGAP	Comprehensive diagnostic characterization of the tumor heterogeneity	*Science*	(2007 or 2016)
IvyGAP-Radiomics [18]	IRAD	Multi-reader segmentation of GBM tumor	*Medical Physics Dataset*	(2007 or 2016)
LGG-1p19qDeletion [19]	LGGD	Predicting 1p/19q Deletion in Low-Grade Gliomas	*Journal of Digital Imaging*	2007
LUMIERE [20]	LUMI	Systematic data collection*	*Scientific data*	2016
QIN-BRAIN-DSC-MRI [21]	QBDM	Multisite/multiplatform analyses of DSC-MR imaging datasets	*American Journal of Neuroradiology*	(2007 or 2016)
QIN GBM Treatment Response [22]	QGTR	Repeatability of relative CBV measurements in newly diagnosed glioblastoma	*American Journal of Neuroradiology}*	(2007)
REMBRANDT [23]	REMB	Connecting clinical information and genomic data	—	(2007)
RHUH-GBM [24]	RGBM	Systematic data collection*	*Data in Brief*	(2016 or 2021)
RIDER NEURO MRI [25]	RIDN	Harmonize data collection and analysis for quantitative imaging	—	2000
ReMIND [26]	RMND	Resource for computational research in brain shift and image analysis	*Scientific data*	2021
TCGA-GBM [27]	TGBM	Connecting phenotypes to genotypes using TCGA clinical images	—	(2007)
TCGA-LGG [28]	TLGG	Connecting phenotypes to genotypes using TCGA clinical images	—	(2007)
Test-retest Reliability Data [29]	TRTR	Activation map quality divergence between brain tumor patients and healthy controls	*PLoS One*	(2007)
UPENN-GBM [30]	UGBM	Systematic data collection*	*Scientific data*	(2016)
UCSF-PDGM [31]	UPDG	Preoperative MRI scans with advanced diffusion and perfusion imaging*	*Radiology: Artificial Intelligence*	2021

Four datasets names were not altered as they already contained less than 4 letters (BITE, BTC1, BTC2, and EGD). The full names of the datasets are according to the TCIA website's collection names. The Test-retest Reliability Data is the name of the file linked to the corresponding paper. The revision of the WHO classification of tumors applied in each dataset is not always available. The AFMB, EGD, LGGD, LUMI, RMND and UPDG are the only ones with clear mention of the WHO classification year. WHO revision years in parenthesis are estimated based on the dataset and corresponding paper publication dates. To avoid any wrong affirmation, the WHO year of the BGBM, CGBM, and RIDN was not added. * Published separately
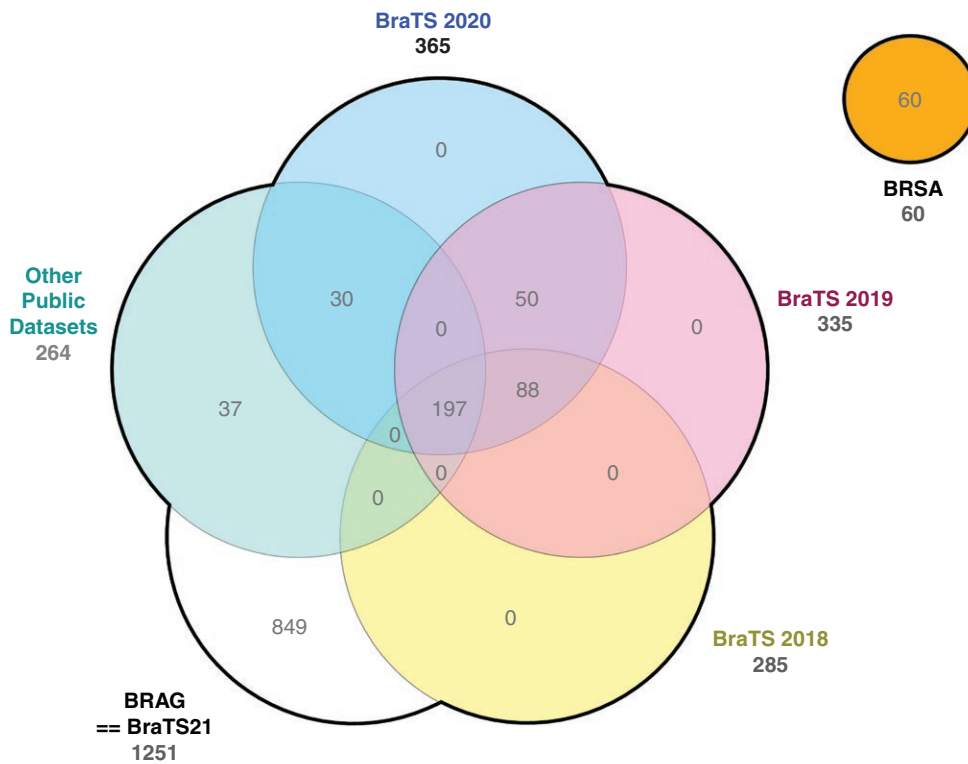
Figure 2. Adult glioma data inclusion relationships over time in BraTS training datasets based on BraTS 2023, which consists of the BRAG and BRSA datasets. Other public datasets include AFMB, TLGG, TGBM, CGBM, and BraTS 2013. BRSA is a new dataset

published to enhance image segmentation methods while others focus on longitudinal problems such as brain shift and post-surgical tumor segmentation (RMND). The purpose of some datasets was to include less conventional MRI modalities, like diffusion and perfusion (eg, UPDG), while others used these modalities to study the repeatability of perfusion measurements across institutions and patients (QGTR, QBDM, and ADMB). Additionally, some datasets were designed with the primary objective of linking imaging data with other types of data, such as genomic, proteomic, and clinical information (LGGD, REMB, CGBM, TGBM, and TLGG).

This comprehensive approach highlights the diverse objectives and applications of the examined datasets in advancing our understanding of glioma imaging and analysis. In the following analysis, we have chosen to focus on the properties of the datasets that are relevant to common medical imaging and, in particular, machine learning approaches.

Patient Number

In total, the datasets gather 5515 patients where the BRAG, the EGD, and the UGBM account for approximately 26.6%, 14.03%, and 11.4% of the total patient number. The BGBM, TLGG, TGBM, DLGG, GLRT, LGGD, REMB, ADMB, and the UPDG cover between 2.2% and 8.9%, and the rest of the datasets cover less than 2% of the total patient number.

Note that these numbers do not account for potential overlaps between datasets, which are analyzed in the next sections.

Dataset Overlaps

We found that some datasets have (inclusion) relationships, that have to be carefully considered in studies based on this data, for example, to avoid model bias by undetected double inclusions of data, leading to data leakage. Such a relationship exists between the IRAD and the IGAP, where IRAD contains the preoperative MRIs of the IGAP datasets with additional segmentations and derived radiomics parameters; however, the IGAP dataset is longitudinal, while the IRAD is not. We note that the IRAD contains 2 less patients than the IGAP (37 instead of 39).

The well-known BraTS challenge datasets are a special case with respect to complex inclusion relationships that evolved over time. For example, *BraTS 2021* contains data that was available in the previous BraTS challenges and other public datasets. *BraTS 2023* further expands on that by extending to 6 different dataset parts, which are used in 9 different challenges.[32] Here, the "Adult Glioma" subdataset is the *BraTS 2021* dataset. The other sub-datasets mostly include non-glioma or pediatric data, except for the *BraTS Sub-Saharan Africa* subset, which is in turn a collection of new glioma imaging data from the African continent.
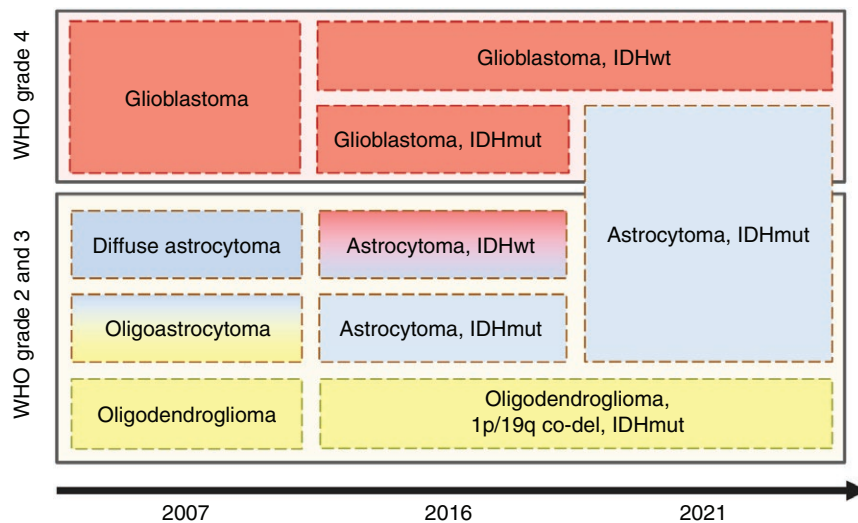
Figure 3. The development of tumor classifications over the last 3 World Health Organization revisions for the main classes of glioma tumors. Note that astrocytoma IDH wild type is considered glioblastoma IDH wild type since the 2021 revision.

BraTS 2021 added 849 new patients compared to the previous version. The overall inclusion relationships for the versions of BraTS are shown in Figure 2. The *BraTS 2021* includes multiple patients from previous datasets: in total, 365 are in *BraTS 2020*, 335 in *BraTS 2019*, 285 in *BraTS 2018*, 285 in *BraTS 2017,* and 264 from other public datasets. More specifically, 30 are in *BraTS 2013*, 65 are in TLGG, 102 are in TGBM, and 30 in IGAP. For consistency purposes, the *BraTS 2024* dataset was excluded from this search as the challenge is still ongoing. The exact inclusions along with the validation dataset were not published yet at the time of the study.

Dataset Release and Update Dates

The datasets in Figure S2 are organized from the most recently released (RGBM and BGBM and UPDG in 2023) to the least recent (RIDN in 2011). Since their initial release dates, 12 datasets have been updated. The CGBM stands out by being updated 14 times from 2018 until 2021. We categorized the updates into 4 distinct types: scans, patient information, metadata files, and external parameters. These updates included lifting access restrictions, modifying file paths, and altering downloaders. Figure S2 summarizes the various updated datasets categorized by the type of update. Note that a dataset may appear in multiple sections if it has undergone different types of updates.

Scans.—The QBDM dataset added 6 new series in its second update, while the LGGD dataset improved the published segmentations and changed the data format from NIfTI to DICOM. The CGBM dataset received a general data cleanup to remove extraneous scans. In the TGBM dataset, a DICOM tag was repaired in 5 series for 1 patient. Finally, 30 DWI MRIs from patients in the AFMB dataset were removed due to inconsistencies in b-value acquisition between GE and Siemens scanners, preventing the reconstruction of ADC maps.

Patient.—The BTC1 and BTC2 excluded 6 patients, whereas the CGBM (v1 to v12) and the TGBM (v2) added more patients.

Metadata.—In the AFMB (v2) a new clinical metadata file including the age, treatment, and health condition information was added along with some row corrections. One patient tumor type was corrected in CGBM (v12) while more general updates were mentioned for TGBM and TLGG. The only update to the UGBM dataset was related to adding histopathology NDPI slides and updating CSV files for mapping Radiology subject IDs to Histopathology patients. As this change might be related to imaging data, we choose to put it with the metadata update as we only consider MRI data updates in the scans paragraph.

External.—The external update type is the one with less impact as it refers to changes in the dataset format or access methods, rather than modifications of the actual content of the dataset. This type of update does not alter the information contained in the dataset itself. For example, the access embargo was lifted in QGTR, the download link for histopathology slides in the CGBM (v13) was changed and the download location was altered for some files in the IRAD.

Dataset WHO Classification Date

The 28 datasets have been collected between 2005 and May 2024. Gliomas are classified into different grades per the WHO. The initial classification was introduced in 1979 with editions in 1993, 2000, 2007[33] (updated 2016[34]), and 2021.[35] These subsequent updates enabled the classification
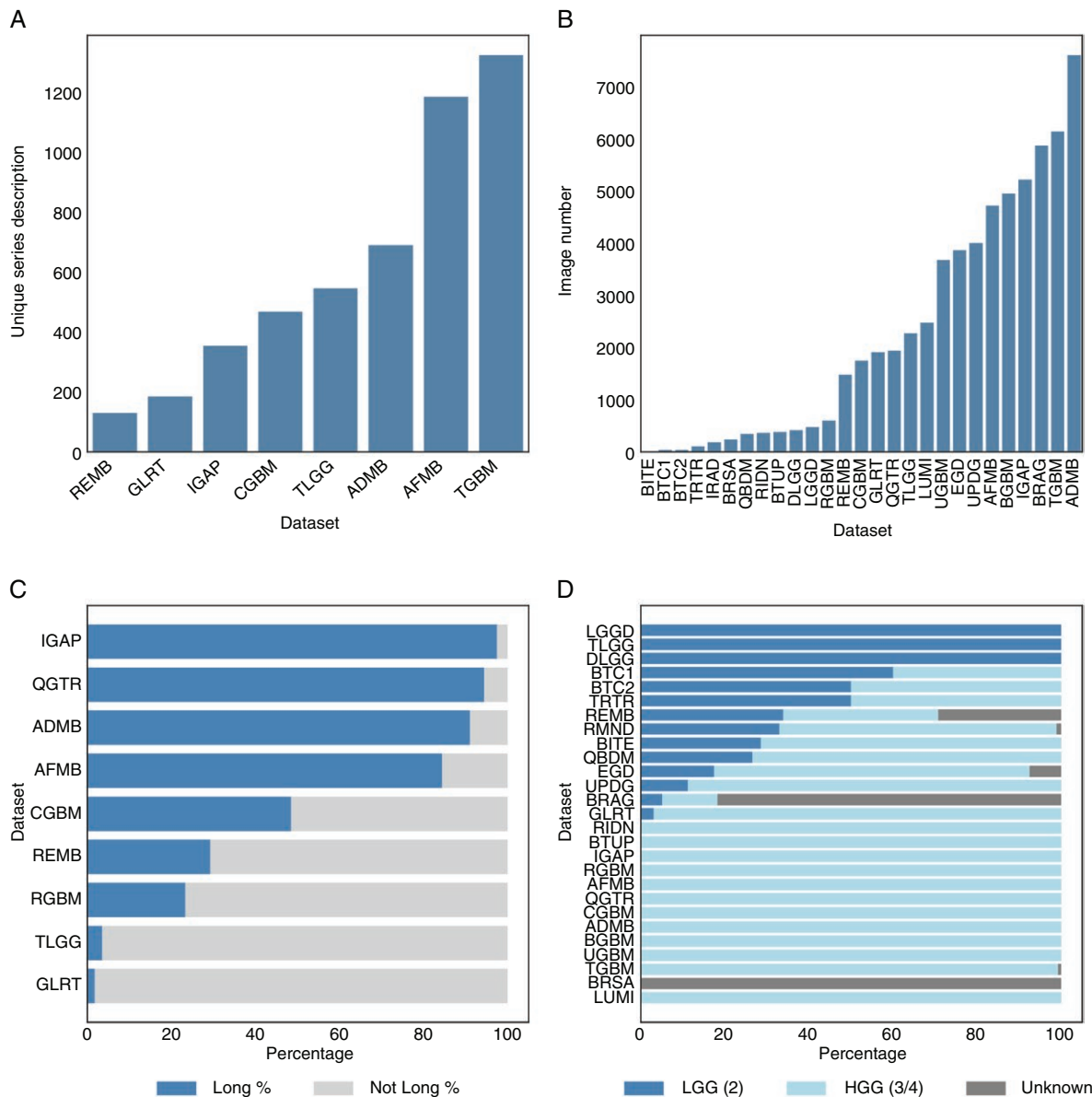
Figure 4. **(A)** Number of distinct magnetic resonance imaging series descriptions of the 8 remaining datasets **(B)** Total image number per dataset **(C)** Percentage of longitudinal studies across mixed datasets. Long % stands for the percentage of longitudinal datasets **(D)** Percentage of tumor grades across datasets.

to evolve with main research breakthrough discoveries namely the 1p/19q chromosome codeletion and the isocitrate dehydrogenase (IDH) mutation status,[36,37] which lead to substantial changes in the 4th and 5th editions published in 2007/2016 and 2021. Figure 3 provides an overview of the main changes in glioma-type classification since the introduction of molecular diagnostics.[38]

The WHO grading system includes grades 2, 3, and 4.[39] In the WHO 2021 revision, grades 2 and 3 include 2 subtypes: oligodendrogliomas (IDH-mutant and 1p/19q chromosome codeleted) and astrocytomas (IDH-mutant with no 1p/19q codeletion). Grade 4 tumors are categorized into astrocytoma (IDH-mutant) and glioblastoma (IDH-wild type)

as the glioblastoma IDH-mutant type does not exist anymore. (Figure 3).

Among the datasets collected, only AFMB (WHO 2007), LGGD (WHO 2007), EGD, LUMI, BGBM (WHO 2016), the RMND, and UPDG (WHO 2021) explicitly state the specific WHO edition version employed for tumor classification. For the remaining datasets, the WHO edition was inferred based on the publication date and associated articles, where available. More specifically, to avoid any wrong affirmation, the WHO year of the CGBM, and RIDN were not added in this study (Table 1).

Finally, most of the datasets are believed to follow the WHO 2007 with respect to tumor classification. However,

this information cannot be confirmed without contacting the authors of the corresponding datasets.

Longitudinal Studies

Out of the 28 datasets, 13 are cross-sectional, 6 are fully longitudinal (LUMI, BGBM, RGBM, BTUP, RIDN, RMND), and the remaining 9, mixed datasets containing both cross-sectional and longitudinal studies, are shown in Figure 4C. In the mixed datasets, the IGAP is the dataset with the maximum percentage of longitudinal data (38 of 39 patients), while the GLRT represents the dataset with the least percentage of longitudinal data (4 of 226 patients).

Within the longitudinal and mixed datasets, 3 datasets have the same number of sessions for all longitudinal patients (2 for the BTUP dataset and the RIDN and 3 for the BGBM dataset).

The IGAP contains the maximum number of timepoints for one patient (27 timepoints) followed by the AFMB and the TGBM with 25 and 23 timepoints.

Tumor Grades

In the rest of the paper, grade 2 gliomas are considered low-grade gliomas (LGG) while grade 3 and 4 gliomas are classified as higher-grade gliomas (HGG) The availability of tumor grade information varied across different datasets. In total, we have 58.5% of HGG, 17% of LGG, and 24.3% labeled as "unknown" including patients from datasets with incomplete or missing tumor grade data. The exact tumor grade for BRAG and BRSA is not publicly available. Therefore, all patients in BRSA were categorized as unknown grade. In contrast, for BRAG, we deduced the grade from a mapping file linking the patients from BRAG to their original source datasets. The naming of these original source datasets often included the keywords "LGG" or "HGG" indicating the tumor grades allowing us to infer the tumor grades. Through this method, we were able to identify at least 193 patients with HGG and 75 patients with LGG in the BRAG dataset.

Additionally, it is worth noting that the patient number and associated tumor grades and types of TRTR were taken from the associated published paper as it was difficult to deduce it from the dataset alone. The tumor grades and types of the QBDM were also extracted from the associated paper. In the 28 datasets, we counted 944 LGG (grade 2), 3233 HGG (grade 3 and 4), and 1344 unknown grades as shown in Figure 4D.

Tumor Types

According to the WHO classification, glioma tumor types are identified according to the IDH mutation and the 1p19q codeletion information.

Regarding the former, the IDH status is easily available for all patients of the UPDG and the RGBM datasets and is partially available for the EGD (467/774 patients), the UGBM (515/630 patients), the BGBM (114/180 patients) and the LUMI (58/91 patients) datasets.

The 1p19q codeletion, on the other hand, is clearly available for all patients of the LGGD datasets, and only partially available

for the EGD (259/774 patients), and the UPDG (405/630 patients) datasets.

Both histopathological and molecular classification criteria are less clear in other datasets which either do not provide them or require a search on the website of the dataset provider. This concerns datasets such as the TLGG, TGBM, or IGAP/IRAD.

Furthermore, the specific tumor type may be guessed from the IDH mutation and 1p19q codeletion information if available. For example, an IDH mutant tumor with no 1p19q codeletion may be denoted as astrocytoma. As such, one can conclude on the WHO 2021 corresponding classification. This process is possible in a few datasets (EGD, UGBM, UPDG, BGBM, and RGBM). On the contrary, when the dataset contains the tumor type along with the exact WHO revision, the IDH mutation and 1p19q codeletion can be deduced (RMND).

Magnetic Resonance Images in the Datasets

In Table 2 we focus on MRI modalities and segmentations in 20 datasets for which the information was available in the description. Note that the number of images and of subjects strictly include glioma and MRI. Thus, we consider a subset of BITE (removing ultrasound imaging), BTC1 and BTC2 (removing all non-glioma MRI), RMND (removing ultrasound imaging and all non-glioma MRI), and TRTR (removing healthy control patients). Observe that 10 datasets (BRAG, BRSA, BGBM, BTUP, EGD, IRAD, RGBM, UGBM, UPDG) provide all 4 conventional MRI modalities (T1, T2, T1 Contrast Enhanced (T1CE) and FLAIR). Diffusion and diffusion-derived (ADC) information is available in 7 Datasets while 6 datasets provide perfusion information (ASL, DSC, DCE) or corresponding derived information such as the rCBV. Three datasets provide fMRI images while SWI and HARDI MRIs are available in UPDG. LGGD provides T2CE MRIs for all patients.

The tumor segmentation information is included in 14 datasets: in BITE and DLGG, the masks are in minc and XML format respectively. Both GLRT and BGBM provide the gross tumor volume and contour tumor volume that are more commonly used in clinics. The RMND dataset comprises presurgical tumor segmentation, intraoperative residual tumor, and automatic segmentations of cerebrum and ventricles for some cases.

MRI information in the 8 remaining datasets was not explicitly documented in the general descriptions. To address this gap, we turned to the metadata files associated with these datasets. Unfortunately, due to the high number of unique MRI series descriptions and images (Figures 4A and 4B), finding the corresponding modalities and segmentations for each of them was out of the scope of this study.

Discussion

In this section, we discuss 4 main points: (1) the practical challenges imposed by the WHO classification updates in the datasets, (2) different problem-specific quality criteria applied across datasets, (3) implications of specific

Table 2. Adult Glioma Magnetic Resonance Imaging Modalities and Segmentations Available Per Dataset

Dataset	Image N.	Subject N.	T1	T1CE	T2	FLAIR	Diffusion	Perfusion	Additional	Tumor	Additional	Format	Skull strip.	Registered
			MRI modality							Segmentation		Preprocessing		
BGBM	4956	180	✓	✓	✓	✓	X	X	X	✓ (GTV-CTV-PTV)	✓	DICOM	X	X
BITE	14	14	X	✓	X	X	X	X	X	X	X	MINC	X	✓,MNI
BRAG	5880	1470	✓	✓	✓	✓	X	X	X	✓3-label	X	NIFTI	✓	✓,SRI
BRSA	240	60	✓	✓	✓	✓	X	X	X	✓3-label	X	NIFTI	✓	✓,SRI
BTC1	40	10	✓	X	X	X	✓ (DWI)	X	✓ (BOLD)	X	X	NIFTI	X	X
BTC2	40	10	✓	X	X	X	✓ (DWI)	X	✓ (BOLD)	X	X	NIFTI	X	X
BTUP	383	20	✓	✓	✓	✓	✓ (ADC)	✓ (nCBF-crCBV-srCBV-DSC)	X	✓	X	DICOM	X	✓,intra-subject
DLGG	420	210	X	X	X	✓	X	X	X	✓ (.xml)	X	NIFTI	X	X
EGD	3870	774	✓	✓	✓	✓	X	X	X	✓	X	NIFTI	X	✓,MNI
IRAD	185	37	✓	✓	✓	✓	X	X	X	✓	X	NIFTI	✓	✓,SRI & MNI
LGGD	478	159	X	✓	X	X	X	X	✓ (T2CE)	✓	X	DICOM	X	X
QBDM	349	49	X	✓	X	X	X	✓ (DSC)	X	✓	X	DICOM	X	✓,intra-subject
QGTR	1942	54	✓	✓	✓	✓	✓	✓ (DCE-DSC)	✓ (MEMRAGE)	X	X	DICOM	X	✓,intra-subject
RGBM	600	40	✓	✓	✓	✓	X	X	X	✓	X	NIFTI	✓	✓,SRI
RIDN	368	19		✓		✓ (P)	✓ (DTI)	✓ (DCE)	X	X	X	DICOM	X	X
TRTR	108	12	X	X	X	X	X	X	✓ (fMRI)	X	X	NIFTI	X	X
UGBM	3680	630	✓	✓	✓	✓	✓ (P)	✓ (P)(DSC)	X	✓(P)	X	NIFTI	✓	✓,unknown atlas
UPDG	4008	495	✓	✓	✓	✓	✓	✓ (ASL)	✓ (SWI-HARDI)	✓	X	NIFTI	✓	✓,intra-subject
LUMI	2478	91	✓	✓	✓	✓	X	X	X	✓3-label	X	NIFTI	✓	✓,unknown atlas
RMND	841	91	(P)	(P)	(P)	(P)	X	X	X	✓	X	DICOM	X	✓,intra-subject

(P) partially available. The datasets with unclear MRI descriptions were excluded from this table. Note that some data sets contain additional non-glioma data not counted here.

research questions on dataset choice and finally, and (4) limitations of this review.

Importance and Practical Implications of the WHO Classification

The WHO's updates to the central nervous system tumor classification edition, driven by an increased understanding of molecular factors, have significant and immediate implications for radiologists and neuropathologists.[40,41] Nevertheless, the practical implementation of these changes in the medical field is not immediate, as it requires the adaptation of jargon and classification systems that have been used daily for more than 25 years since the first edition was published in 1979.[42,43]

The nature of this process may appear insignificant, yet its implications for patient care, clinical trials, and training of artificial intelligence models are undeniable. Specifically, the transition delay could significantly impact public glioma datasets and all downstream development done based on them. For example, very popular and large datasets appear to still adhere to the 2007 edition of the WHO classification. Consequently, machine learning models trained on them will produce diagnoses aligned with the 2007 standard. In particular, such models—like the ones trained on BraTS—would likely misclassify grade 4 astrocytomas as glioblastomas (Figure 3).

Updating the classification labels in existing datasets is often infeasible due to the absence of molecular and/or histopathological tests as shown in Tumor Types. For example, LGGD adheres to the 2007 revision. However, even though the 1p19q deletion status is included and would suggest oligodendroglioma, the unknown IDH mutation status prevents us from confidently providing an updated WHO classification. The EGD dataset is an additional example, with 291 patients lacking both the IDH mutation and 1p19q codeletion information. Another problem regarding WHO classification updates lies within longitudinal datasets. For patients with recurrent tumors and long follow-up, the WHO classification might have changed during the course of the disease. None of the longitudinal datasets was updated accordingly which might affect longitudinal studies of treatment response. Also, 61 % of the tumors in TLGG have been classified as "oligoastrocytoma" which are tumors considered as a mixture of cells that originated from oligodendrocytes and astrocytes. This term was removed in the WHO revision in 2016, and additional information would be required to re-classify these tumor types (1p19q codeletion and IDH mutation).

Problem-Specific Quality Criteria

The most appropriate dataset for a specific problem changes depending on the intended use. For example, if the study focuses on a specific tumor type, appropriate datasets are limited to sets that include detailed type classification, including IDH mutation status. For this case, the EGD, UGBM, UPDG, BGBM, and RGBM datasets would be the preferred choice.

If the intended study focuses on specific differences between low-grade and high-grade tumors, the imbalance of datasets needs to be considered. As introduced previously, more than 50% of patients across all the datasets have a high-grade tumor.

For observing tumor evolution or treatment effect on tumor volume over time, datasets are restricted to the ones that provide longitudinal MRIs. For that purpose, Figure 4C illustrates the percentage of patients with longitudinal MRIs within all longitudinal datasets. The most adequate choice in that case might be the IGAP dataset.

Besides the WHO classification date, the study date itself might be crucial as well. Overall MRI quality is expected to be better in 2023 than in 1995. This needs to be taken into account when training models intended to have clinical impact.

Potential population shifts are also to be considered. For example, different healthcare systems across the globe might lead to different stages where MR imaging takes place, different scanner hardware generations used, and different therapy regimes, in turn, influencing the imaging phenotype. It is important to note that most medical imaging datasets are centered in the United States or Europe. In this context, BRSA stands out as the only dataset specifically designed for Africa.

Datasets that include tumor masks along with the intended MRI modality facilitate the training of models for tumor segmentation.

Research Questions Examples and Available Public Datasets

In this section, we discuss the relevance of dataset use in the scope of 3 different research questions: tumor growth, malignant transformation (MT), and tumor segmentation.

Tumor growth.—One of the central questions in cancer research is understanding the causes of tumor evolution over time. This is necessary for predicting tumor growth, customizing treatments, and potentially preventing the tumor from reaching a critical stage. These research questions require longitudinal patient data. In our study, within the mixed datasets, only 4 datasets are primarily longitudinal (IGAP, QGTR, ADMB, and AFMB) and all of them include already grade 4 glioblastomas with treatment and resection surgeries. These would be appropriate to answer research questions related to multisite analysis of treatment response. Nonetheless, postoperative MRIs might be present in these datasets affecting tumor growth and heterogeneity prediction. In addition, brain shift makes it very difficult to identify spatial differences in tumor evolution. According to surgeons' opinion, 6 weeks approximately are needed for the brain shift's effect to be negligible.[44] However, the RMND dataset might help tackle this issue as it contains preoperative MRIs along with registered intraoperative MRIs and, in some cases, residual tumor segmentations.

Malignant transformation.—Tumor location and size are not the only factors that evolve through time. Tumor heterogeneity and aggressivity also change. More specifically, LGG with IDH mutation almost always recurs as a higher more aggressive grade through MT.[45,46] This process occurs gradually, through changes in the tumor micro-environment.[47,48] However, the reasons leading to MT are not yet fully understood. This is why, using available patient data, mathematical models attempt to describe tumor growth and heterogeneity until the MT is diagnosed.[49] In this case, longitudinal datasets alone are not sufficient. Additional criteria are needed: IDH mutant grade 2 gliomas with at least 3 timepoints serving to initialize, calibrate, and evaluate a patient-specific mathematical model as described in[50,51] and[52] on brain and breast tumors. Furthermore, MT must have been histopathologically confirmed at the third time point. In the studied longitudinal datasets, only 1 patient with a grade 2 to 3 astrocytoma was found (TCGA- LGG) that meets this criterion. However, only 2 imaging sessions were identified, making the evaluation step of mathematical modeling not possible for this patient. Conventional MRIs (T1, T2, FLAIR, T1CE) are sufficient for tumor segmentation. However, more MRI modalities are needed to inform about the tumor micro-environment during MT, such as the Apparent Diffusion Coefficient (ADC) map describing diffusion of water molecules[53,54] and the (relative Cerebral Blood Volume) rCBV for tumor vascularization.[55–57] To address this gap, a solution would be to predict the missing MRI modalities, leading to new research questions.[58–60]

Tumor segmentation.—The BraTS datasets are the most prominent ones specifically published for segmentation and were released in a challenging context. As such, the MRIs are highly preprocessed: Skull stripped, registered to the same template, in the NIfTI format, and resampled to 1 *mm*3. From Table 2, 8 datasets may potentially be added to BraTS to train a model for tumor segmentation (BGBM, BTUP, IRAD, LUMI, RGBM, RMND, UGBM, and UPDG). Segmentation models trained on such datasets may be used for tumor segmentations in datasets that include the 4 classical MRIs such as QGTR but no segmentation maps.

Limitation of the Review

We recognize several limitations of our review that might be addressed in further work. Firstly, due to the WHO classification changes, it is challenging to apply the current tumor type specification without information about IDH and 1p19q status in existing datasets, thus we cannot provide a complete and detailed overview of the covered tumor types according to the recent 2021 specification. Secondly, 8 datasets out of the 28 were excluded from Table 2 due to unclear MRI descriptions, which might be partially recoverable by the original dataset authors. Finally, the field is moving fast and new datasets are published regularly. Therefore, this review can only be considered as a snapshot of adult glioma datasets until May 2024.

Conclusion

In the context of the data scarcity faced by computational researchers in medical imaging, publicly available datasets represent a valuable asset. In this article, we provided a comprehensive overview of MRI public adult glioma datasets and highlighted their potential and challenges. The resources would serve as a foundation for researchers studying adult glioma tumors.

Across the 28 gathered datasets, only the UPDG and the RMND datasets follow the current WHO classification criteria. For the UPDG dataset, all classical MRIs are available along with diffusion and perfusion MRI modalities. A tumor mask is also provided in NIfTI format and the main preprocessing steps are performed (skull stripping and co-registration). Additionally, the corresponding histopathological and molecular characteristics are available along with treatment-relevant information such as MGMT methylation status. Regarding RMND, classical preoperative MRIs are available for almost all patients in DICOM format. Intraoperative MRIs are also available. The tumor type along with its corresponding WHO classification are informed for all patients except one. In the Discussion, we discussed the possible usage of the different datasets for specific research purposes. A careful selection is crucial, and researchers must define their study objectives precisely. The IGAP, QGTR, ADMB, and AFMB datasets are relevant for studying glioblastoma response to treatment, while BGBM, BRAG, BRSA, BTUP, EGD, IRAD, QGTR, RGBM, UGBM, and UPGD could be relevant for T1CE contrast agent uptake analysis. Model training for tumor segmentation could benefit from the BGBM, BTUP, IRAD, RGBM, UGBM, and UPDG datasets.

As part of our ongoing work, we are developing tools that facilitate easier and standardized access to these datasets within the research community. The search for datasets for this study stopped in May 2024, and new datasets are being released regularly. A natural continuity of this work would involve tracking the emergence of new datasets and updating this review accordingly. More generally, it is important for medical and research communities to collaborate in the creation of such datasets to respect as much as possible the quality criteria expected from both fields—imaging data alone is not sufficient for many research questions, but coupling with molecular information is needed. An ideal scenario would be publishing such datasets while finding a mechanism that leaves the door open for updates in case of changes such as the WHO classification. Another idea would be to agree on minimum specific publication quality criteria for such datasets to reach a standard dataset format in the future.

Supplementary material

Supplementary material is available online at *Neuro-Oncology Advances* (https://academic.oup.com/noa).

Keywords

computer-aided diagnosis | medical image analysis | MRI glioma datasets

Conflict of interest statement

All research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authorship statement

M.A.A. Data curation, Investigation, Validation, Methodology, Visualization, Writing—original draft; R.M. Software, Methodology; F.H. Conceptualization, Writing—review & editing; F.L. Methodology, Investigation, Supervision, Validation, Writing—original draft; A.H. Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing—review & editing.

Data availability

The datasets analyzed in this study are publicly available. All relevant metadata used to conduct the dataset review are published in the main text or supplementary files.

Affiliations

Imaging AI Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg (M.A.A., R.M., F.L., A.D.H.); Faculty of Science, Technology and Medicine, University of Luxembourg, University of Luxembourg, Belvaux, Luxembourg (M.A.A.); National Department of Neurosurgery, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg (F.H.)

References

1. Pati S, Baid U, Edwards B, et al; others. Federated learning enables big data for rare cancer boundary detection. *Nat Commun.* 2022;13(1):7346.
2. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* 2014;34(10):1993–2024.
3. Yearley AG, Iorgulescu JB, Chiocca EA, et al. The current state of glioma data registries. *Neurooncol Adv.* 2022;4(1):vdac099.
4. Moher D, Shamseer L, Clarke M, et al; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4(1):1.
5. Boxerman JL, Zhang Z, Safriel Y, et al. Early post-bevacizumab progression on contrast-enhanced MRI as a prognostic marker for overall survival in recurrent glioblastoma: Results from the ACRIN 6677/RTOG 0625 Central Reader Study. *Neuro Oncol.* 2013;15(7):945–954.
6. Gerstner ER, Zhang Z, Fink JR, et al; ACRIN 6684 Trial Group. ACRIN 6684: Assessment of tumor hypoxia in newly diagnosed glioblastoma using 18F-FMISO PET and MRI. *Clin Cancer Res.* 2016;22(20):5079–5086.
7. Mercier L, Del Maestro RF, Petrecca K, et al. Online database of clinical MR and ultrasound images of brain tumors. *Med Phys.* 2012;39(6Part1):3253–3261.
8. Schmainda K, Prah M. Data from brain-tumor-progression. *Cancer Imaging Arch.* 2018;21.
9. Baid U, Ghodasara S, Mohan S, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:210702314.* 2021.
10. Adewole M, Rudie JD, Gbdamosi A, others. The brain tumor segmentation (brats) challenge 2023: Glioma segmentation in sub-saharan Africa patient population (brats-africa). *ArXiv.* 2023.
11. Aerts H, Schirner M, Jeurissen B, et al. Modeling brain dynamics in brain tumor patients using the virtual brain. *eNeuro.* 2018;5(3):ENEURO.0083–ENEU18.2018.
12. Aerts H, Schirner M, Dhollander T, et al. Modeling brain dynamics after tumor resection using the virtual brain. *Neuroimage.* 2020;213:116738.
13. NCICPTAC (CPTAC). Radiology data from the clinical proteomic tumor analysis consortium glioblastoma multiforme [CPTAC-GBM] collection [Data set]. *Cancer Imaging Arch.* 2018.
14. Parisot S, Darlix A, Baumann C, et al. A probabilistic atlas of diffuse WHO grade II glioma locations in the brain. *PLoS One.* 2016;11(1):e0144200.
15. van der Voort SR, Incekara F, Wijnenga MMJ, et al. The erasmus glioma database (EGD): Structural MRI scans, WHO 2016 subtypes, and segmentations of 774 patients with glioma. *Data Brief.* 2021;37:107191.
16. Shusharina N, Bortfeld T, Cardenas C, et al. Cross-modality brain structures image segmentation for the radiotherapy target definition and plan optimization. In: Segmentation, Classification, and Registration of Multi-Modality Medical Imaging Data: MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 23; 2021:3–15.
17. Puchalski RB, Shah N, Miller J, et al. An anatomic transcriptional atlas of human glioblastoma. *Science.* 2018;360(6389):660–663.
18. Pati S, Verma R, Akbari H, et al. Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the ivy glioblastoma atlas project (Ivy GAP) dataset. *Med Phys.* 2020;47(12):6039–6052.
19. Akkus Z, Ali I, Sedlář J, et al. Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence. *J Digit Imaging.* 2017;30(4):469–476.
20. Suter Y, Knecht U, Valenzuela W, et al. The LUMIERE dataset: Longitudinal glioblastoma MRI with expert RANO evaluation. *Sci Data.* 2022;9(1):768.
21. Schmainda KM, Prah MA, Rand SD, et al. Multisite concordance of DSC-MRI analysis for brain tumors: Results of a national cancer institute quantitative imaging network collaborative project. *AJNR Am J Neuroradiol.* 2018;39(6):1008–1016.
22. Prah MA, Stufflebeam SM, Paulson ES, et al. Repeatability of standardized and normalized relative CBV in patients with newly diagnosed glioblastoma. *AJNR Am J Neuroradiol.* 2015;36(9):1654–1661.
23. Scarpace L, Flanders AE, Jain R, Mikkelsen T, Andrews DW. Data From REMBRANDT [Data set]. *Cancer Imaging Arch.* 2019.
24. Cepeda S, Garcia-Garcia S, Arrese I, et al. The R\'io Hortega University Hospital Glioblastoma dataset: a comprehensive collection of preoperative, early postoperative and recurrence MRI scans (RHUH-GBM). *arXiv preprint arXiv:230500005.* 2023.
25. Barboriak D. Data from RIDER_NEURO_MRI. *Cancer Imaging Arch.* 2015:577.
26. Juvekar P, Dorent R, Kögl F, et al. ReMIND: The Brain Resection Multimodal Imaging Database. *Sci Data.* 2024;11(1):494.
27. Scarpace L, Mikkelsen T, Cha S, et al. The cancer genome atlas glioblastoma multiforme collection (TCGA-GBM)(Version 4)[Data set]. *Cancer Imaging Arch.* 2016.
28. Pedano N, Flanders AE, Scarpace L, et al. The Cancer Genome Atlas Low Grade Glioma Collection (TCGA-LGG)(version 3)[Data set]. *Cancer Imaging Archive.* 2016.
29. Morrison MA, Churchill NW, Cusimano MD, et al. Reliability of task-based fMRI for preoperative planning: A test-retest study in brain tumor patients and healthy controls. *PLoS One.* 2016;11(2):e0149547.

30. Bakas S, Sako C, Akbari H, et al. The University of Pennsylvania Glioblastoma (UPenn-GBM) cohort: Advanced MRI, clinical, genomics, & radiomics. *Sci Data.* 2022;9(1):453.

31. Calabrese E, Villanueva-Meyer JE, Rudie JD, et al. The university of California san francisco preoperative diffuse glioma mri dataset. *Radiol Artif Intell.* 2022;4(6):e220058.

32. Li HB, Conte GM, Anwar SM, et al. The brain tumor segmentation (BraTS) challenge 2023: brain MR image synthesis for tumor segmentation (BraSyn). *ArXiv.* 2023.

33. Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 2007;114(2):97–109.

34. Louis DN, Perry A, Reifenberger G, et al. The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* 2016;131(6):803–820.

35. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro Oncol.* 2021;23(8):1231–1251.

36. Reifenberger J, Reifenberger G, Liu L, et al. Molecular genetic analysis of oligodendroglial tumors shows preferential allelic deletions on 19q and 1p. *Am J Pathol.* 1994;145(5):1175–1190.

37. Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science.* 2008;321(5897):1807–1812.

38. Barthel FP, Johnson KC, Wesseling P, Verhaak RGW. Evolving insights into the molecular neuropathology of diffuse gliomas in adults. *Neurol Clin.* 2018;36(3):421–437.

39. van den Bent MJ, Chang SM. Grade II and III oligodendroglioma and astrocytoma. *Neurol Clin.* 2018;36(3):467–484.

40. Johnson DR, Guerin JB, Giannini C, et al. 2016 updates to the WHO brain tumor classification system: What the radiologist needs to know. *Radiographics.* 2017;37(7):2164–2180.

41. Martinez-Lage M, Sahm F. Practical implications of the updated WHO classification of brain tumors. *Semin Neurol.* 38.; 2018:11–18.

42. Zulch KJ. Histological typing of tumours of the central nervous system. *International histological classification of tumours No 21.* 1979:43–62.

43. Nakazato Y. Revised WHO classification of brain tumours. *Brain Nerve.* 2008;60(1):59–77.

44. Gerard IJ, Kersten-Oertel M, Petrecca K, et al. Brain shift in neuronavigation of brain tumors: A review. *Med Image Anal.* 2017;35:403–420.

45. Nakasu S, Nakasu Y. Malignant progression of diffuse low-grade gliomas: A systematic review and meta-analysis on incidence and related factors. *Neurol Med Chir (Tokyo).* 2022;62(4):177–185.

46. Bogdańska MU, Bodnar M, Piotrowska MJ, et al. A mathematical model describes the malignant transformation of low grade gliomas: prognostic implications. *PLoS One.* 2017;12(8):e0179999.

47. Lima GL de O, Zanello M, Mandonnet E, et al. Incidental diffuse low-grade gliomas: From early detection to preventive neuro-oncological surgery. *Neurosurg Rev.* 2016;39(3):377–384.

48. Bready D, Placantonakis DG. Molecular pathogenesis of low-grade glioma. *Neurosurg Clin N Am.* 2019;30(1):17–25.

49. Jalbert LE, Neill E, Phillips JJ, et al. Magnetic resonance analysis of malignant transformation in recurrent glioma. *Neuro Oncol.* 2016;18(8):1169–1179.

50. Hormuth DA, Weis JA, Barnes SL, et al. A mechanically coupled reaction–diffusion model that incorporates intra-tumoural heterogeneity to predict in vivo glioma growth. *J R Soc Interface.* 2017;14(128):20161010.

51. Urcun S, Baroli D, Rohan PY, et al. Non-operable glioblastoma: Proposition of patient-specific forecasting by image-informed poromechanical model. *Brain Multiphysics.* 2023;4:100067.

52. Jarrett AM, Hormuth DA, Barnes SL, et al. Incorporating drug delivery into an imaging-driven, mechanics-coupled reaction diffusion model for predicting the response of breast cancer to neoadjuvant chemotherapy: Theory and preliminary clinical results. *Phys Med Biol.* 2018;63(10):105015.

53. Ignjatović J, Stojanov D, Živković V, et al. Apparent diffusion coefficient in the evaluation of cerebral gliomas malignancy. *Vojnosanit Pregl.* 2015;72(10):870–875.

54. Wong AMC, Siow TY, Wei KC, et al. Prediction of malignant transformation of WHO II astrocytoma using mathematical models incorporating apparent diffusion coefficient and contrast enhancement. *Front Oncol.* 2021;11(Sep 29):744827.

55. Danchaivijitr N, Waldman AD, Tozer DJ, et al. Low-grade gliomas: Do changes in rCBV measurements at longitudinal perfusion-weighted MR imaging predict malignant transformation? *Radiology.* 2008;247(1):170–178.

56. Upadhyay N, Waldman AD. Low-grade gliomas: role of relative cerebral blood volume in malignant transformation. In: M.A. Hayat, ed. *Tumors of the Central Nervous System, Volume 2: Gliomas: Glioblastoma (Part 2).* New Jersey: Kean University, 2011:417–423.

57. Bobek-Billewicz B, Stasik-Pres G, Hebda A, et al. Anaplastic transformation of low-grade gliomas (WHO II) on magnetic resonance imaging. *Folia Neuropathol.* 2014;52(2):128–140.

58. Islam M, Wijethilake N, Ren H. Glioblastoma multiforme prognosis: MRI missing modality generation, segmentation and radiogenomic survival prediction. *Comput Med Imag Graph.* 2021;91:101906.

59. Azad R, Khosravi N, Dehghanmanshadi M, Cohen-Adad J, Merhof D. Medical image segmentation on mri images with missing modalities: A review. *arXiv preprint arXiv:220306217.* 2022.

60. Khalid F, Goya-Outi J, Escobar T, et al. Multimodal MRI radiomic models to predict genomic mutations in diffuse intrinsic pontine glioma with missing imaging modalities. *Front Med (Lausanne).* 2023;10:1071447.

61. Varrette S, Bouvry P, Cartiaux H, Georgatos F. Management of an academic HPC cluster: the UL experience. In: *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014).* Bologna, Italy: IEEE; 2014:959–967.