scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Glioblastoma MRI Dataset with Standardized Preprocessing, Expert-Validated Segmentation, and MGMT Profiling

Elena Filimonova *et al.*[#]

Glioblastoma research increasingly relies on large, well-curated imaging datasets that combine standardized MRI data, accurate tumor segmentations, and molecular profiling. We constructed a multi-center dataset of preoperative MRI scans from 337 patients with histologically confirmed primary glioblastoma collected across eight hospitals. All cases include T1-weighted (pre- and post-contrast), T2-weighted, and FLAIR sequences. Images underwent systematic quality assessment, BIDS organization, defacing, skull stripping, and linear registration to the MNI152 template. Tumor segmentation was performed using a SegResNet CNN model following the BraTS labeling convention, with all masks reviewed and manually refined by neuroradiologists. MGMT promoter methylation status was determined for all patients. This dataset provides a robust, clinically representative resource for radiomics, deep learning, and radiogenomic research in glioblastoma, supporting concrete downstream tasks including automated segmentation benchmarking (mean Dice = 0.94) and MGMT methylation prediction (baseline ACC = 0.60). Its multi-center origin, comprehensive preprocessing, expert-refined segmentations, and complete MGMT annotations address limitations of existing datasets and support the development and validation of reproducible imaging biomarkers.

Background & Summary

Glioblastoma (GBM) remains the most frequent and aggressive primary malignant brain tumor in adults, characterized by extensive heterogeneity in morphology and molecular profile[1]. Magnetic resonance imaging (MRI) remains a crucial part of GBM clinical diagnosis, treatment planning, and treatment response assessment[2,3]. On the other hand, molecular biomarkers have also become essential for diagnosis, prognosis, and therapeutic decision-making[4–6]. Integrating multiparametric MRI with molecular markers, such as O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation, enables the creation of non-invasive "virtual biopsy" tools that predict molecular features directly from radiological data[3,7,8]. For instance, MGMT promoter methylation status holds a crucial role in the therapeutic/prognostic framework of GBM, as its methylation is linked to both improved overall survival and therapeutic response to temozolomide[4,5,9,10]. With the growing application of machine learning and radiogenomics, there is an increasing need for large, well-curated, multimodal datasets that combine raw MRI data, molecular data, standardized preprocessing, and ground-truth tumor segmentations.

Over the past decade, several open-source datasets have played a central role in advancing research in this field. The Brain Tumor Segmentation (BraTS) Challenge dataset[2] provides expertly annotated pre-operative MRI scans with standardized segmentation labels, and has become the primary benchmark for algorithm development. Similarly, the UPenn/ The Cancer Imaging Archive (TCIA) GBM collections offer curated, multi-sequence MRIs with varying degrees of clinical annotation[11]. These resources have significantly advanced automated tumor segmentation, radiomic modeling, and deep learning research.

[#]A full list of authors and their affiliations appears at the end of the paper.

Institution	Number of included cases	Main MRI Systems*
AOU Policlinic of Bari, Bari, Italy	3	1.5 T Philips Achieva 1.5 T GE Signa HDxt
IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy	172	3 T Siemens Skyra 3 T GE Signa HDxt 1.5 T Siemens Magnetom Aera 1.5 T Philips Achieva
Policlinico "Riuniti" Foggia, Foggia, Italy	9	3 T Siemens Skyra 1.5 T Philips Achieva dStream
Hospital Joan XXIII, Tarragona, Spain	12	1.5 T GE Signa 1.5 T Philips Ingenia
Städtisches Klinikum Karlsruhe, Karlsruhe, Germany	80	3 T Siemens Magnetom Vida 1.5 T Siemens Magnetom Avanto
Department of Neurosurgery, University of Turin, Turin, Italy	14	3 T Philips Ingenia
Neurosurgery Unit, University Hospital "G. Martino", Messina, Italy	7	1.5 T Philips Ingenia
Federal Neurosurgery Center, Novosibirsk, Russia	41	3 T Philips Ingenia Elition

Table 1. Summary of the glioblastoma cohort. *The full list of MRI systems used to acquire the dataset is provided in the "participants.csv" file on OpenNeuro.

However, despite their impact, the literature highlights several limitations of currently available open-source datasets[12]. First, clinical diversity remains limited, as many datasets originate from a relatively small number of academic centers with similar imaging protocols, potentially restricting generalizability to broader clinical practice. Second, heterogeneity in acquisition parameters, scanner vendors, contrast timing, and spatial resolution is often underrepresented relative to real-world clinical populations. Third, molecular data, such as MGMT promoter methylation, are incomplete or unavailable for a substantial proportion of cases, limiting the potential for radiogenomic studies. These gaps motivate the continued development of large, high-quality datasets that more accurately reflect molecular and clinical variability.

To address these limitations, we introduce a multi-institutional glioblastoma dataset comprising 337 subjects collected from eight hospitals. All cases include T1-weighted (pre- and post-contrast), T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) sequences, each subjected to a unified preprocessing workflow. Tumor segmentation was performed using a SegResNet convolutional neural network (CNN) trained on BraTS conventions, followed by expert visual review and manual correction when necessary. The resulting dataset provides both raw MRI data and segmentation masks, spatially aligned in MNI152 space and ready for radiomic or computer vision analyses. Importantly, MGMT promoter methylation status, determined via methylation-specific PCR for all cases, provides a molecular annotation for multimodal investigations linking imaging phenotypes with molecular features.

This dataset aims to support reproducible research in GBM imaging, including algorithm benchmarking, radiogenomic correlation studies, and the development of clinically interpretable deep learning models. By combining standardized preprocessing, expert quality control, and multimodal data availability, it offers a valuable reference resource for both methodological development and translational neuro-oncology research.

Methods

Participants and study design. This multi-center retrospective study was conducted between April 2024 and November 2025 as part of an ongoing neuro-oncology research initiative aimed at creating a standardized MRI dataset of primary GBM suitable for radiomics, automated tumor segmentation, and radiogenomic analyses. The final cohort included 337 patients with a histopathologically confirmed diagnosis of isocitrate dehydrogenase (IDH)-wildtype GBM who underwent preoperative MRI at eight collaborating hospitals. Formal ethical committee approval for anonymized data collection and publication was obtained from the University of Foggia and all participating centers (46-CE-2023). A summary of the cohort is presented in Table 1.

Patients were included if they met the following criteria: (1) availability of a complete preoperative MRI protocol containing T1-weighted images before and after contrast enhancement, T2-weighted images, and FLAIR; (2) diagnosis of primary, IDH wildtype GBM according to WHO CNS5 criteria[1]; (3) adequate image quality without severe artifacts; and (4) availability of MGMT promoter methylation status, determined via methylation-specific PCR (MSP). Exclusion criteria included incomplete MRI protocols, images acquired after biopsy or surgical intervention, severe artifacts impairing tumor delineation, or missing molecular data.

To ensure clinical representativeness, no restrictions were applied regarding age, sex, tumor location, or symptom severity at presentation. The participating hospitals used diverse MRI scanners and acquisition protocols, reflecting real-world clinical variability and providing a heterogeneous dataset valuable for algorithm development and validation. All imaging data were anonymized at each site before transfer and processing. The average time between MRI scan acquisition and tissue sampling (either biopsy or tumor resection) was 14.3 days (range 3–20 days; SD 5.58 days).

The study complied with all relevant regulations and with the principles of the Declaration of Helsinki. All resources generated and utilized in this study are publicly accessible to support transparency and reproducibility. The methods and tools used in this study are described in technical detail to facilitate the validation and reproducibility of the results presented. All included raw and preprocessed MRI data were converted into the Brain

Imaging Data Structure BIDS (https://bids.neuroimaging.io/) using dcm2bids software (https://github.com/UNFmontreal/Dcm2Bids). To ensure patient privacy, all raw series were defaced using PyDeface (https://github.com/poldracklab/pydeface). Tumor segmentation was performed via inference using the pretrained nnUNet model weights originally developed for brain metastasis screening by Luu *et al.*[13].

MRI acquisition. MRI data were collected across eight collaborating institutions using both 1.5 T and 3 T scanners from major vendors (Siemens, Philips, and GE Healthcare), reflecting the heterogeneity of real-world clinical imaging practice. All examinations were performed as part of routine preoperative evaluation for newly diagnosed GBM. Each patient underwent a standard multiparametric brain MRI protocol that included the following sequences: 1) T1-weighted (T1-WI) pre-contrast; 2) T1-weighted post-contrast (T1-CE) acquired after intravenous gadolinium administration; 3) T2-weighted (T2-WI); 4) FLAIR sequence. These sequences represent the widely accepted minimal diagnostic set for GBM assessment and are essential for reliable tumor segmentation using deep learning models.

Despite variations in vendor-specific implementations, all imaging protocols adhered to typical clinical ranges for echo time (TE), repetition time (TR), flip angle, voxel size, and field of view (FOV), ensuring comparability across sites. Slice thickness typically ranged from 1.0–5.0 mm, depending on local clinical standards. A summary of the acquisition parameters for each sequence and scanner type is presented in Table 2. An example of raw MRI scans from the three contributing centers with the highest number of cases is shown in Fig. 1.

MRI preprocessing. All raw MRI scans were first evaluated by an experienced neuroradiologist to identify intraaxial tumors and ensure concordance between imaging findings and histopathological diagnosis. In the second step, structural images, including T1-weighted (pre- and post-contrast), T2-weighted, and FLAIR series, were visually inspected to exclude scans with prominent artifacts such as motion-related blurring, distortions from dental hardware, incomplete brain coverage, or severe noise. Only cases with all required sequences of acceptable diagnostic quality proceeded to the next stage.

For each subject, T1-weighted, T2-weighted, and FLAIR images were linearly registered to the MNI152 template (https://github.com/Jfortin1/MNITemplate) using 3D Slicer v5.6.2 (https://www.slicer.org) to achieve a standardized spatial orientation. Brain extraction was then performed using the FSL Brain Extraction Tool – BET to remove non-brain tissues[14]. The accuracy of spatial registration and brain extraction was verified by an experienced neuroradiologist. Quantitative registration quality control was also performed and the results are provided in Supplementary File 1 (Supplementary Tables S1 and S2). Subsequently, N4 bias field correction was applied using AFNI software (https://afni.nimh.nih.gov/), followed by z-score intensity normalization using FSL (https://fsl.fmrib.ox.ac.uk/fsl/docs/).

All raw and preprocessed MRI data were converted to the Brain Imaging Data Structure BIDS (https://bids.neuroimaging.io/) using dcm2bids software (https://github.com/UNFmontreal/Dcm2Bids). To ensure patient privacy, all raw series were defaced using PyDeface (https://github.com/poldracklab/pydeface).

Tumor segmentation. Tumor segmentation was performed using a deep learning–based pipeline[15]. For each subject, the tumor was delineated using a SegResNet CNN trained according to the BraTS labeling convention[16]. Under this convention, GBM comprises three biologically and radiologically distinct components: (1) enhancing tumor – regions demonstrating contrast enhancement on post-contrast T1-weighted images; (2) non-enhancing/edematous tumor – hyperintense regions on FLAIR that exclude both the enhancing component and the necrotic core; (3) necrotic core – centrally located, non-enhancing areas that appear hypointense on post-contrast T1-weighted images.

The segmentation pipeline operated on skull-stripped and MNI152-registered T1-WI, T1-CE, T2-WI, and FLAIR images. All automated segmentations underwent visual quality control by experienced neuroradiologists. When algorithmic outputs did not fully correspond to visible tumor margins – for example, in cases with atypical enhancement patterns or hemorrhage - manual corrections were performed in 3D Slicer to ensure anatomical accuracy. Agreement between automatically generated SegResNet tumor masks and final expert-validated tumor masks was assessed using the Dice similarity coefficient and HD95 metrics (Supplementary Tables S5 and S6). The final segmentation masks, after quality control and any required manual refinement, are included in the dataset alongside raw and preprocessed MRI images. Figure 2 illustrates an example of tumor segmentation.

Tumor segmentation was performed via inference using the pretrained nnUNet model weights originally developed for brain metastasis screening by Luu *et al.*[13]. The code execution scripts are available via the public repository at https://github.com/luumsk/BrainMetaSeg.

MGMT methylation analysis. Molecular characterization of all included cases was performed using MSP to determine MGMT promoter methylation status, a key biomarker associated with temozolomide responsiveness and overall prognosis in glioblastoma[8,17,18].

For each patient, tumor tissue samples were obtained during surgical resection or stereotactic biopsy. Genomic DNA was isolated from tumor specimens and MSP analysis was performed according to local laboratory protocols. All participating laboratories followed validated protocols that adhered to national quality assurance standards. Only samples with definitive MGMT methylation assessment were included in the dataset.

MGMT methylation status was subsequently linked to each patient's imaging data within the BIDS structure, enabling integrated radiogenomic analyses and supporting future studies investigating imaging correlates of tumor biology.

Site	Scanner Model	Field Strength	Sequence	2D / 3D	TR	TE	Slice Thickness	Matrix	FOV
AOU Policlinic of Bari, Bari, Italy	Philips Achieva	1.5	T1-WI	2D ax	630	12.5	5	336*336	240*240
			T1-CE	3D	25	4.6	1	256*256	256*256
			FLAIR	3D	4800	280	1	256*256	256*256
			T2-WI	3D	2800	263	1	256*256	256*256
IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy	Siemens Skyra	3	T1-WI	3D	600	7.2	1	256*256	256*256
			T1-CE	3D	600	7.2	1	256*256	256*256
			FLAIR	3D	5000	428	1	256*256	256*256
			T2-WI	2D ax	4000	110	4	512*448	240*210
	Siemens Magnetom Aera	1.5	T1-WI	3D	700	11	1	256*256	256*256
			T1-CE	3D	700	11	1	256*256	256*256
			FLAIR	3D	5000	335	1	256*256	256*256
			T2-WI	2D ax	4000	98	4	320*250	240*188
Policlinico "Riuniti" Foggia, Foggia, Italy	Philips Achieva dStream	1.5	T1-WI	2D ax	220	2.4	5	640*640	230*230
			T1-CE	3D	7.5	3.4	1	288*288	256*256
			FLAIR	3D	4800	290	1.14	240*240	250*250
			T2-WI	2D ax	5670	100	5	560*560	230*230
	Siemens Skyra	3	T1-WI	3D	2300	2.28	0.9	256*256	240*240
			T1-CE	3D	18	3.7	0.9	256*256	240*240
			FLAIR	3D	5000	394	1	512*480	240*225
			T2-WI	2D cor	8000	105	3	384*384	220*220
Hospital Joan XXIII, Tarragona, Spain	GE Signa	1.5	T1-WI	2D ax	560	13	5	320*224	220*175
			T1-CE	3D	12.7	5.3	1	256*256	256*256
			FLAIR	2D ax	8000	128	5	320*256	220*220
			T2-WI	2D ax	5180	104	5	512*512	220*220
Städtisches Klinikum Karlsruhe, Karlsruhe, Germany	Siemens Magnetom Vida	3	T1-WI	3D	2300	2.98	1.1	256*248	256*248
			T1-CE	3D	2300	2.98	1.1	256*248	256*248
			FLAIR	3D	7000	384	1	256*256	256*256
			T2-WI	2D ax	4500	80	4	512*291	230*186
	Siemens Magnetom Avanto	1.5	T1-WI	2D ax	607	11	5	256*208	230*186
			T1-CE	2D ax	750	17	5	256*192	230*172.5
			FLAIR	2D cor	8500	117	6	320*256	250*184
			T2-WI	2D sag	9320	112	3	320*320	240*240
Department of Neurosurgery, University of Turin, Turin, Italy	Philips Ingenia	3	T1-WI	2D ax	2000	20	4	640*640	230*230
			T1-CE	3D	6.2	2.7	1	256*256	240*240
			FLAIR	3D	4800	300	1.2	384*384	230*230
			T2-WI	2D cor	4170	100	4	432*432	230*230
Neurosurgery Unit, University Hospital "G. Martino", Messina, Italy	Philips Ingenia	1.5	T1-WI	2D ax	634	15	5	672*672	227*227
			T1-CE	3D	25	4.6	1	288*288	230*230
			FLAIR	3D	5200	291	1.6	336*336	229*229
			T2-WI	2D ax	4436	100	5	672*672	220*220
Federal Neurosurgery Center (Novosibirsk, Russia)	Philips Ingenia Elition	3	T1-WI	3D	6.7	3	1	512*512	240*240
			T1-CE	3D	6.7	3	1	512*512	240*240
			FLAIR	3D	4800	340	1.2	240*240	240*240
			T2-WI	3D	3500	300	1	252*252	250*250

Table 2. Overview of MRI acquisition parameters across participating centers. Abbreviations: 2D ax - two-dimensional acquisition, axial plane; 2D cor - two-dimensional acquisition, coronal plane; 2D sag – two-dimensional acquisition, sagittal plane; 3D – three-dimensional acquisition; FLAIR – fluid-attenuated inversion recovery; FOV – field of view; GE – General Electric; T1-CE - T1-weighted images after contrast injection; T1-WI – T1-weighted images before contrast injection; T2-WI – T2-weighted images; TE – echo time; TR – repetition time.

Data Records

The dataset consists of BIDS-organized MRI data and corresponding tumor segmentation masks for 337 patients with histologically confirmed IDH-wildtype GBM. It is freely available on the open-access OpenNeuro platform in BIDS format (https://doi.org/10.18112/openneuro.ds007045.v1.0.1) and includes all required components[19].

Each participant directory (sub-***) contains the "anat" folder with defaced structural images (pre- and post-contrast T1-weighted, T2-weighted, and FLAIR sequences), organized according to the BIDS specification.

All derived data are stored under the derivatives/ directory:

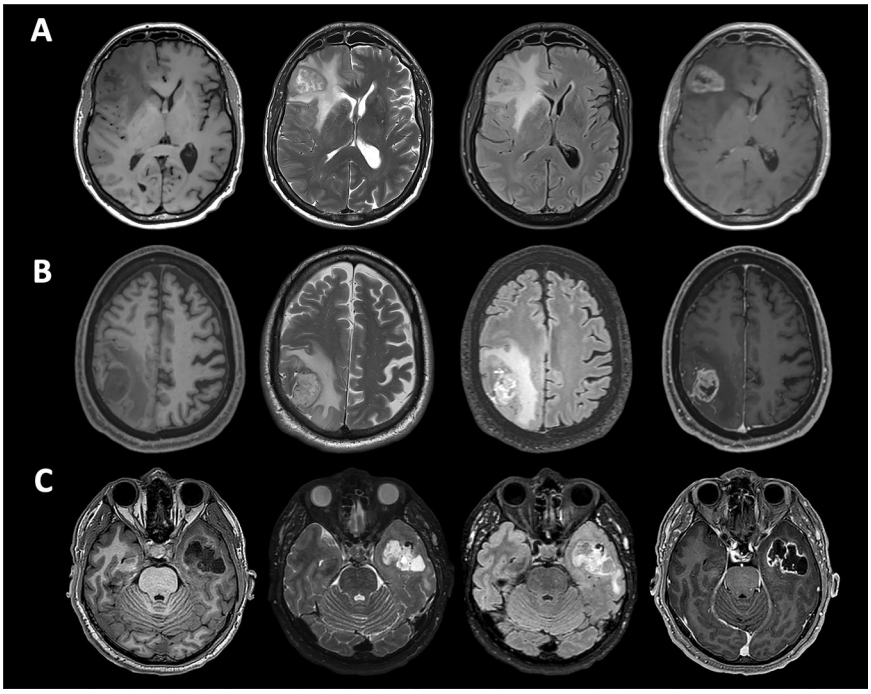

Fig. 1 An example of raw MRI data from the three centers with the highest number of cases. (**A**) – IRCCS Istituto delle Scienze Neurologiche di Bologna, Italy; (**B**) – Städtisches Klinikum Karlsruhe, Karlsruhe, Germany; (**C**) – Novosibirsk Federal Neurosurgery Center, Russia.

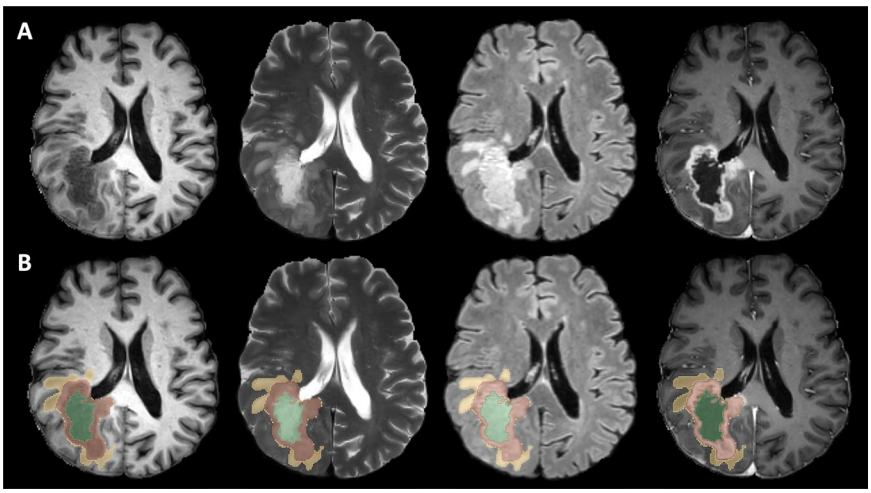

Fig. 2 An example of automatic tumor segmentation using the SegResNet CNN model. Panel A shows co-registered and skull-stripped axial T1-WI, T2-WI, FLAIR, and T1-CE images in MNI152 space (from left to right); panel B shows the corresponding segmentation output. Color coding: green – necrotic core; red – contrast-enhancing tumor; yellow - non-enhancing/edematous tumor.

- derivatives/processing/sub-*** includes preprocessed structural images, such as skull-stripped T1-WI, T2-WI, and FLAIR scans registered to the MNI152 standard space.
- derivatives/processing_MNI152_skull-stripped_N4/sub-*** contains preprocessed structural MRI data (all modalities), skull-stripped and registered to the MNI152 template, with intensity non-uniformity correction performed using N4 bias field correction (AFNI).
- derivatives/processing_MNI152_skull-stripped_N4_z-scored/sub-*** includes preprocessed structural MRI data (all modalities), skull-stripped and registered to the MNI152 template, with N4 bias field correction (AFNI) followed by z-score intensity normalization (FSL). This directory represents the fully normalized, radiomics-ready variant of the dataset.
- derivatives/segmentation/sub-*** contains the tumor segmentation masks, stored as single NIfTI (.nii.gz) files spatially aligned to the corresponding preprocessed MRI volumes in MNI152 space. Each voxel in the

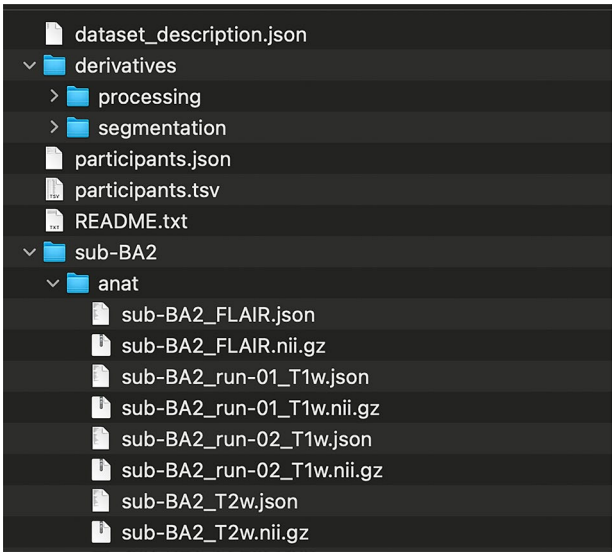

Fig. 3 Structure of the dataset shown for one participant.

mask file contains an integer label following the BraTS convention: 0 – background/non-tumor tissue, 1 – necrotic core, 2 – non-enhancing/edematous tumor, and 3 – contrast-enhancing tumor. These data can be used directly for: volumetric subregion analysis (enhancing tumor, necrotic core, edematous tumor) relevant to surgical planning; segmentation benchmarking against the SegResNet baseline (mean Dice = 0.94); MGMT promoter methylation prediction using radiomic or deep learning approaches; and cross-site generalization studies exploiting the multi-vendor, multi-field-strength acquisition diversity.

MGMT promoter methylation status is provided in the "participants.tsv" file. Further details about the dataset can be found in the "README.txt" file. Figure 3 shows the data structure.

Technical Validation

Quality control procedures were implemented at each stage of dataset construction. The validation pipeline included (1) visual quality assessment of raw MRI data, (2) inspection of skull stripping and spatial normalization outputs, and (3) visual review of tumor segmentations with manual corrections when required.

All raw MRI data were reviewed by experienced neuroradiologists. Each sequence was assessed for motion artifacts, distortions caused by dental implants, incomplete brain coverage, and severe noise. Scans that failed quality checks were excluded from further processing.

After skull stripping and spatial normalization to the MNI152 template, all preprocessed images were visually inspected to confirm correct alignment across modalities and accurate removal of extracranial tissues without loss of tumor margins. Scans showing misregistration or inadequate skull stripping were reprocessed. Overall, 19 cases (6%) required reprocessing, and all issues were resolved before segmentation. Quantitative registration quality metrics are provided in Supplementary File 1 (Supplementary Tables S1 and S2).

Subsequently, automated tumor segmentations generated by the CNN-based SegResNet model were examined by neuroradiologists with expertise in neuro-oncology imaging. Cases with segmentation inaccuracies were manually edited in 3D Slicer. A total of 87 cases (from 337, 25.8%) required minor corrections (e.g., boundary adjustments), whereas extensive manual segmentation was needed in another 11 cases (3.2%). Quantitative agreement between automated SegResNet tumor masks and final expert-validated masks was assessed using the Dice similarity coefficient and HD95 metrics (Supplementary Tables S5 and S6).

Because the dataset integrates MRI scans from eight hospitals using different scanner models and acquisition protocols, additional checks were implemented for cross-site consistency. Voxel dimensions and orientation were standardized during preprocessing; histograms and intensity distributions were reviewed across sites. Additional quantitative inter-site intensity analyses across preprocessing stages are provided in Supplementary File 1 (Supplementary Tables S3 and S4; Supplementary Figs. S1 and S2). After preprocessing, the images demonstrated improved cross-site consistency and spatial alignment suitable for downstream radiomics and deep learning applications.

Ethical statement. This research was conducted in compliance with the Declaration of Helsinki. Formal ethical committee approval for anonymized data collection and publication was obtained from the University of Foggia and all participating centers (46-CE-2023). Further ethical approval was required from the University of Bologna and was received prior to data collection (CE24128). Both ethical committees approved the study and the sharing of data.

Data availability

The complete dataset is available at: (https://doi.org/10.18112/openneuro.ds007045.v1.0.1).

Code availability

The tumor segmentation pipeline was implemented using in-house code, which is publicly available at GitHub (https://github.com/luumsk/BrainMetaSeg).

Received: 11 January 2026; Accepted: 20 July 2026;
Published online: 07 August 2026

References

1. Louis, D. N. *et al*. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* **23**, 1231–1251 (2021).
2. Menze, B. H. *et al*. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2015).
3. Singh, G. *et al*. Radiomics and radiogenomics in gliomas: a contemporary update. *Br J Cancer* **125**, 641–657 (2021).
4. Butler, M. *et al*. MGMT Status as a Clinical Biomarker in Glioblastoma. *Trends in Cancer* **6**, 380–391 (2020).
5. Shaw, R., Basu, M., Karmakar, S. & Ghosh, M. K. MGMT in TMZ-based glioma therapy: Multifaceted insights and clinical trial perspectives. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1871**, 119673 (2024).
6. Weller, M., Stupp, R. & Reifenberger, G. MGMT promoter methylation in malignant gliomas: Ready for personalized medicine? *Nature Reviews Neurology* **6**, 39–51 (2010).
7. Le Guillou Horn, X. M. *et al*. From Voxel to Gene: A Scoping Review on MRI Radiogenomics' Artificial Intelligence Predictions in Adult Gliomas and Glioblastomas-The Promise of Virtual Biopsy? *Biomedicines* **12** (2024).
8. Leone, A. *et al*. Virtual Biopsy for the Prediction of MGMT Promoter Methylation in Gliomas: A Comprehensive Review of Radiomics and Deep Learning Approaches Applied to MRI. *Diagnostics* **15**, 251 (2025).
9. Caccese, M. *et al*. Definition of the Prognostic Role of MGMT Promoter Methylation Value by Pyrosequencing in Newly Diagnosed IDH Wild-Type Glioblastoma Patients Treated with Radiochemotherapy: A Large Multicenter Study. *Cancers* **14**, 2425 (2022).
10. Mur, P. *et al*. Impact on prognosis of the regional distribution of MGMT methylation with respect to the CpG island methylator phenotype and age in glioma patients. *J Neurooncol* **122**, 441–450 (2015).
11. Bakas, S. *et al*. The University of Pennsylvania glioblastoma (UPenn-GBM) cohort: advanced MRI, clinical, genomics, & radiomics. *Sci Data* **9**, 453 (2022).
12. Zenk, M. *et al*. Towards fair decentralized benchmarking of healthcare AI algorithms with the Federated Tumor Segmentation (FeTS) challenge. *Nat Commun* **16**, 6274 (2025).
13. Luu, M. S. K. *et al*. Transfer Learning Approaches for Brain Metastases Screenings. *Biomedicines* **12**, 2561 (2024).
14. Smith, S. M. Fast robust automated brain extraction. *Human Brain Mapping* **17**, 143–155 (2002).
15. Myronenko, A. 3D MRI brain tumor segmentation using autoencoderregularization. arXiv. Preprint posted online November 19 https://doi.org/10.48550/arXiv.1810.11654 (2018).
16. Bonato, B., Nanni, L. & Bertoldo, A. Advancing Precision: A Comprehensive Review of MRI Segmentation Datasets from BraTS Challenges (2012–2025). *Sensors (Basel)* **25** (2025).
17. Cheng, J. *et al*. Prediction of Glioma Grade Using Intratumoral and Peritumoral Radiomic Features From Multiparametric MRI Images. *IEEE/ACM Trans. Comput. Biol. and Bioinf.* **19**, 1084–1095 (2022).
18. Le Rhun, E. *et al*. Molecular targeted therapy of glioblastoma. *Cancer Treatment Reviews* **80**, 101896 (2019).
19. Filimonova, E. *et al*. Real-world Multi-Institutional Glioblastoma MRI Dataset with Standardized Preprocessing, Expert-Validated Segmentation, and MGMT Promoter Profiling. *OpenNeuro* https://doi.org/10.18112/openneuro.ds007045.v1.0.1 (2025).

Author contributions

All authors have read and approved the manuscript, including the revisions. The data collection was done by E.F., A.L., F.C., M.Z., A.C., A.B., A.R., F.Co., V.D.N., N.P.F., R.L., A.R., V.I., G.G., and A.Cu. The data analysis was done by E.F., A.L., F.C., J.R., M.S., K.L.M.S., M.B., A.S., and B.T. The manuscript was written mainly by E.F., F.C., and A.L. The project was supervised by E.F., F.C., A.L., U.S., S.R., D.M., D.G., B.T., and A.C.

Funding

This research was supported by a grant for research centers, provided by the Ministry of Economic Development of the Russian Federation in accordance with the subsidy agreement with the Novosibirsk State University dated 17 April 2025 No. 139-15-2025-006: IGK 000000C313925P3S0002

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-026-07953-2.

Correspondence and requests for materials should be addressed to F.C. or B.T.

Elena Filimonova[1,2,22], Augusto Leone[3,22], Francesco Carbone[3,4] ✉, Matteo Zoli[5,6], Jamil Rzaev[1,2], Mariya Schukina[1,2], Alessandro Carretta[5,6], Andrea Bianconi[7,8], Fabio Cofano[9,10], Alberto Morello[9,10], Daniele Armocida[9,10], Uwe Spetzger[3,11], Safwan Roumia[12], Veronica Di Napoli[10], Nicola Pio Fochi[10], Ruth Lau[13,14], Valeria Internò[15], Guido Giordano[16], Antonello Curcio[17], Arianna Rustici[18], Flavio Angileri[19], Diego Mazzatenta[5,6], Francesco Signorelli[20], Camillo Porta[21], Diego Garbossa[9,10], Minh Sao Khue Luu[2], Margaret Benedichuk[2], Ahsan Shakoor[2], Bair Tuchinov[2,23] ✉ & Antonio Colamaria[4,23]

[1]Federal Neurosurgical Center, 132/1 Nemirovicha-Danchenko St, Novosibirsk, 630087, Russia. [2]The Artificial Intelligence Research Center of Novosibirsk State University, 1 Pirogov St, Novosibirsk, 630090, Russia. [3]Department of Neurosurgery, Städtisches Klinikum Karlsruhe, Karlsruher Neurozentrum, Karlsruhe, Germany. [4]Division of Neurosurgery, University Hospital of Foggia, Foggia, Italy. [5]Department of Biomedical and Neuromotor Sciences (DIBINEM), "Alma Mater Studiorum" University of Bologna, Bologna, Italy. [6]IRCCS Istituto delle Scienze Neurologiche di Bologna, Programma Neurochirurgia Ipofisi-Pituitary Unit, Bologna, Italy. [7]Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), University of Genova, 16132, Genova, Italy. [8]Department of Neurosurgery, IRCCS Ospedale Policlinico San Martino, Genova, 16132, Genoa, Italy. [9]Department of Neurosurgery, AOU Città della Salute e della Scienza, Turin, Italy. [10]Department of Neurosurgery, University of Turin, Turin, Italy. [11]Institute for Anthropomatics and Robotics, Karlsruhe Institute for Technology - KIT, Karlsruhe, Germany. [12]Department of Radiology, Division of Neuroradiology, Städtisches Klinikum Karlsruhe, Karlsruhe, Germany. [13]Neurosurgery Department, Hospital Universitari Joan XXIII, Tarragona, Spain. [14]Universitat Rovira i Virgili, Tarragona, Spain. [15]Division of Medical Oncology, A.O.U. Consorziale Policlinico Di Bari, Bari, Italy. [16]Unit of Medical Oncology and Biomolecular Therapy and CREATE Center for Research and Innovation Medicine, Department of Medical and Surgical Sciences, University of Foggia, Policlinico Riuniti, Foggia, Italy. [17]Fondazione IRCCS Casa Sollievo della Sofferenza, UOC di Neurochirurgia, San Giovanni Rotondo, Italy. [18]Neuroradiology Unit, Ospedale Maggiore, IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy. [19]Department of Biomedical Sciences, Dental Sciences and Morpho-functional Imaging, University of Messina, Messina, Italy. [20]Division of Neurosurgery, Department of Translational Biomedicine and Neurosciences, University "Aldo Moro" of Bari, Bari, Italy. [21]Interdisciplinary Department of Medicine, University of Bari "Aldo Moro" and Division of Medical Oncology, A.O.U. Consorziale Policlinico Di Bari, Bari, Italy. [22]These authors contributed equally: Elena Filimonova, Augusto Leone. [23]These authors jointly supervised this work: Bair Tuchinov, Antonio Colamaria. ✉e-mail: francesco.carbone615@gmail.com; bairt@nsu.ru