The Brain Tumor Segmentation (BraTS) Cluster of Challenges: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

Use the title to convey the essential information on the challenge mission.

The Brain Tumor Segmentation (BraTS) Cluster of Challenges

Challenge acronym

Preferable, provide a short acronym of the challenge (if any).

BraTS

Challenge abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

This document describes the experimental design of the 'International Brain Tumor Segmentation (BraTS) Cluster of Challenges', in partnership with the 'A.I. for Response Assessment in Neuro-Oncology' (AI-RANO) cooperative group.

Since 2012, the annual BraTS challenge has focused on the generation of a fair benchmarking environment and an associated dataset for the delineation of adult brain tumors. After 10 years, we conducted the RSNA-ASNR-MICCAI BraTS 2021 challenge, which spearheaded a partnership of MICCAI with two major clinical societies in the US (RSNA & ASNR) and contributed to extending the BraTS dataset to >2,000 cases. Building upon this effort, in 2023, we conducted the first BraTS Cluster of Challenges, with a substantially expanded dataset of >4,500 cases and a scope to address additional i) underserved patient populations, ii) tumor types, and iii) clinical concerns (e.g., missing data).

This year the focus of the BraTS 2024 Cluster of Challenges remains the generation of a common benchmarking environment, but with a further expanded 1) clinical relevance, 2) scope, and 3) dataset. Specifically, the BraTS 2024 Cluster of Challenges partners with the AI-RANO group to present newly proposed clinically relevant challenges, in a synergistic attempt to maximize the potential clinical impact of the innovative algorithmic contributions made by the participating teams. The scope extends further to address additional i) underserved populations (i.e., sub-Saharan Africa patients), ii) timepoints (i.e., pre- & post-treatment), iii) tumor types (e.g., meningioma), iv) modalities (i.e., histology samples), v) clinical concerns (e.g., missing data), and iv) technical considerations (e.g., generalizability). Finally, the BraTS 2024 datasets describes a further contribution to the community of additional well-curated manually-annotated cases, comprising a) MRI scans from 4,000 previously unseen patients & 280,000 histology samples.

The focus of each challenge (referred to as "task" here onwards - based on predefined terminology by the submission platform) of the BraTS 2024 Cluster of Challenges is to identify the current state-of-the-art algorithms for addressing (Task 1) Post-Treatment Adult Glioma, (Task 2) Post-treatment Intracranial Meningioma, (Task 3) Pre- and Post-Treatment Brain Metastases, (Task 4) Brain Glioma in the underserved sub-Saharan African patient population, (Task 5) Pre-Treatment Pediatric Tumor Patients in partnership with multiple related societies, (Task 6) Generalizability of Segmentation Methods Across (Pre-treated) Tumors, (Task 7) Evaluation of Augmentation Techniques, in partnership with FDA, (Task 8) MRI Synthesis, (Task 9) MRI Inpainting, as well as (Task 10) Assessing the Heterogeneous Histologic Landscape of Glioma. Worth highlighting that 6/10 tasks (1-3,6,7,10) are newly introduced tasks.

Details for each 'Task' are listed in the rest of this document. Notably, all data for tasks 1-9 are routine clinically acquired, multi-site multiparametric magnetic resonance imaging (mpMRI) scans of brain tumor patients. The BraTS 2024 challenge participants are able to obtain the training and validation data of the challenge at any point from the Synapse platform. These data will be used to develop, containerize, and evaluate their algorithms in unseen validation data until August 2024, when the organizers will stop accepting new submissions and evaluate the submitted algorithms in the hidden testing data. Ground truth reference annotations for all datasets are created and approved by expert neuroradiologists/neuropathologists for every subject included in the training, validation, and testing datasets to quantitatively evaluate the performance of the participating algorithms.

Due to local space constraints in MICCAI 2024, the BraTS CLuster of Challenges will be split into 2 half-day events. The first half-day event will be identified by the acronym "BraTS" and describe all the segmentation challenges, and the second half-day event will be identified as "Beyond-BraTS" and will describe all the non-segmentation tasks.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

BraTS, Brain Tumor, Neuro Oncology, Radiology, Pathology, Digital Pathology, Pre-treatment, Post-treatment, Segmentation, Classification, Generalizability, Augmentation, Synthesis, Inpainting, Infill, resection cavity, Cancer, Challenge, Glioma, Glioblastoma, Brain metastasis, diffuse glioma, Meningioma, MRI, brain, radiotherapy, contrast enhancing, peritumoral edema, Sub-Saharan Africa, Pediatric, Rare Diseases, Diffuse Midline Glioma, CBTN, Data Centric, Health Disparities, Health Inequities, NIH, MICCAI, NCI, DREAM, RSNA, ASNR, PrecisionFDA, Synapse

Year

The challenge will take place in 2024

FURTHER INFORMATION FOR CONFERENCE ORGANIZERS

Workshop

If the challenge is part of a workshop, please indicate the workshop.

The BraTS 2024 Cluster of Challenge is coordinating publication of its own proceedings with Springer.

Duration

How long does the challenge take?

Full day.

Expected number of participants

Please explain the basis of your estimate (e.g. numbers from previous challenges) and/or provide a list of potential participants and indicate if they have already confirmed their willingness to contribute.

We can conservatively estimate approximately 500 participating teams for this year's cluster of challenges.

This is based on the continuously increasing number of teams participating in the BraTS challenge during its initial 10 years (2012:n=10, 2013: n=10, 2014: n=10, 2015: n=12, 2016: n=19, 2017: n=53, 2018: n=63, 2019: n=72, 2020: n=78, 2021: n>2,300). Notably, we strongly believe that the 2021 participation increase (from 78 to 2,300) is a result of multiple factors (challenge's maturity, involvement of RSNA & ASNR, professional evaluation through Synapse and Kaggle) that has been carried forward since 2021 up to this year's cluster of challenges thereby guaranteeing broad participation. It is difficult to estimate exact numbers of BraTS participation since 2021, as we have been running continuous evaluation (open call) challenges ever since.

We also advertise the event in related mailing lists (e.g., CVML; visionlist@visionscience.com; cvnet@mail.ewind.com; MIPS@LISTSERV.CC.EMORY.EDU), NCI's CBIIT blog posts and tweets, and we intend to send an email to all the above and notify them about this year's challenge.

Finally, since we will specifically focus on assessing generalizability across brain tumors and patient populations (including the Africa-BraTS challenge), we will also advertise the event in ML communities in Africa to strengthen local participation. Communities we will consider include the "Data Science Nigeria" (DSN, https://www.datasciencenigeria.org), the "African Institute of Mathematical Sciences" (https://aims.edu.gh), the "African Centre of Excellence in Data Science" (ACE-DS, https://aceds.ur.ac.rw/), and the "INDABA" (https://deeplearningindaba.com).

Publication and future plans

Please indicate if you plan to coordinate a publication of the challenge results.

We intend to coordinate 2 specific publication plans immediately after the challenge.

Plan 1:
The configuration of coordinating the BraTS proceedings with Springer provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings. We have already been performing this configuration for BraTS, since 2015.

Plan 2:
We will coordinate journal manuscripts focusing on publishing and summarizing the results of each BraTS 2024 challenge, making a comprehensive meta-analysis for each to inform the community about the obtained results, findings, and insights.

Space and hardware requirements

Organizers of on-site challenges must provide a fair computing environment for all participants. For instance, algorithms should run on the same computing platform provided to all.

Hardware requirements for the in-person meeting: 1 projector, 3 microphones, loudspeakers

BraTS 2024 Cluster of Challenges describes off-site challenges, where 1) during the training phase algorithms are trained using the participants' computing infrastructure, and 2) during the validation and final testing/ranking phase using the organizers' infrastructure (i.e., Synapse.org - SAGE Bionetworks).

TASK 1: BraTS-GLIOMA: Segmentation of Post-Treatment Glioma

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Brain tumors are among the deadliest types of cancer. Specifically, glioblastoma, and diffuse astrocytic glioma with molecular features of glioblastoma (WHO Grade 4 astrocytoma), are the most common and aggressive malignant primary tumor of the central nervous system in adults, with extreme intrinsic heterogeneity in appearance, shape, and histology, with a median survival of approximately 12 months. Brain tumors in general are challenging to diagnose, hard to treat and inherently resistant to conventional therapy because of the challenges in delivering drugs to the brain. Years of extensive research to improve diagnosis, characterization, and treatment have decreased mortality rates in the U.S. by 7% over the past 30 years. Although modest, these research innovations have not translated to improvements in survival.

Considering the clinical impact, the BraTS 2024 Glioma challenge will differ from previous years BraTS-Glioma challenge, as it is will 1) utilize a new dataset consisting only of post-treatment diffuse gliomas, and 2) will include a new tissue class consisting of the resection cavity. Notably this cavity is what most algorithms trained on previous BraTS-Glioma pre-operative challenge data are failing. The goal of the BraTS-2024-Glioma challenge will be to identify current state-of-the-art segmentation algorithms for brain diffuse glioma patients and their sub-regions in the post-treatment timepoint, and hence target a more clinically relevant question of monitoring tumor progression. All challenge data are routine clinically-acquired, multi-institutional multiparametric magnetic resonance imaging (mpMRI) scans of brain glioma patients.

These data will be used by participants to develop, containerize, and evaluate their algorithms in unseen validation data until July 2024, when the organizers will stop accepting new submissions and evaluate the submitted algorithms in the hidden testing data. Ground truth reference annotations for all datasets are created and approved by expert neuroradiologists for every subject included in the training, validation, and testing datasets to quantitatively evaluate the performance of the participating algorithms.

Keywords

List the primary keywords that characterize the task.

Segmentation, Brain Tumors, Post-treatment, resection cavity, Cancer, Challenge, Glioma, Glioblastoma, MICCAI, NCI, DREAM, diffuse glioma

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Jeffrey Rudie, MD PhD [Lead Organizer - Contact Person]
University of California San Diego

Maria Correia De Verdier, MD
University of California San Diego

Spyridon Bakas, PhD & Ujjwal Baid, PhD
Indiana University

Raymond Huang, MD PhD
Mass General Hospital

Evan Calabrese, MD PhD
Duke University

Dominic LaBella, MD
Duke University Medical Center

Rachit Saluja, MS
Cornell University

Louis Gagnon, MD PhD
Université Laval

Mariam Aboian, MD PhD
Childrens Hospital of Philadelphia

Aly Abayazeed, MD
Neosoma Inc.

Keyvan Farahani, PhD.
National Institutes of Health

Jake Albrecht, PhD
Sage Bionetworks

Verena Chung
Sage Bionetworks

Clinical Evaluators and Annotation Approvers:
=================================
Jeffrey Rudie, MD, PhD
Maria Correia De Verdier, MD

Evan Calabrese, MD PhD
Andreas Rauschecker, MD PhD (UCSF)

b) Provide information on the primary contact person.

Jeffrey Rudie, MD PhD
[Lead Organizer of the "BraTS-GLIOMA: Segmentation of Post-Treatment Glioma"]
University of California, San Diego, CA, USA
Email: jeff.rudie@gmail.com

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NCI (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data

Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data, but if they do so, they MUST also discuss the potential difference in their results after using only the BraTS 2024 data, since our intention is to solve the particular segmentation problem, but also to provide a fair comparison among the participating methods.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel 2) Neosoma Inc, and 3) Cortechs.ai we have informal confirmation for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge.
Note that Intel has been offering monetary awards during each of BraTS 2018-2023, Neosoma for BraTS 2021-2022, and Cortechs.ai during 2023. NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams’ members qualifies as author

- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the output of their methods to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it in the hidden testing data. The participants will be provided guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility, thereby maximizing solutions in solving the problem of brain tumor segmentation.

During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Tool [7] (https://fets-ai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).

[1] C .Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

[2] S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

[3] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in May after the release of the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We are already in close coordination with The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH), to release the training and validation data following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions). The TCIA has already approved this, and we are now in the process of submission (includes a detailed curation process specific to TCIA). The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Googles AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Informed consent was obtained from all subjects at their respective institutions, and the protocol for releasing the data was approved by the institutional review board of the data-contributing institution.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Platform [7] (https://fets-ai.github.io/Front-End/).

[4] C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

[5] S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

[6] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS 2021 and 2022 challenges.

All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

The National Cancer Institute takes special interest in the BraTS 2023 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel, Neosoma Inc, and Cortechs.ai

Challenge Organizers, SAGE Bionetworks, and the clinical evaluators will have access to the validation, and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients diagnosed with diffuse gliomas of the brain and have already undergone treatment, which may include surgery, radiation, and/or chemotherapy.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with de novo diffuse gliomas of the brain and have already undergone treatment which may include surgery, radiation and/or chemotherapy. They will have been clinically scanned with mpMRI acquisition protocol including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region volumes)

Data Contributors:
==================================

Jeffrey Rudie, MD PhD
University of California San Diego, Department of Radiology

Evan Calabrese MD PhD
Duke University Medical Center, Department of Radiology

Javier Villanueva-Meyer, MD Andreas Rauschecker, MD PhD
University of California San Francisco, Department of Radiology & Biomedical Imaging

Spyridon Bakas, PhD, & Ujjwal Baid, PhD
Indiana University, Department of Pathology & Laboratory Medicine

Adam Flanders, MD
Thomas Jefferson University Medical Center, Department of Radiology

Mariam Aboian ,MD PhD

Yale University Medical Center, Department of Radiology

Ayman Nada, MD
Missouri University Medical Center, Department of Radiology

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Dice, Hausdorff 95th percentile, Sensitivity, Precision, Specificity. - per lesion evaluation

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in our related manuscripts [1,2,4]. Since then, multiple institutions have contributed data to create the BraTS 2024 Glioma dataset and these will be listed in a BraTS arXiv paper following acceptance of the challenge. We are currently in coordination with TCIA to make the complete BraTS 2024 dataset permanently available through their portal. All the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, including Google and AWS Marketplaces, as part of their Public Datasets Programs.

[7] U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314

[8] S.Bakas, 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

[9] Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe MRI scans, acquired with different clinical protocols and various scanners from: UCSF, UCSD, Duke University, Indiana University, Thomas Jefferson University, Yale University, etc.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at a single timepoint. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There is no inclusion/exclusion criterion applied that related to 3d acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

Training data: 1,050 cases
Validation data: 150 cases
Testing data: 300 cases
Note: None of these cases have been used in previous challenges.

We are already coordinating with other institutions that have expressed interest in contributing more data, towards increasing the total number of cases assessed in the challenge.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability.

The data will be split in these numbers between training, validation, and testing based on a standard split (70:10:20) used in machine learning research.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

N/A

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Reference approved from at least 2 experienced neuroradiologists, following initial annotations from other annotation volunteers.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists and/or radiation oncologists), which is provided to all clinical annotators, describing in detail instructions on what the gross tumor volume segmentation should and should not include. The annotators are given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions:
i) Enhancing Tissue (ET): This delineates the hyperintense signal of the T1-Gd, after excluding the vessels. Any areas of thick or nodular enhancement are included in the ET class, though typical treatment related thin linear enhancement along and within resection cavities and along the dura is not included in the ET class.
ii) Nonenhancing tumor core (NETC): The NETC class consists of necrotic/nonenhancing tissue surrounded by ET and not otherwise clearly represented by a prior resection cavity the necrotic core (when present). The tumor core (TC) is the union of the enhancing tumor and the necrotic core described in (i) and (ii) here.
iii) Surrounding nonenhancing FLAIR hyperintensity (SNFH): This tissue typically includes edema and infiltrating tumor. Given the post-treatment nature of the scans, any T2/FLAIR signal abnormalities, including radiation-related hyperintensity, gliosis, edema, and non-enhancing tumor, are included in the SNFH label. The Whole Tumor is the union of ET, NETC and SNFH. .
iv) Resection cavity (RC): The RC class consists of both recent and chronic resection cavities. Chronic resection cavities, which are typically older than 3-6 months, were considered those with signal intensity isointense to CSF. More recent resection cavities often contained air, blood, and/or proteinaceous materials, and otherwise exhibited variable signal characteristics.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case is assigned to a pair of annotator-approver. Annotators span across various experience levels and clinical/academic ranks, while the approvers are 2 board-certified neuroradiologists or neuro-focused radiation oncologists, listed in the Organizers section as clinical evaluators and annotation approvers. The annotators are given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators are satisfied with the produced annotations, these are passed to an approver. The approver is then responsible for signing off on these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scan, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach is followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The exact preprocessing pipeline applied to all the data considered in the BraTS 2024 challenge is identical with the one evaluated and followed by the BraTS 2017-2022 challenges. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format [10], we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 [9]) and interpolating to the same resolution as this atlas (1 mm^3). The exact registration process comprises the following steps:

STEP 1: N4 Bias field correction (notably the application of N4 bias field correction is a temporary step. Taking into consideration we have previously [4] shown that use of non-parametric, non-uniform intensity normalization (i.e., N4) to correct for intensity non-uniformities caused by the inhomogeneity of the scanners magnetic field during image acquisition obliterates the MRI signal relating to the abnormal/tumor regions, we intentionally use N4 bias field correction in the preprocessing pipeline to facilitate a more optimal rigid registration across the difference MRI sequences. However, after obtaining the related information (i.e., transformation matrices), we discard the bias field corrected scans, and we apply this transformation matrix towards the final co-registered output images used in the challenge).

STEP 2: Rigid Registration of T1, T2, T2-FLAIR to the T1-Gd scan, and obtain the corresponding transformation matrix.

STEP 3: Rigid Registration of T1-Gd scan to the SRI-24 atlas [9], and obtain the corresponding transformation matrix.

STEP 4: Join the obtained transformation matrices and applying aggregated transformation to the LPS-oriented scans.

STEP 5: After completion of the registration process, we perform brain extraction to remove any apparent nonbrain tissue (e.g., neck fat, skull, eyeballs) based on a deep-learning approach we developed in house, focusing on scans with apparent brain tumors and exhaustively evaluated it in both private and public multi-institutional data [11]. We then manually assessed all scans for confirming the correct brain extraction (i.e., skull stripping), where the complete brain region is included, and all non-brain tissue is excluded.

This whole pipeline, and its source code are available through the CaPTk and FeTS ttps://github.com/CBICA/CaPTk) (https://fets-ai.github.io/Front-End/) platforms)

[9] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117, 2017. DOI: 10.1038/sdata.2017.117

[10] T. Rohlfing, et al. The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798-819, 2010.

[11] R.Cox, J.Ashburner, H.Breman, K.Fissell, C.Haselgrove, C.Holmes, J.Lancaster, D.Rex, S.Smith, J.Woodward, A (Sort of) new image data format standard: NIfTI-1: WE 150, Neuroimage, 22, 2004.

[12] S.Thakur, J.Doshi, S.Pati, S.Rathore, C.Sako, M.Bilello, S.M.Ha, G.Shukla, A.Flanders, A.Kotrotsou, M.Milchenko, S.Liem, G.S.Alexander, J.Lombardo, J.D.Palmer, P.LaMontagne, A.Nazeri, S.Talbar, U.Kulkarni, D.Marcus, R.Colen, C.Davatzikos, G.Erus, S.Bakas, 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. DOI: 10.1016/j.neuroimage.2020.117081

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [8] and is outside the scope of the BraTS 2024 Glioma challenge.

[13] R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise
Sensitivity, Lesionwise
Specificity, Lesionwise
Precision, Lesionwise
The regions evaluated using these metrics describe the aggregate of abnormal signal (whole tumor plus resection cavity), the tumor core (ET and NETC (when present)) . Note that the tumor core includes the part of the tumor that is typically resected, and the whole tumor describes all tumor sub-regions (i.e., tumor core and SNFH).

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of the assessed and evaluated tumor sub-regions:
i) the enhancing tissue describes the regions of active tumor as well as potentially active tumor in the post-treatment setting, which in linical practice characterizes the areas to longitudinall assess and potentially

re-resect. ii) the tumor core (incl. the NETC) also what is typically resected during a surgical procedure. iii) the whole tumor as it defines the whole extent of the tumor, including the SNFH.

In terms of evaluation metrics, we use:
i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance, ii) the 95% Hausdorff distance as opposed to standard HD, which are computed on a lesionwise bases per BraTS 2023 iii) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or undersegment. iv) Precision to complement the metric of Sensitivity (also known as recall).

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics at Indiana University School of Medicine), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (0 for the DSC and the image diagonal for the HD).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot [6].

[14] Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2023, uncertainties in rankings will be assessed using permutational analyses [3]. Performance for the segmentation task will be assessed based on relative performance of each team on each tumor tissue class and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

[2] S. Bakas et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, arXiv:1811.02629 [cs, stat], Apr. 2019, Accessed: Dec. 10, 2020. [Online]. Available: http://arxiv.org/abs/1811.02629.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 2: BraTS-Meningioma: Segmentation of Post-treatment Intracranial Meningioma

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Meningioma is the most common primary intracranial tumor and can result in significant morbidity and mortality for affected patients. Most meningiomas are benign (approximately 80%) and are typically well controlled with surgical resection and/or radiation therapy. However, higher grade meningiomas (World Health Organization [WHO] grades 2 and 3) are associated with significantly higher morbidity and mortality rates and often recur despite optimal management. Essentially all grade 3 and many grade 2 meningiomas will be treated with radiation therapy, either as a primary treatment modality, as an adjunct in the immediate postoperative setting, and/or in the setting of meningioma recurrence. Accurate segmentation of the gross tumor volume (GTV), the portion of the tumor visible on postcontrast MRI, is essential for radiation therapy planning. Unfortunately, GTV segmentation is complex, time-consuming, and requires considerable expertise. To date, no reliable automated methods for meningioma GTV segmentation exist.

Automated tumor segmentation on brain MRI has matured into a clinically viable tool that can provide objective assessments of tumor volume and can assist in surgical planning, radiotherapy, and treatment response assessment. However, to date most tumor segmentation studies, including all prior BraTS challenges, have focused exclusively on preoperative tumors, which limits clinical utility. Segmenting postoperative/posttreatment tumors is a considerably more complex challenge but is also considerably more clinically relevant.

The purpose of the BraTS 2024 Meningioma challenge is to create a community benchmark for automated segmentation of meningioma GTV based on pre-radiation therapy planning brain MRI exams. This task, if successful, will provide an important tool for the objective delineation of meningioma GTV, which will be immediately relevant for radiotherapy planning. In addition, this algorithm will provide a starting point for future studies focused on distinguishing residual/recurrent meningioma from post-treatment changes and predicting risk of progression and future recurrence.

The BraTS 2024 Meningioma challenge will feature several important changes compared to prior BraTS challenges, even to the BraTS 2023 Meningioma challenge that focused on pre-operative cases. All these changes are implemented to substantially increase the clinical relevance. First, challenge data will consist exclusively of radiotherapy planning brain MRI exams. In addition, image data will consist of 1) a single series (3D postcontrast T1-weighted spoiled gradient echo imaging) and 2) in native acquisition space, which mimics the data available for most radiotherapy planning scenarios, rather than 4 MRI scans co-registered to a canonical atlas space. Furthermore, previous BraTS challenges have utilized skull-stripping, whereas here we will preserve extracranial structures and instead use automated face randomization algorithms to preserve patient anonymity (i.e., defacing). Finally, target labels will consist of a single tumor region (the GTV) in the native acquisition space. To reduce barriers to participation, we will provide software to automatically convert challenge data to a canonical reference space similar to previous BraTS challenges, as well as to convert resulting labels back to patient native space for subsequent evaluation.

Keywords

List the primary keywords that characterize the task.

Meningioma, brain tumor, MRI, brain, neuro-oncology, radiotherapy, segmentation, MICCAI

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Evan Calabrese MD PhD [Lead Organizer Main Contact evan.calabrese@duke.edu]
Dominic LaBella MD [Co-Organizer dominic.labella@duke.edu]
Duke University Medical Center

Zachary Reitman MD PhD & John Kirkpatrick MD PhD & Chunhao Wang PhD
Duke University Medical Center

Keyvan Farahani
NIH

Verena Chang
Sage Bionetworks

Jake Albrecht
Sage Bionetworks

Clinical Evaluators and Annotation Approvers:
Evan Calabrese MD PhD
Dominic LaBella MD

Annotation Volunteers
BraTS 2023 Meningioma Annotators Pool

Data Contributors:
Evan Calabrese MD PhD & Zachary Reitman MD PhD & Chunhao Wang PhD & John Kirkpatrick MD PhD
Duke University Medical Center, Department of Radiology

Javier Villanueva-Meyer MD
University of California San Francisco

Spyridon Bakas PhD, & Ujjwal Baid PhD
Indiana University

Adam Flanders MD
Thomas Jefferson University Medical Center

Mariam Aboian MD PhD
Yale University Medical Center

Ayman Nada MD
Missouri University Medical Center

b) Provide information on the primary contact person.

Evan Calabrese, MD PhD
[Lead Organizer of the "BraTS-Meningioma: Segmentation of Post-treatment Intracranial Meningioma"]
Duke University Medical Center
Email: evan.calabrese@duke.edu

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NCI (represented by Dr. Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr. Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data, but if they do so, they MUST also discuss the potential difference in their results after using only the BraTS 2024 data, since our intention is to solve the particular segmentation problem, but also to provide a fair comparison among the participating methods.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel and 2) Neosoma Inc, we are currently coordinating for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge.
Note that Intel has been offering monetary awards during each of BraTS 2018-2022, and Neosoma for BraTS 2021. NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the output of their methods to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it in the three hidden testing data. The participants will be provided guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility, thereby maximizing solutions in solving the problem of brain tumor segmentation.

During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Tool [7] (https://fets-ai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).

5 C .Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

6 S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor

research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in May after the release of the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of

presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We are already in close coordination with The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH), to release the training and validation data following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions). The TCIA has already approved this, and we are now in the process of submission (includes a detailed curation process specific to TCIA). The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Google s AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Informed consent (or an approved waiver of informed consent) was obtained for all subjects at their respective institutions, and the protocol for releasing the data was approved by the institutional review board of the data contributing institution.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Platform [7] (https://fets-ai.github.io/Front-End/).

5 C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

6 S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS 2021-2023 challenges.

All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Challenge organizers, data contributors, SAGE Bionetworks, and the clinical evaluators will have access to the validation, and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval

- Segmentation
- Tracking

Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients diagnosed with intracranial meningioma and have already undergone treatment, which may include surgery, radiation, and/or chemotherapy.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with meningioma, clinically scanned for radiation planning purposes including 3D post-contrast T1-weighted spoiled gradient echo imaging at 1.5-3 Tesla.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

MRI.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

Gross tumor volume label will be included along with the MRI images. No additional patient clinical information will be provided.

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Radiotherapy planning brain MRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Dice, Hausdorff 95th percentile, Sensitivity, Precision, Specificity. - per lesion evaluation

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in our related manuscripts [1,2,4]. Since then, multiple institutions have contributed data to create the current BraTS 2023 Meningioma dataset and these will be listed in the latest BraTS arXiv paper following acceptance of the challenge. We are currently in coordination with TCIA to make the complete BraTS 2021-2023 dataset permanently available through their portal. All the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, including Google and AWS Marketplaces, as part of their Public Datasets Programs.

1 U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
2 S.Bakas, 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
4 S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient (where available) will be published as supplementary material together with the challenge meta-analysis manuscript. Image pre-processing will consist only of automated face randomization to preserve patient anonymity.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe MRI scans, acquired with different clinical protocols and various scanners from: Duke University, University California San Fransisco, Indiana University, Thomas Jefferson University, Yale University, etc.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

Patients diagnosed with intracranial meningioma undergoing MRI brain for radiotherapy planning during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes an MRI scan for a single patient at a single timepoint. The exact scan included for one case is 3D post-contrast T1-weighted spoiled gradient echo imaging at 1.5-3 Tesla.
Please note that images included for each case of the provided dataset represent the sequences with the best image quality available in the acquiring institution for this particular case. All images will be maintained in their original (native) orientation and spacing.

b) State the total number of training, validation and test cases.

The following are estimates based on currently available data. We are working with other institution to further increase the number of cases for the 2024 challenge:
Training data: 300 cases
Validation data: 50 cases
Testing data: 150 cases
Note: None of these cases have been used in previous challenges.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

N/A

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Reference approved from at least 2 experienced neuroradiologists/radiation oncologists, following annotations from 60 clinical neuroradiologists (volunteers from the BraTS 2023 annotators pool)

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists and/or radiation oncologists), which is provided to all clinical annotators, describing in detail instructions on what the gross tumor volume segmentation should and should not include. The annotators are given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions:
Gross tumor volume (GTV) is defined as the entire extent of imaging abnormality represented by the tumor itself including any tumor extending into or past the cranial vault but not including any surrounding edema within the adjacent brain parenchyma.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case is assigned to a pair of annotator-approver. Annotators span across various experience levels and clinical/academic ranks, while the approvers are 2 board-certified neuroradiologists or neuro-focused radiation oncologists, listed in the Organizers' section as clinical evaluators and annotation approvers. The annotators are given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators are satisfied with the produced annotations, these are passed to an approver. The approver is then responsible for signing off on these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scan, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach is followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

In an effort to improve clinical relevance of the 2024 BraTS meningioma challenge, only the minimum necessary preprocessing will be applied to preserve patient anonymity. Specifically, images will be converted from DICOM to NIfTI format and patient faces will be digitally replaced with a randomized average face to prevent 3D renderings of real patient faces. No additional pre-processing methods will be employed.

STEP 1: Convert from DICOM to NIfTI using the publicly available dcm2niix tool [9].

STEP 2: Digitally randomize patient faces using the publicly available mri_dreface tool [10].

9 Li X, Morgan PS, Ashburner J, Smith J, Rorden C (2016) The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J Neurosci Methods. 264:47-56. doi: 10.1016/j.jneumeth.2016.03.001. PMID: 26945974

10 Schwarz CG, Kremers WK, Wiste HJ, Gunter JL, Vemuri P, Spychalla AJ, Kantarci K, Schultz AP, Sperling RA, Knopman DS, Petersen RC, Jack CR, 2021. Changing the face of neuroimaging research: comparing a new MRI de-facing technique with popular alternatives. Neuroimage 231. doi: 10.1016/j.neuroimage.2021.117845

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [8] and is outside the scope of the BraTS 2024 challenge.

8 R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise

Sensitivity, Lesionwise
Specificity, Lesionwise
Precision, Lesionwise
The regions evaluated using these metrics describe the gross tumor volume label and are calculated in the patient native acquisition space per lesion.

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

The gross tumor volume (GTV) is the basis for meningioma radiotherapy planning and is required for effective radiation treatment.

In terms of evaluation metrics, we use:
i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance, ii) the 95% Hausdorff distance as opposed to standard HD, in order to avoid outliers having too much weight, iii) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or undersegment. iv) Precision to complement the metric of Sensitivity (also known as recall).

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics at Indiana University School of Medicine), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (0 for the DSC and the image diagonal for the HD).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot [6].

11 Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,

- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2023, uncertainties in rankings will be assessed using permutational analyses [3]. Performance for the segmentation task will be assessed based on relative performance of each team and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

2 S. Bakas et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, arXiv:1811.02629 [cs, stat], Apr. 2019, Accessed: Dec. 10, 2020. [Online]. Available: http://arxiv.org/abs/1811.02629.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 3: BraTS-Metastasis: Segmentation of Pre- and Post-Treatment Brain Metastases

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Background: Clinical monitoring of metastatic disease in the brain is laborious and time-consuming, especially in the setting of multiple metastases and when performed manually. Response assessment in brain metastases based on maximal unidimensional diameter as per the RANO-BM guideline is commonly performed, however, accurate volumetric lesion and peri-lesional edema estimates can be crucial for clinical decision-making and enhance outcome prediction. The unique challenge of performing segmentations in brain metastases is that they are commonly small and detection and segmentation of lesions that are smaller than 10 mm has not demonstrated high dice similarity coefficients in prior publications. We propose that machine learning based auto-segmentation approach of brain metastases and perilesional edema will improve the time-efficiency, reproducibility and robustness against inter-rater variability. These efficiencies will provide critical algorithms that are translatable to the post-treatment setting.

Impact: Brain metastases are the most common CNS malignancy in adults and evaluation of brain metastases in clinical practice is commonly limited to comparison to one prior imaging study due to common presentation of multiple metastases in single patient. Detailed analysis of multiple patient lesions on multiple serial scans is impossible in current clinical practice because of the time it requires to assess a study. Therefore, development of automated segmentation tools for brain metastases are critical for providing precision based patient care. In addition, accurate detection of small metastatic lesions that are smaller than 10 mm and are an average of 1-2 mm is critical for patient prognosis and missing even a single lesion can result in the patient requiring repeat interventions, and experience delays in treatment. In addition, gross total volume of brain metastases in a patient is an important predictor of patient outcomes and is not currently available in clinical practice due to the lack of volumetric segmentation tools that can be translated. Therefore, it is critical to develop novel segmentation algorithms for small brain metastases that detect and accurately volumetrically segment all lesions in pre-treatment and post-treatment setting, which is different from the 2023 challenge that was solely focused on pre-treatment segmentations. In BraTS 2024, we are expanding the pre-treatment dataset with more diverse patients and we are adding post-treatment imaging studies, including post-radiation and post-surgical cases. Many of the algorithms that were developed for gliomas, such as nnUnet, demonstrate high dice scores for larger metastases but their performance significantly drops off for small metastases. This challenge will be critical for the development of novel segmentation and detection algorithms for brain metastases that are common in clinical practice and will provide algorithms that can be readily translated into clinical practice.

Datasets:

1) Cluster of datasets from the following sites (Duke University-103 patients, National cancer institute-41 patients, Missouri University-165 patients, Washington University-112 patients, Yale University-225 patients, Northwestern

University-339 patients, and Heidelberg University-300 patients). All these patients have multiparametric MRI with T1, post-gadolinium T1, T2, and FLAIR sequences available. The post-gadolinium T1 of these MRI are thin slices 0.5-1 mm to ensure inclusion of small metastasis. These studies are split into training (n=900), validation (n=128), and test (n=257) sets. Manually generated segmentations of contrast enhancing lesion, areas of internal necrosis and peri-lesional edema were revised by two board certified neuro-radiologists and will serve as ground truth.
2) NYU dataset - publicly available dataset containing 1147 patients. All these patients have multiparametric MRI with T1, post-gadolinium T1, T2 and FLAIR sequences are available. The NYU cases will be used for training only and will not be included in validation and test sets due to technical difficulties. NYU studies as well as the ground truth segmentation mask will be hosted on the NYU website only. A website hyperlink will be available on the main challenge website where all participants can freely download the dataset.
3) UCSF (n=425 patients) and Stanford University (n=151) datasets. All patients have MRI with T1, post gadolinium T1, T2, and FLAIR sequences available. Due to absence of T2 in UCSF and Stanford datasets, these cases will be separated from the main dataset cohort and will be available only for training.

Keywords

List the primary keywords that characterize the task.

Segmentation, Brain metastasis, pre-treatment, post-treatment, contrast enhancing lesion, peritumoral edema

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Mariam Aboian [Lead Organizer - Contact Person]
Childrens Hospital of Philadelphia
Co-lead: Aly Abayazeed, MD Neosoma
Co-lead: Philipp Lohman, PhD Research Center Juelich (FZJ), Germany
Trainee Lead: Ahmed Moawad, M.D.
Trainee Annotation chief: Anastasia Janas, MD/PhD

Additional coorganizers
Spyridon Bakas
Kiril Krantchev
Gian Marco Conte
Fatima Memon
Florian Kofler
Ujjwal Baid
Yury Velichko
Elizabeth Schrickel
Katie Link
Hongwei Li
Sanjay Aneja
Ryan Maresca
Ayman Nada

Philipp Vollmuth
Víctor Manuel Pérez
Keyvan Farahani
Matthew W Pease
Devon Godfrey
Scott Floyd
Jeffrey Rudie
Jake Albrecht
Verena Chung

The trainee annotator group is continuously growing, currently encompassing around 150 individuals from over 15 countries. These trainees have various backgrounds, including medical students in the latter stages of their education, radiology residents, and researchers. To ensure high-quality annotations, each one undergoes thorough training before beginning the actual annotation.

b) Provide information on the primary contact person.

Mariam Aboian, MD PhD
[Lead Organizer of the "BraTS-Metastasis: Segmentation of Pre- and Post-Treatment Brain Metastases"]
Childrens Hospital of Philadelphia
Email: aboianm@chop.edu

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org
Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NCI

(represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2023 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for "a sustainable medical imaging challenge cloud infrastructure," to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data, but if they do so, they MUST also discuss the potential difference in their results after using only the BraTS 2023 and 2024 data, since our intention is to solve the particular segmentation problem, but also to provide a fair comparison among the participating methods.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel and 2) Neosoma Inc, we are currently coordinating for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge.
Note that Intel has been offering monetary awards during each of BraTS 2018-2022, and Neosoma for BraTS 2021.
NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the output of their methods to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it in the three hidden testing data. The participants will be provided guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility, running of these algorithms to the previous BraTS instances and comparison with results obtained by algorithms of previous years, thereby maximizing solutions in solving the problem of brain tumor segmentation.

During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk, https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Tool (https://fets-ai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).

C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and

predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018.
https://doi.org/10.1117/1.jmi.5.1.011018

S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019.
https://doi.org/10.1007/978-3-030-46643-5_38

S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. https://doi.org/10.1088/1361-6560/ac9449

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in May after the release of the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the

camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We are already in close coordination with The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH), to release the training and validation data following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions).
The TCIA has already approved this, and we are now in the process of submission (includes a detailed curation process specific to TCIA). The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Googles AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Informed consent was obtained from all subjects at their respective institutions, and the protocol for releasing the data was approved by the institutional review board of the data-contributing institution.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)

- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Cancer Imaging Phenomics Toolkit (CaPTk, https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Platform (https://fets-ai.github.io/Front-End/).

C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018, 2018.
https://doi.org/10.1117/1.jmi.5.1.011018

S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS 2021-2023 challenges.

All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for "a sustainable medical imaging challenge cloud infrastructure," to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as

Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Spyridon Bakas, Ujjwal Baid, SAGE Bionetworks, Mariam Aboian, MD/PhD and the clinical evaluators will have access to the validation, and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling

- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients diagnosed with metastasis, agnostic to whether they have undergone treatment or not.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with metastasis, clinically scanned with mpMRI acquisition protocol during pre-treatment and post-treatment including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

... directly to the image data (i.e., tumor sub-region volumes)

1 Jeffrey Rudie MD/PhD University of California San Diego
2 Mariam Aboian MD/PhD Yale University School of Medicine
3 Philipp Volmouth MD University of Heidelberg
4 Nourel hoda Tahon MD, Msc Ayman Nada MD/PhD University of Missouri
5 Devon Godfrey PhD Scott Floyd MD/PhD Duke University
6 Satrajit Chakrabart Washington University
7 Ahmed Moawad MD Oleg Teytelboym Mercy Hospital
8 Katie Link New York University

9 Ayuda Youseff MD National Cancer Institute

10 Greg Zaharchuk MD/PhD Stanford University

11 Aly Abayazeed Neosoma Inc.

12 Yury Velichko Northwestern University

13 Spyridon Bakas PhD, Ujjwal Baid PhD, Matthew W. Pease MD Indiana University

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans in patients with brain metastases before and after initiation of treatment.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Dice, Hausdorff 95th percentile, Sensitivity, Precision, Specificity. - per lesion evaluation

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in our related manuscripts. Since then, multiple institutions have contributed data to create the current BraTS 2023 Metastasis dataset and these will be listed in the latest BraTS arXiv paper following acceptance of the challenge. We are currently in coordination with TCIA to make the complete BraTS 2021-2023 dataset permanently available through their portal. All the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, including Google and AWS Marketplaces, as part of their Public Datasets Programs.

U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314

S.Bakas, 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

S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe mpMRI scans, acquired with different clinical protocols and various scanners from:

Yale University School of Medicine
Northwestern University
University of Heidelberg
University of Missouri
Duke University
Washington University
Mercy Hospital
National Cancer Institute
Stanford University
New York University (NYU)
University of California San Francisco (UCSF)
Indiana University (IU)

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at multiple timepoints. Pre-treatment and posttreatment scans are included in the datasets. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI. Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There was no inclusion/exclusion criterion applied that related to 3d acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

The currently available dataset include:
Total data: 3,008
Training data: 1,285 (additional optional 1,723 cases available for training only)
Validation data: 128 cases
Test data: 257
Note: The above reported numbers have been used in the 2023 challenge.
We are working with other institutions to further increase the number of cases for BraTS 2024 challenge.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on our preliminary data in glioblastoma and brain metastasis segmentation, the plateau for training segmentations of brain metastases is reached at approximately 150 cases using the nnUnet algorithm (Merkaj et al, 2021). Training data of 413 cases is sufficient to train the algorithm. We have accumulated over 2000 cases of unlabeled data and the focus of this year s challenge is to increase the number of high-quality annotations and add about 250 cases per month from February to late April 2024.

Merkaj S, Bousabarah K, Zeevi T, Lin M, Aboian MS. PACS based glioma segmentation and grade prediction for clinical implementation, ASFNR-ASNR AI workshop, 2021

Based on availability.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

N/A

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Reference approved from at least 2 experienced neuroradiologists, following annotations from over 200 annotators (students, residents, postgraduate fellows).

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The data considered in this task of the BraTS-METS 2024 challenge follows the paradigm of the BraTS 2021-2022 and BraTS-METS 2023 challenge data. The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists), which was provided to all clinical annotators, describing in detail instructions on what the segmentations of each tumor sub-region should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations with preference placed on ITK-SNAP, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. All segmentations are checked in the final step by one final annotator.

Summary of specific instructions:
i) the enhancing tumor (when present) delineates the hyperintense signal of the T1-Gd, after excluding the vessels.
ii) the necrotic core (when present) outlines regions appearing dark in both T1 and T1-Gd images (denoting necrosis/cysts), and darked regions in T1-Gd that appear brighter in T1.
iii) the tumor core, which is the union of the enhancing tumor and the necrotic core described in (i) and (ii) above.
iv) the farthest tumor extent including the edema (what is called the whole tumor), delineates the tissue represented by the abnormal T2-FLAIR envelope.
v) resection cavity delineates the resection of region within the brain in post-treatment cases

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers are experienced board-certified neuroradiologists (with >5 years of experience), listed in the "Organizers" section as "clinical evaluators and annotation approvers". The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators were satisfied with the produced annotations, they were passing these to two different approvers. Approver 1 is then responsible for signing off these annotations. Specifically, approver 1 would review the tumor annotations, in tandem with the corresponding MRI scans, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to approver 1) . The segmentation mask from Approver 1 is passed blindly to Approver 2 for further refinement. The whole dataset is finally approved by "Final approver" to ensure consistency. The "final approver" make the segmentation available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The exact preprocessing pipeline applied to all the data considered in the BraTS 2024 challenge is similar with the one evaluated and followed by the BraTS 2017-2022 challenges. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format (Cox et al, 2004), we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 (Rohlfing et al 2010)) and interpolating to the same resolution as this atlas (1 mm3). The exact registration process comprises the following steps:

STEP 1: N4 Bias field correction (notably the application of N4 bias field correction is a temporary step. Taking into consideration we have previously (Bakas et al 2017) shown that use of non-parametric, non-uniform intensity normalization (i.e., N4) to correct for intensity non-uniformities caused by the inhomogeneity of the scanner s magnetic field during image acquisition obliterates the MRI signal relating to the abnormal/tumor regions, we intentionally use N4 bias field correction in the preprocessing pipeline to facilitate a more optimal rigid registration across the difference MRI sequences. However, after obtaining the related information (i.e., transformation matrices), we discard the bias field corrected scans, and we apply this transformation matrix towards the final co-registered output images used in the challenge).

STEP 2: Perform brain extraction to remove any apparent non-brain tissue (e.g., neck fat, skull, eyeballs) based on a deep-learning approach we developed in house, focusing on scans with apparent brain tumors and exhaustively evaluated it in both private and public multi-institutional data (Thakur et al 2020). We then manually assessed all scans for confirming the correct brain extraction (i.e., skull stripping), where the complete brain region is included, and all non-brain tissue is excluded.

This whole pipeline, and its source code are available through the CaPTk (Bakas et al, 2017) (https://github.com/CBICA/CaPTk) and FeTS [7] (https://fets-ai.github.io/Front-End/) platforms.

S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117, 2017. DOI: 10.1038/sdata.2017.117

T. Rohlfing, et al. The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp.

31(5):798-819, 2010.

R.Cox, J.Ashburner, H.Breman, K.Fissell, C.Haselgrove, C.Holmes, J.Lancaster, D.Rex, S.Smith, J.Woodward, A (Sort of) new image data format standard: NIfTI-1: WE 150, Neuroimage, 22, 2004.

S.Thakur, J.Doshi, S.Pati, S.Rathore, C.Sako, M.Bilello, S.M.Ha, G.Shukla, A.Flanders, A.Kotrotsou, M.Milchenko, S.Liem, G.S.Alexander, J.Lombardo, J.D.Palmer, P.LaMontagne, A.Nazeri, S.Talbar, U.Kulkarni, D.Marcus, R.Colen, C.Davatzikos, G.Erus, S.Bakas, 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. DOI: 10.1016/j.neuroimage.2020.117081

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) (Mehta et al, 2022) and is outside the scope of the BraTS 2022 challenge.

R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise
Sensitivity, Lesionwise
Specificity, Lesionwise
Precision Lesionwise
The regions evaluated using these metrics describe the whole tumor, the tumor core, and the enhancing tumor (when present). Note that the tumor core includes the part of the tumor that is typically resected (i.e., enhancing, non-enhancing, and necrotic tumor), and the whole tumor describes all tumor sub-regions (i.e., tumor core and

edema/invasion).

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of the assessed and evaluated tumor sub-regions:
i) the enhancing tumor describes the regions of active tumor with blood brain barrier breakdown and based on this, clinical practice characterizes the extent of resection based on removal of the contrast enhancing region.
ii) the tumor core (incl. the necrotic component) describes what is typically resected during a surgical procedure.
iii) the whole tumor as it defines the whole extent of the tumor, including the peritumoral edematous tissue and non-enhancing infiltrated area.

In terms of evaluation metrics, we use:
i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance,
ii) the 95% Hausdorff distance as opposed to standard HD, in order to avoid outliers having too much weight,
iii) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or undersegment.
iv) Precision to complement the metric of Sensitivity (also known as recall).

All metrics are performed on per lesion basis.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics at Indiana University School of Medicine), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (0 for the DSC and the image diagonal for the HD).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot (Duan et al, 2020).

Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,

- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2022, uncertainties in rankings will be assessed using permutational analyses (Bakas et al, 2019). Performance for the segmentation task will be assessed based on relative performance of each team on each tumor tissue class and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

S. Bakas et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, arXiv:1811.02629 [cs, stat], Apr. 2019, Accessed: Dec. 10, 2020. [Online]. Available: http://arxiv.org/abs/1811.02629.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 4: BraTS-Africa: Segmentation of Brain Glioma in Sub-Saharan Africa patient population

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Brain tumors are among the deadliest type of cancer. Approximately 80% of individuals with Glioblastoma (GB) die within two years of diagnosis 1. Brain tumors in general are challenging to diagnose, hard to treat and inherently resistant to conventional therapy. Years of extensive research to improve diagnosis and treatment of GB have decreased mortality rates in the U.S by 7% over the past 30 years 2. Although modest, these research innovations have not translated to improvements in survival for adults and children in low- and middle-income countries (LMICs), particularly in African populations where death rates in Sub-Saharan Africa (SSA) rose by approximately 25% on average while decreasing by up to 30% in the Global North 2. Long-term survival with GB is associated with identification of appropriate pathological features on brain MRI and confirmation by histopathology. Since 2012, the BraTS Challenge have evaluated state-of-the art machine learning methods to detect, characterize, and classify brain GB. In 2023, BraTS featured African data (BraTS-Africa) as a sub-challenge, with 18 teams participating from around the globe 3.

This year, the BraTS-Africa Challenge provides a renewed opportunity to expand the brain MRI GB cases from Sub-Saharan Africa in global efforts to develop and evaluate computer-aided-diagnostic (CAD) methods for detection and characterization of GB in resource-limited settings, where the potential for CAD tools to transform healthcare are more likely 4.

1 M. Poon, et al., Longer-term (>= 2 years) survival in patients with glioblastoma in population-based studies pre- and post-2005: a systematic review and meta-analysis. Sci Rep. 2020 Jul 15;10(1):11622. https://doi.org/10.1038/s41598-020-68011-4

2 WHO GBD 2016 Brain and Other CNS Cancer Collaborators. Global, regional, and national burden of brain and other CNS cancer, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016, Lancet Neurol. 2019 Apr;18(4):376-393. https://doi.org/10.1016/S1474-4422(18)30468-X

3 M. Adewole, et al., The Brain Tumor Segmentation (BraTS) Challenge 2023: Glioma Segmentation in Sub-Saharan Africa Patient Population (BraTS-Africa) (arXiv:2305.19369), https://doi.org/10.48550/arXiv.2305.19369

4 U. Anazodo, et al., AI for Population and Global Health in Radiology. Radiology: Artificial Intelligence, 2022. https://doi.org/10.1148/ryai.220107

Keywords

List the primary keywords that characterize the task.

Segmentation, Glioma, Challenge, Sub-Saharan Africa, BraTS, MRI

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Udunna Anazodo, Ph.D. - Lead Organizer
McGill University

Maruf Adewole, MSc
Medical Artificial Intelligence Laboratory (MAI Lab), Lagos, Nigeria

Jeffrey Rudie, MD PhD
University of California, San Diego

Spyridon Bakas PhD & Ujjwal Baid PhD
Indiana University

Farouk Dako
University of Pennsylvania

Keyvan Farahani, Ph.D.
NIH

Jake Albrecht & Verena Chung
Sage Bionetworks

Clinical Evaluators and Annotation Approvers:
Jeffrey Rudie, MD PhD
University of California, San Diego

Oluyemisi Toyobo
Crestview Radiology Ltd., Nigeria.

Olubukola Omidiji
Lagos University Teaching Hospital, Nigeria

Annotation Volunteers
Yewande Gbadamosi & Afolabi Ogunleye
Lagos State University Teaching Hospital, Lagos

Nancy Ojo
Federal Medical Centre, Abeokuta

Kator Iorpagher
Benue State University, Makurdi

Gabriel Babatunde
Lagos University Teaching

Kenneth Aguh
Federal Medical Center, Umuahia

Adaobi Emegoakor
Nnamdi Azikiwe University Hospital, Nnewi

Chinasa Kalaiwo
National Hospital Abuja

b) Provide information on the primary contact person.

Udunna Anazodo PhD
[Lead Organizer of the "BraTS-Africa: Segmentation of Brain Glioma in Sub-Saharan Africa patient population"]
Montreal Neurological Institute, McGill University
Email: udunna.anazodo@mcgill.ca

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-

MICCAI BraTS 2021 challenge 5 and support from NCI (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group), we have coordinated with Synapse to use their platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2024 Challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data. If they do so, they MUST fully describe the additional datasets and discuss the potential difference in their results after using only the BraTS 2024 data. Since our intention is to solve the particular segmentation problem, but also to provide a fair comparison among the participating methods, enhancing the training data is conditionally permitted.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel and 2) Neosoma Inc, we are coordinating for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge. Note that Intel has been offering monetary awards during each of BraTS 2018-2023, and Neosoma for BraTS

2021-2023.

NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the output of their methods to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it in the three hidden testing data. The participants will be provided guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility, running of these algorithms to the previous BraTS instances and comparison with results obtained by algorithms of previous years, thereby maximizing solutions in solving the problem of brain tumor segmentation.

During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk 5-6, https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Tool 7 (https://fets-ai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).

All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

5 C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

6 S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in June after the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

The brain MRI for Africa BraTS challenge is specifically retrospectively collected images where patient informed consent was not feasible. However, the study has been approved by the Institution Review Board of Western University (ID: 121287), College of Medicine of the University of Lagos (ID: CMUL/HREC/04/22/1090), Lagos State University Teaching Hospital (ID: LREC/06/10/1952), Lily Hospital Benin (ID: LH/HREC-MA/0050-23) and National Hospital Abuja (ID: NHA/EC/049/2023). Ethics approvals will also be obtained from the IRBs of all data contributing centers.

We are already in close coordination with The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH), to release the training and validation data following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions).
The TCIA has already approved this, and we are now in the process of submission (includes a detailed curation process specific to TCIA). The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data

available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Google s AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Cancer Imaging Phenomics Toolkit (CaPTk 5-6, https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Platform 7 (https://fets-ai.github.io/Front-End/).

5 C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

6 S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS 2021-2023 challenges.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc.

Jeff Rudie, Spyridon Bakas, Ujjwal Baid, Maruf Adewole, SAGE Bionetworks, and the clinical evaluators will have access to the validation, and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Sub-Saharan Africa patients diagnosed with de novo diffuse gliomas of the brain.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients from Africa, diagnosed with de novo diffuse gliomas of the brain, clinically scanned with mpMRI acquisition protocol including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region volumes)

Data Contributors:
Abiodun Fatade, MBBS
Crestview Radiology, Lagos, Nigeria

Olubukola Omidiji, MBBS
Lagos University Teaching Hospital, Lagos Nigeria.

Rachel Akinola, MBBS
Lagos State University Teaching Hospital, Lagos Nigeria

Feyisayo Daji
National Hospital, Abuja, Nigeria

M.A Suwaid, MBBS
Aminu Kano Teaching Hospital, Lagos, Nigeria

Kenneth Aguh
Medhub Africa

Mayomi Onuwaje
Lily Hospitals Benin

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Dice, Hausdorff 95th percentile, Sensitivity, Precision, Specificity. - per lesion evaluation

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the BraTS 2024 cohort has been listed in the data reference published in our related manuscripts 1,2,4,5. Since then, multiple institutions have contributed data to create the current MICCAI BraTS dataset, and these are listed in the latest BraTS-Africa arxiv paper 5. We are currently in coordination with TCIA to make the complete BraTS-Africa 2023 dataset permanently available through their portal. All the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, including Google and AWS Marketplaces, as part of their Public Datasets Programs.

1 U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
2 S.Bakas, 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
4 S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117
(2017) DOI: 10.1038/sdata.2017.117
5 Adewole, M., Rudie, J. D., Gbadamosi, A., Toyobo, O., Raymond, C., Zhang, D., Omidiji, O., Akinola, R., Suwaid, M. A., Emegoakor, A., Ojo, N., Aguh, K., Kalaiwo, C., Babatunde, G., Ogunleye, A., Gbadamosi, Y., Iorpagher, K., Calabrese, E., Aboian, M., et Int, Anazodo, U. C. (2023). The Brain Tumor Segmentation (BraTS) Challenge 2023: Glioma Segmentation in Sub-Saharan Africa Patient Population (BraTS-Africa) (arXiv:2305.19369). arXiv.
https://doi.org/10.48550/arXiv.2305.19369

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe mpMRI scans, acquired with different clinical protocols and various scanners across sub-Saharan Africa from:
Crestview Radiology, Lagos, Nigeria (1.5 T Siemens)

Lagos University Teaching Hospital, Lagos Nigeria (1.5T Toshiba/Canon)
Lagos State University Teaching Hospital, Lagos Nigeria (1.5T Philips)
The National Hospital, Abuja, Nigeria (1.5 T Toshiba)
Aminu Kano Teaching Hospital, Lagos, Nigeria (1.5 T Siemens)
Lily Hospital, Benin (1.5T GE)
Medhub Africa (multi-scanner)

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at a single timepoint. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There was no inclusion/exclusion criterion applied that related to 3D acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

Training data: 120 cases
Validation data: 40 cases
Testing data: 60 cases
Note: 50% of the above reported numbers have been used in the 2023 challenge.

These numbers are expected to increase as we continuously collect and annotate more cases.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

N/A

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Reference approved from at least 2 experienced neuroradiologists, following annotations from 10 clinical neuroradiologists (volunteers from ASNR, ARIN, or other African Imaging Societies)

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The data considered in this task of the BraTS-Africa 2024 challenge follows the paradigm of the BraTS 2021-2023 challenge data. The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists), which was provided to all clinical annotators, describing in detail instructions on what the segmentations of each tumor sub-region should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions:
i) the enhancing tumor (when present) delineates the hyperintense signal of the T1-Gd, after excluding the vessels.
ii) the necrotic core (when present) outlines regions appearing dark in both T1 and T1-Gd images (denoting necrosis/cysts), and darked regions in T1-Gd that appear brighter in T1.
iii) the tumor core, which is the union of the enhancing tumor and the necrotic core described in (i) and (ii) above.
iv) the farthest tumor extent including the edema (what is called the whole tumor), delineates the tissue represented by the abnormal T2-FLAIR envelope.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers were the 2 experienced board-certified neuroradiologists (with >15 years of experience), listed in the Organizers' section as clinical evaluators and annotation approvers. The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators were satisfied with the produced annotations, they were passing these to the corresponding approver. The approver is then responsible for signing off these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scans, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The exact preprocessing pipeline applied to all the data considered in the BraTS 2023 challenge is identical with the one evaluated and followed by the BraTS 2017-2022 challenges. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format 10, we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 9) and interpolating to the same resolution as this atlas (1 mm^3). The exact registration process comprises the following steps:

STEP 1: N4 Bias field correction (notably the application of N4 bias field correction is a temporary step. Taking into consideration we have previously 4 shown that use of non-parametric, non-uniform intensity normalization (i.e., N4) to correct for intensity non-uniformities caused by the inhomogeneity of the scanner's magnetic field during image acquisition obliterates the MRI signal relating to the abnormal/tumor regions, we intentionally use N4 bias field correction in the preprocessing pipeline to facilitate a more optimal rigid registration across the difference MRI sequences. However, after obtaining the related information (i.e., transformation matrices), we discard the bias field corrected scans, and we apply this transformation matrix towards the final co-registered output images used in the challenge).

STEP 2: Rigid Registration of T1, T2, T2-FLAIR to the T1-Gd scan, and obtain the corresponding transformation matrix.

STEP 3: Rigid Registration of T1-Gd scan to the SRI-24 atlas 9, and obtain the corresponding transformation matrix.

STEP 4: Join the obtained transformation matrices and applying aggregated transformation to the LPS-oriented scans.

STEP 5: After completion of the registration process, we perform brain extraction to remove any apparent non-brain tissue (e.g., neck fat, skull, eyeballs) based on a deep-learning approach we developed in house, focusing on scans with apparent brain tumors and exhaustively evaluated it in both private and public multi-institutional data 11. We then manually assessed all scans for confirming the correct brain extraction (i.e., skull stripping), where the complete brain region is included, and all non-brain tissue is excluded.

This whole pipeline, and its source code are available through the CaPTk [5-6] (https://github.com/CBICA/CaPTk) and FeTS [7] (https://fets-ai.github.io/Front-End/) platforms.

4 S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117, 2017. DOI: 10.1038/sdata.2017.117

9 T. Rohlfing, et al. The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798-819, 2010.

10 R.Cox, J.Ashburner, H.Breman, K.Fissell, C.Haselgrove, C.Holmes, J.Lancaster, D.Rex, S.Smith, J.Woodward, A (Sort of) new image data format standard: NIfTI-1: WE 150, Neuroimage, 22, 2004.

11 S.Thakur, J.Doshi, S.Pati, S.Rathore, C.Sako, M.Bilello, S.M.Ha, G.Shukla, A.Flanders, A.Kotrotsou, M.Milchenko, S.Liem, G.S.Alexander, J.Lombardo, J.D.Palmer, P.LaMontagne, A.Nazeri, S.Talbar, U.Kulkarni, D.Marcus, R.Colen, C.Davatzikos, G.Erus, S.Bakas, 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. DOI: 10.1016/j.neuroimage.2020.117081

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) 8 and is outside the scope of the BraTS 2024 challenge.

8 R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise
Sensitivity, Lesionwise
Specificity, Lesionwise
Precision , Lesionwise
The regions evaluated using these metrics describe the whole tumor, the tumor core, and the enhancing tumor (when present). Note that the tumor core includes the part of the tumor that is typically resected (i.e., enhancing, non-enhancing, and necrotic tumor), and the whole tumor describes all tumor sub-regions (i.e., tumor core and edema/invasion).

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of the assessed and evaluated tumor sub-regions:
i) the enhancing tumor describes the regions of active tumor and based on this, clinical practice characterizes the extent of resection.
ii) the tumor core (incl. the necrotic component) describes what is typically resected during a surgical procedure.
iii) the whole tumor as it defines the whole extent of the tumor, including the peritumoral edematous tissue and highly infiltrated area.

In terms of evaluation metrics, we use:
i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance,
ii) the 95% Hausdorff distance as opposed to standard HD, in order to avoid outliers having too much weight,
iii) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or under-segment.
iv) Precision to complement the metric of Sensitivity (also known as recall).

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of the Dept of Biostatistics and Health Data Science), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (0 for the DSC and the image diagonal for the HD).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot 6.

12 Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2023, uncertainties in rankings will be assessed using permutational analyses 3. Performance for the segmentation task will be assessed based on relative performance of each team on each tumor tissue class and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

2 S. Bakas et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, arXiv:1811.02629 cs, stat, Apr. 2019, Accessed: Dec. 10, 2020. Online. Available: http://arxiv.org/abs/1811.02629.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 5: BraTS-PEDs: Segmentation of Pre-Treatment Pediatric Tumors

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Brain tumors are among the deadliest types of cancer and the BraTS Challenge [1-3] has a successful history of resource creation for the segmentation and analysis of most common and aggressive malignant primary tumor of the central nervous system in adults, namely the glioblastoma multiforme (GBM). Although rare, pediatric tumors of the brain and central nervous system are the most common cause of disease related death in children. Brain tumors in general are challenging to diagnose, hard to treat and inherently resistant to conventional therapy because of the challenges in delivering drugs to the brain. While pediatric tumors may share certain similarities with adult tumors, their imaging and clinical presentations differs. For example, GBMs and pediatric diffuse midline gliomas (DMGs) are both high grade gliomas with short overall survival of about 11-13 months on average. GBMs are found in 3 in 100,000 people, DMGs are about three times rarer. While GBMs are usually found in the frontal or/and temporal lobes at an average age of 64 years, DMGs are usually located in the pons and often diagnose between 5 and 10 years of age. Enhancing tumor region on post-gadolinium T1-weighted MRI and necrotic region are common imaging findings in GBM. But these imaging characteristics are less common or clear in DMGs. Thus, pediatric brain tumors require dedicated imaging tools that help in their characterization and facilitate their diagnosis/prognosis. In 2022, we organized the first initiative to include pediatric brain tumors, specifically DMGs in the test set of the BraTS challenge and results were promising. These findings encouraged us to organize a larger and more diverse initiative in 2023 with multi-institutional pediatric data, leading to BraTS-PEDs 2023 challenge [4]. In the BraTS-PEDs 2024 challenge, we will extend the pediatric brain tumor cohort to a larger set, collected through a few consortiums, including Childrens Brain Tumor Network (CBTN) [5], DIPG/DMG-registry, and across multiple institutions. The challenge participants will have access to the pediatric training and validation data at any point from the Synapse platform. These data will be used to develop, containerize, and evaluate their algorithms in unseen validation data until August 2024, when the organizers will stop accepting new submissions and evaluate the submitted algorithms in the pediatric patient population.

[1] U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
[2] S.Bakas, 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
[3] B. H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
[4] Kazerooni AF, Linguraru MG,. The brain tumor segmentation (BRATS) challenge 2023: Focus on pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs). ArXiv. 2023 May 26.
[5] Familiar AM, Kazerooni AF, et al., A multi-institutional pediatric dataset of clinical radiology MRIs by the Children's Brain Tumor Network. arXiv preprint arXiv:2310.01413. 2023 Oct 2.

Keywords

List the primary keywords that characterize the task.

Segmentation, Brain Tumor, Pediatric, Rare Diseases, Challenge, Diffuse Midline Glioma, CBTN, MICCAI

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Leads:
Marius George Linguraru, D.Phil., M.A., M.Sc. [Lead Organizer - Contact Person]
Childrens National Hospital / George Washington University

Anahita Fathi Kazerooni, PhD, MSc [Co-Lead Organizer]
Childrens Hospital of Philadelphia / University of Pennsylvania

Additional members of the Organizing Team:

Zhifan Jiang, Ph.D.
Childrens National Hospital

Xinyang Liu, Ph.D.
Childrens National Hospital

Deep Gandhi
Childrens Hospital of Philadelphia

Nastaran Khalili
Childrens Hospital of Philadelphia

Spyridon Bakas, PhD
Indiana University, IN, USA

Ujjwal Baid, PhD
Indiana University, IN, USA

Keyvan Farahani, PhD.
National Institutes of Health

Jake Albrecht, PhD
Sage Bionetworks

Verena Chung
Sage Bionetworks

Clinical Evaluators and Annotation Approvers:
==================================
Arastoo Vossough, MD & Mariam Aboian, MD
Childrens Hospital of Philadelphia

Ali Nabavizadeh, MD & Jeffrey B Ware, MD
University of Pennsylvania

b) Provide information on the primary contact person.

Marius George Linguraru, DPhil
[Lead Organizer of the "BraTS-PEDs: Segmentation of Pre-Treatment Pediatric Tumors"]
Childrens National Hospital / George Washington University
Email: mlingura@childrensnational.org

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1] and of BraTS-PEDS 2023, we have coordinated with them and following the support from NCI (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (James Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data, but if they do so, they MUST also discuss the potential difference in their results after using only the BraTS-PEDs 2024-2025 data, since our intention is to solve the particular segmentation problem, but also to provide a fair comparison among the participating methods.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel and 2) Neosoma Inc, we are coordinating for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge.
Note that Intel has been offering monetary awards during each of BraTS 2018-2023, and Neosoma for BraTS 2021-2023.
NIH/NCI will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the output of their methods to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it on testing data. The participants will be provided guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility, running of these algorithms to the previous BraTS instances and comparison with results obtained by algorithms of previous years, thereby maximizing solutions in solving the problem of brain tumor segmentation.

During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk [6-7], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Tool [8] (https://fets-ai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).

All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

[6] C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018

[7] S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978-3-030-46643-5_38
[8] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in June after the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all

participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We are already in close coordination with The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH), to release the training and validation data following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions).
The TCIA has already approved this, and we are now in the process of submission (includes a detailed curation process specific to TCIA). The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Googles AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Informed consent or assent has been obtained from all subjects at their respective institutions or a waiver of informed consent was approved by the local institutional review board. The protocol for releasing the data was approved by the institutional review board of the data-contributing institution.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)

- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

This whole preprocessing pipeline, and its source code are available through the CaPTk [6-7](https://github.com/CBICA/CaPTk) and FeTS [8] (https://fets-ai.github.io/Front-End/) platforms. Pediatric-specific automated defacing and tumor subregion segmentation methods [9-10] will be available on https://github.com/d3b-center/peds-brain-auto-seg-public and https://tinyurl.com/2ksfd9yv. [11]

[6] C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018
[7] S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978-3-030-46643-5_38
[8] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449
[9] Fathi Kazerooni A, et al. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. Neuro-Oncology Advances. 2023 Jan 1;5(1):vdad027.
[10] Vossough A, Fathi Kazerooni A, Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors. Under Review (American Journal of Neuroradiology)
[11] Liu X, Linguraru MG. From adult to pediatric: deep learning-based automatic segmentation of rare pediatric brain tumors. InMedical Imaging 2023: Computer-Aided Diagnosis 2023 Apr 7 (Vol. 12465, pp. 15-19). SPIE.

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS-PEDs 2023 challenge.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Marius George Linguraru, Anahita Fathi Kazerooni, Spyridon Bakas, Ujjwal Baid, SAGE Bionetworks, and the clinical evaluators will have access to the validation and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration

- Retrieval
- Segmentation
- Tracking

Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients, diagnosed with de novo pediatric brain tumors.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with pediatric brain tumors, clinically scanned with mpMRI acquisition protocol including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region volumes)

Data Contributors:
==================================
Childrens Brain Tumor Network (CBTN)
DIPG/DMG Registry
Boston Childrens Hospital
Yale University
Indiana University
University Childrens Hospital Zürich

We already have data available from BraTS-PEDs 2023 that was collected from Childrens National Hospital and Childrens Hospital of Philadelphia (as parts of the international Childrens Brain Tumor Network (CBTN) consortium), Boston Childrens Hospital, and Yale University. For BraTS-PEDs 2024, we also collected a large multi-institutional set of data from DIPG/DMG Registry, additional subjects from CBTN, and a few other institutions. With the support of these initiatives, we aim to double to triple the sample size of our data for the 2024 BraTS-PEDs challenge.

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Dice, Hausdorff 95th percentile, Sensitivity, Precision, Specificity. - per lesion evaluation

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The pediatric brain tumor images collected through CBTN have been acquired on multiple scanners, including but not limited to 1.5T and 3T Siemens and GE scanners. We expect to receive data from other institutions across CBTN and non-CBTN institutes and will provide their technical specifications in the final BraTS manuscript. Furthermore, all the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, including Google and AWS Marketplaces, as part of their Public Datasets Programs.

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe mpMRI scans, acquired with different clinical protocols and various scanners from:
Childrens National Hospital (CBTN site)
Childrens Hospital of Philadelphia (CBTN site)
Other CBTN sites
Boston Childrens Hospital
Yale University
Other sites in DIPG/DMG registry

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at a single timepoint. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI.Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There was no inclusion/exclusion criterion applied that related to 3D acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

The following estimates represent the minimum amount of data we intend to use for the challenge; we expect to increase these numbers through additional cohorts.
Training data: 300 cases
Validation data: 100 cases
Testing data: 100 cases

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

N/A

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Reference approved by at least 2 experienced neuroradiologists, following annotations from over 30 clinical neuroradiologists (volunteers from ASNR)

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The data considered in BraTS-PEDs 2024 challenge follows the paradigm of the BraTS-PEDs 2023 challenge data. The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by a consensus of experienced pediatric neuroradiologists from the Childrens Hospital of Philadelphia, with the annotation method published in [9]). This was provided to all clinical annotators, describing in detail instructions on what the segmentations of each tumor sub-region should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach [9-10] is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions (also can be found in [9]):
Enhancing Tumor: This subregion is described by areas with enhancement (brightness) on T1 post-contrast images as compared to T1 pre-contrast. In case of mild enhancement, checking the signal intensity of normal brain structure can be helpful.
Cystic Component: The appearance of the cystic region is hyperintense (very bright) on T2 and hypointense (dark) on T1CE. The cystic portion should be within the tumor (versus edema which is peritumoral). The brightness is comparable to CSF.
Non-enhancing Tumor: Any abnormal signal intensity within the tumoral region that cannot be defined as enhancing or cystic. For example, the abnormal signal intensity on T1, FLAIR and T2 that is not enhancing on T1CE should be considered as non-enhancing portion.
Edema: This sub-region is defined by the abnormal hyperintense signal (very bright) on FLAIR scans. Edema is

finger-like spreading that preserves underlying brain structure and surrounds the tumor.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers were the 2 experienced board-certified neuroradiologists (with >7 years of experience), listed in the Organizers section as clinical evaluators and annotation approvers. The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators were satisfied with the produced annotations, they were passing these to the corresponding approver. The approver is then responsible for signing off these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scans, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The image preprocessing pipeline applied to all the data considered in the BraTS-PEDs 2024 challenge is identical with the one evaluated and followed by the BraTS 2017-2023 challenges, with the difference in applying pediatric-specific tumor subregion segmentation and automated defacing tools. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format, we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 [9]) and interpolating to the same resolution as this atlas (1 mm3).

After completion of the registration process, we will perform automated defacing based on an in-house pediatric-specific deep-learning tool, to remove some face features that may risk re-identification of the subjects. We will then manually review all scans for confirming the correct defacing, where the complete brain region is included, and all non-brain tissue is excluded.

This whole preprocessing pipeline, and its source code are available through the CaPTk [6-7](https://github.com/CBICA/CaPTk) and FeTS [8] (https://fets-ai.github.io/Front-End/) platforms. Pediatric-specific automated defacing and tumor subregion segmentation methods [9-10] will be available on https://github.com/d3b-center/peds-brain-auto-seg-public and https://tinyurl.com/2ksfd9yv. [11]

[6] C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI:

10.1117/1.jmi.5.1.011018
[7] S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978-3-030-46643-5_38
[8] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449
[9] Fathi Kazerooni A, et al. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. Neuro-Oncology Advances. 2023 Jan 1;5(1):vdad027.
[10] Vossough A, Fathi Kazerooni A, Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors. Under Review (American Journal of Neuroradiology)
[11] Liu X, Linguraru MG. From adult to pediatric: deep learning-based automatic segmentation of rare pediatric brain tumors. InMedical Imaging 2023: Computer-Aided Diagnosis 2023 Apr 7 (Vol. 12465, pp. 15-19). SPIE.

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [12] and is outside the scope of the BraTS 2024-2025 challenge.

[12] R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise
Sensitivity, Lesionwise
Specificity, Lesionwise
Precision Lesionwise

The regions evaluated using these metrics describe the whole tumor, the tumor core, and the enhancing tumor (when present). Note that the tumor core includes the part of the tumor that is typically resected (i.e., enhancing, non-enhancing, and necrotic tumor), and the whole tumor describes all tumor sub-regions (i.e., tumor core and edema/invasion).

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of the assessed and evaluated tumor sub-regions:
i) the enhancing tumor describes the regions of active tumor and based on this, clinical practice characterizes the extent of resection.
ii) the tumor core (incl. the necrotic component) describes what is typically resected during a surgical procedure.
iii) the whole tumor as it defines the whole extent of the tumor, including the peritumoral edematous tissue and highly infiltrated area.

In terms of evaluation metrics, we use:
i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance,
ii) the 95% Hausdorff distance as opposed to standard HD, in order to avoid outliers havings too much weight,
iii) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or undersegment.
iv) Precision to complement the metric of Sensitivity (also known as recall).

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics and Health Data Science), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (0 for the DSC and the image diagonal for the HD).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot [13].

[13] Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or

- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2023, uncertainties in rankings will be assessed using permutational analyses [2]. Performance for the segmentation task will be assessed based on relative performance of each team on each tumor tissue class and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

[2] S. Bakas et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, arXiv:1811.02629 [cs, stat], Apr. 2019, Accessed: Dec. 10, 2020. [Online]. Available: http://arxiv.org/abs/1811.02629.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 6: BraTS-GoAT: Segmentation Generalizability Across (Pre-treated) Tumors

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

The International BraTS challenge has been focusing, since its inception in 2012, on the generation of a benchmarking environment and a dataset for the delineation of adult brain gliomas. The focus of BraTS2024 challenge remained the same in terms of generating the common benchmark environment, while the datasets expands into explicitly addressing 1) the same adult glioma population, as well as 2) the underserved sub-Saharan African brain glioma patient population, 3) brain/intracranial meningioma, 4) brain metastasis, and 5) pediatric brain tumor patients. Although segmentation is the most widely investigated medical image processing task, the various challenges

have been organized to focus only on specific clinical tasks. That is, each segmentation method was evaluated exclusively on the patients population it was trained on in each sub-challenge. In this challenge, we aim to organize the Generalizability Assessment of Segmentation Algorithms Across Brain Tumors. The hypothesis is that a method capable of performing well on multiple segmentation tasks will generalize well on unseen tasks. Specifically, in this task, we will be focusing on assessing the algorithmic generalizability beyond each individual patient population and focus across all of them. Importantly, although each MR exam will undergo the same preprocessing pipeline, including an intensity normalization step, there are characteristics of each exam that will not be affected (I.e., different number of lesions per exam, different location within the brain, etc.) preserving the generalizability aspect of the challenge.

Keywords

List the primary keywords that characterize the task.

Generalizability, Segmentation, Brain Tumors, Cancer, Challenge, NIH, DREAM

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Lead Organizers

Gian Marco Conte, MD, PhD
Mayo Clinic, Rochester, MN, USA

Ujjwal Baid, PhD
Indiana University

Spyridon Bakas, PhD
Indiana University

Associate organizing committee (alphabetical)

Mariam Aboian,
CHOP

Maruf Adewole,
Medical Artificial Intelligence (MAI) Lab, Crestview Radiology Ltd., Lagos, Nigeria

Jake Albrecht,
Sage Bionetworks

Udunna Anazodo,
McGill University, Montreal, Canada / MAI Lab

Evan Calabrese,
Duke University Medical Center

Verena Chung,
Sage Bionetworks

Anastasia Janas,
Yale University

Anahita Fathi Kazerooni,
Childrens Hospital of Philadelphia / University of Pennsylvania

Dominic Labella,
Duke University Medical Center

Marius George Linguraru,
Childrens National Hospital / George Washington University

Bjoern Menze
University of Zurich

Ahmed Moawad
Mercy Catholic Medical Center

Jeffrey Rudie

Scripps Clinic and University of California, San Diego

b) Provide information on the primary contact person.

Gian Marco Conte, PhD
[Lead Organizer of the "BraTS-GoAT: Generalizability Across (Pre-treated) Tumors"]
Mayo Clinic
Email: Conte.gianmarco@mayo.edu

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NIH (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this challenge.

The NIH takes special interest in the BraTS 2024 cluster of challenges and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through the ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are not allowed to use additional data neither from publicly available datasets nor their own institutions.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

The lead organizers of the challenge are in communication with 1) Intel and 2) Neosoma Inc, to sponsor monetary awards for the top 3 teams.
Formal confirmation can only be provided after the acceptance of the challenge.

Note that Intel has been offering monetary awards during each of BraTS 2018-2023, and Neosoma for BraTS 2021-2023.

NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author

- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the output of their methods to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it in unseen and permanently hidden (from future challenges too) testing data. The participants will be provided detailed guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility.

Submission of a containerized method will take place after the participants agree to release this container publicly available with either a CC-BY, or CC-BY-NC license as part of the challenge repository and meta-analysis manuscript.

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in May after the release of the training set, allowing participants to tune their methods in the unseen validation data . The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. However, only 2 submissions will be allowed in the final testing/ranking data/phase to avoid potential tuning of the submitted approach to the testing data.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period

- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training and Validation data release
Training data (with ground truth labels) and validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We have already released the training and validation data in TCIA following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions)

The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Googles AI resources. IDC data are also expected to be available through the AWS (Amazon
Web Services) Marketplace.

Informed consent was obtained from all subjects at their respective institutions, and the protocol for releasing the data was approved by the institutional review board of the data-contributing institution.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Generally Nuanced Deep Learning Framework (GaNDLF - https://github.com/mlcommons/GaNDLF).

Pati S, Thakur SP, Hamamc E, Baid U, Baheti B, Bhalerao M, et al. GaNDLF: the generally nuanced deep learning framework for scalable end-to-end clinical workflows. Communications Engineering. 2023;2(1):1-17.

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS 2021-2023 challenges.

All participants of the challenge they will be required to accept an agreement through the synapse.org website that participation in the testing phase of the challenge, will automatically mean that we can make their containerized method publicly available through our challenge webpage.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in several ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Spyridon Bakas, Ujjwal Baid, Gian Marco Conte, SAGE Bionetworks (synapse.org) will be the only ones who will have access to the validation, and test case ground truth labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Treatment planning, Intervention planning, Assistance, Research, Surgery, Training, Diagnosis, CAD, Education, Decision support.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients, diagnosed with a de novo brain tumor (i.e., glioma, meningioma, metastasis, pediatric).

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with brain tumor, clinically scanned with mpMRI acquisition protocol including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region class)

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Dice, Hausdorff 95th percentile, Sensitivity, Precision, Specificity. - per lesion evaluation

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in our related manuscripts [1,2,4]. Since then, multiple institutions have contributed data to create the current RSNA-ASNR-MICCAI BraTS dataset and these are listed in the latest BraTS arxiv paper [1]. We have made the complete BraTS 2021-2023 dataset permanently available through the TCIA portal. All the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, when available .

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript when available.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe mpMRI scans, acquired with different clinical protocols and various scanners as described in the previous tasks.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at a single timepoint. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There was no inclusion/exclusion criterion applied that related to 3d acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

Training data: 2,000+ cases
Validation data: 300+ cases

Testing data: 200+
Note: All the cases have been used in previous BraTS challenges.
We are currently working on extending the testing sets.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability. The data was split in these numbers between training, validation, and testing after considering the number of cases used as test cases in previous instances of BraTS and the fact that the organizers did not want to reveal ground truth labels of previous test cases, to avoid compromising ranking the participants.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

N/A

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Automated segmentations were generated using the FeTS tool and annotators corrected them to create ground truth dataset. These annotations were approved by radiologists.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The data considered in this task of the BraTS 2024 challenge follows the paradigm of the BraTS 2021-2023 challenge data. The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists), which was provided to all clinical annotators, describing in detail instructions on what the segmentations of each tumor sub-region should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions:
i) the enhancing tumor (when present) delineates the hyperintense signal of the T1-Gd, after excluding the vessels.
ii) the necrotic core (when present) outlines regions appearing dark in both T1 and T1-Gd images (denoting necrosis/cysts), and darked regions in T1-Gd that appear brighter in T1.
iii) the tumor core, which is the union of the enhancing tumor and the necrotic core described in (i) and (ii) above.
iv) the farthest tumor extent including the edema (what is called the whole tumor), delineates the tissue represented by the abnormal T2-FLAIR envelope.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers were the 2 experienced board-certified neuroradiologists (with >15 years of experience), listed in the Organizers section as clinical evaluators and annotation approvers. The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators were satisfied with the produced annotations, they were passing these to the corresponding approver. The approver is then responsible for signing off these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scans, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The exact preprocessing pipeline applied to all the data considered in the BraTS 2024 challenge is identical with the one evaluated and followed by the BraTS 2017-2023 challenges. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format [10], we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 [9]) and interpolating to the same resolution as this atlas (1 mm^3). The exact registration process comprises the following steps:

STEP 1: N4 Bias field correction (notably the application of N4 bias field correction is a temporary step. Taking into consideration we have previously [4] shown that use of non-parametric, non-uniform intensity normalization (i.e., N4) to correct for intensity non-uniformities caused by the inhomogeneity of the scanners magnetic field during image acquisition obliterates the MRI signal relating to the abnormal/tumor regions, we intentionally use N4 bias field correction in the preprocessing pipeline to facilitate a more optimal rigid registration across the different MRI sequences. However, after obtaining the related information (i.e., transformation matrices), we discard the bias field corrected scans, and we apply this transformation matrix towards the final co-registered output images used in the challenge).
STEP 2: Rigid Registration of T1, T2, T2-FLAIR to the T1-Gd scan, and obtain the corresponding transformation matrix.
STEP 3: Rigid Registration of T1-Gd scan to the SRI-24 atlas [9], and obtain the corresponding transformation matrix.
STEP 4: Join the obtained transformation matrices and applying aggregated transformation to the LPS-oriented scans.
STEP 5: After completion of the registration process, we perform brain extraction to remove any apparent nonbrain tissue (e.g., neck fat, skull, eyeballs) based on a deep-learning approach we developed in house, focusing on scans with apparent brain tumors and exhaustively evaluated it in both private and public

multi-institutional data [11]. We then manually assessed all scans for confirming the correct brain extraction (i.e., skull stripping), where the complete brain region is included, and all non-brain tissue is excluded. This whole pipeline, and its source code are available through the CaPTk [5-6](https://github.com/CBICA/CaPTk) and FeTS [7] (https://fetsai.github.io/Front-End/) platforms.

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [8] and is outside the scope of the BraTS 2024 challenge.

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise
Sensitivity, Lesionwise
Specificity, Lesionwise
Precision Lesionwise

The regions evaluated using these metrics describe the whole tumor, the tumor core, and the enhancing tumor (when present). Note that the tumor core includes the part of the tumor that is typically resected (i.e., enhancing, non-enhancing, and necrotic tumor), and the whole tumor describes all tumor sub-regions (i.e., tumor core and edema/invasion).

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of the assessed and evaluated tumor sub-regions:
i) the enhancing tumor describes the regions of active tumor and based on this, clinical practice characterizes the extent of resection.
ii) the tumor core (incl. the necrotic component) describes what is typically resected during a surgical procedure.
iii) the whole tumor as it defines the whole extent of the tumor, including the peritumoral edematous tissue and highly infiltrated area.

In terms of evaluation metrics, we use:

i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance,
ii) the 95% Hausdorff distance as opposed to standard HD, in order to avoid outliers havings too much weight,
iii) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or undersegment.
iv) Precision to complement the metric of Sensitivity (also known as recall).

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics & Health Data Science), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (0 for the DSC and the image diagonal for the HD).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team.

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2023, uncertainties in rankings will be assessed using permutational analyses [2]. Performance for the segmentation task will be assessed based on relative performance of each team on each tumor tissue class and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 7: BraTS-Augment: Evaluation of Augmentation Techniques for BraTS

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

In the broader machine learning community, the concept of Data Centric machine learning has emerged to improve the performance of models with more meaningful training data. Data augmentation has been shown to improve the robustness of machine learning models, but the types of augmentations that may be useful for biomedical imaging are unknown. Conventional challenges ask participants to submit a model for evaluation on test data. This data-centric challenge will invert the process, asking participants to submit a method to augment training data such that a baseline model will show improved robustness on new data. Participants will submit a container that will augment training data (while keeping the number of training cases fixed) from the RSNA-ASNR-MICCAI BraTS 2021 (which represents the BraTS 2023 GLIOMA) challenge such that a common baseline U-Net model architecture can be trained on the container output. The trained model will be evaluated on the BraTS 2023 GLIOMA test data for Dice coefficient and Hausdorff95 measures of accuracy, per lesion, with emphasis on the consistency across the test set cases. Top performing methods may offer insight to augmentation approaches that could be used to generate robust state-of-the-art segmentation models.
This challenge task will be promoted by Sage Bionetworks and PrecisionFDA, in consultation with the NCI/NIH, and the FDA Center for Devices and Radiological Health.

Keywords

List the primary keywords that characterize the task.

Segmentation, Augmentation, BraTS, Data Centric

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Jake Albrecht [Lead Organizer - Contact Person]
Affiliation: Sage Bionetworks

Elaine Johanson
precisionFDA

Spyridon Bakas
Indiana University

Zeke Meier

Booz Allen Hamilton

Weijie Chen
Center for Devices and Radiological Health, U.S. Food and Drug Administration

Nicholas Petrick
Center for Devices and Radiological Health, U.S. Food and Drug Administration

Berkman Sahiner
Center for Devices and Radiological Health, U.S. Food and Drug Administration

Keyvan Farahani
National Institutes of Health

Ujjwal Baid
Indiana University

Rong Chai
Sage Bionetworks

Verena Chung
Sage Bionetworks

Clinical Evaluators, Annotation Approvers, & AnnotationVolunteers:
The same 65 people that facilitated the RSNA-ASNR-MICCAI BraTS 2021 challenge.

Data Contributors
The RSNA-ASNR-MICCAI BraTS 2021 challenge data contributors
(not included here as the space provided does not allow sufficient characters to be entered.)

b) Provide information on the primary contact person.

Jake Albrecht, PhD
[Lead Organizer of the "BraTS-Augment: Evaluation of Augmentation Techniques for BraTS"]
Sage Bionetworks
Email: jake.albrecht@sagebionetworks.org

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NCI (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2023 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are not allowed to use additional data from publicly available datasets or their own institutions.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel and 2) Neosoma Inc, we are coordinating with them for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge.
Note that Intel has been offering monetary awards during each of BraTS 2018-2022, and Neosoma for BraTS 2021. NIH/NCI will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the augmentation training pipeline to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the submission they would like to be evaluated in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it in the three hidden testing data. The participants will be provided guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility, running of these algorithms to the previous BraTS instances and comparison with results obtained by algorithms of previous years, thereby maximizing solutions in solving the problem of brain tumor segmentation.

All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Tool [7] (https://fets-ai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).

[5] C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

[6] S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

[7] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in June after the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We have released the data in The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH), following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions). The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Googles AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Informed consent was obtained from all subjects at their respective institutions, and the protocol for releasing the data was approved by the institutional review board of the data-contributing institution.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Platform [7] (https://fets-ai.github.io/Front-End/).

[5] C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

[6] S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

[7] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS 2021-2023 challenges.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Spyridon Bakas, Ujjwal Baid, SAGE Bionetworks, and the clinical evaluators will have access to the validation, and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients, diagnosed with de novo diffuse gliomas of the brain.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with de novo diffuse gliomas of the brain, clinically scanned with mpMRI acquisition protocol including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region volumes)

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Dice, Hausdorff 95th percentile , Gini index
Additional points: Sensitivity, Precision, Specificity.

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in our related manuscripts [1,2,4]. Since then, multiple institutions have contributed data to create the current RSNA-ASNR-MICCAI BraTS dataset and these are listed in the latest BraTS arxiv paper [1]. We are currently in coordination with TCIA to make the complete BraTS 2021-2023 dataset permanently available through their portal. All the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, including Google and AWS Marketplaces, as part of their Public Datasets Programs.

[1] U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and

Radiogenomic Classification, arXiv preprint arXiv:2107.02314
[2] S.Bakas, 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
[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript.

The acquisition protocol is different for each different institution as these scans we use are representative of real clinical protocols. The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in the related manuscripts of ours [1,2,4]. Since then multiple institutions have contributed data to create the current BraTS dataset and these are listed in the latest BraTS arxiv paper [1]. We are currently in coordination with TCIA to make the complete BraTS dataset permanently available through them. All the acquisition details will be included together with the data availability in TCIA.

[1] U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314

[2] S.Bakas, 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

[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117
(2017) DOI: 10.1038/sdata.2017.117

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe mpMRI scans, acquired with different clinical protocols and various scanners from:
1. University of Pennsylvania (PA, USA),
2. University of Alabama at Birmingham (AL, USA),
3. Heidelberg University (Germany),
4. University of Bern (Switzerland),
5. University of Debrecen (Hungary),
6. Henry Ford Hospital (MI, USA),
7. University of California (CA, USA),
8. MD Anderson Cancer Center (TX, USA),

9. Emory University (GA, USA),
10. Mayo Clinic (MN, USA),
11. Thomas Jefferson University (PA, USA),
12. Duke University School of Medicine (NC, USA),
13. Saint Joseph Hospital and Medical Center (AZ, USA),
14. Case Western Reserve University (OH, USA),
15. University of North Carolina (NC, USA),
16. Fondazione IRCCS Instituto Neuroligico C. Besta, (Italy),
17. Ivy Glioblastoma Atlas Project,
18. MD Anderson Cancer Center (TX, USA),
19. Washington University in St. Louis (MO, USA),
20. Tata Memorial Center (India),
21. University of Pittsburg Medical Center (PA, USA),
22. University of California San Francisco (CA, USA),
23. Unity Health,
24. University Hospital of Zurich.

Note that data from institutions 6-17 are provided through The Cancer Imaging Archive (TCIA - http://www.cancerimagingarchive.net/), supported by the Cancer Imaging Program (CIP) of the National Cancer Institute (NCI) of the National Institutes of Health (NIH).

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at a single timepoint. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There was no inclusion/exclusion criterion applied that related to 3d acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

Training data: 1,251 cases
Validation data: 219 cases
Testing data: 570 cases
Note: All the cases have been used in previous BraTS challenges.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

N/A

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Reference approved from at least 2 experienced neuroradiologists, following annotations from 60 clinical neuroradiologists (volunteers from ASNR)

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The data considered in this task of the BraTS 2023 challenge follows the paradigm of the BraTS 2021-2022 challenge data. The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists), which was provided to all clinical annotators, describing in detail instructions on what the segmentations of each tumor sub-region should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions:
i) the enhancing tumor (when present) delineates the hyperintense signal of the T1-Gd, after excluding the vessels.

ii) the necrotic core (when present) outlines regions appearing dark in both T1 and T1-Gd images (denoting necrosis/cysts), and darked regions in T1-Gd that appear brighter in T1.

iii) the tumor core, which is the union of the enhancing tumor and the necrotic core described in (i) and (ii) above.

iv) the farthest tumor extent including the edema (what is called the whole tumor), delineates the tissue represented by the abnormal T2-FLAIR envelope.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers were the 2 experienced board-certified neuroradiologists (with >15 years of experience), listed in the Organizers section as clinical evaluators and annotation approvers. The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators were satisfied with the produced annotations, they were passing these to the corresponding approver. The approver is then responsible for signing off these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scans, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The exact preprocessing pipeline applied to all the data considered in the BraTS 2023 challenge is identical with the one evaluated and followed by the BraTS 2017-2022 challenges. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format [10], we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 [9]) and interpolating to the same resolution as this atlas (1 mm^3). The exact registration process comprises the following steps:

STEP 1: N4 Bias field correction (notably the application of N4 bias field correction is a temporary step. Taking into consideration we have previously [4] shown that use of non-parametric, non-uniform intensity normalization (i.e., N4) to correct for intensity non-uniformities caused by the inhomogeneity of the scanners magnetic field during image acquisition obliterates the MRI signal relating to the abnormal/tumor regions, we intentionally use N4 bias field correction in the preprocessing pipeline to facilitate a more optimal rigid registration across the difference MRI sequences. However, after obtaining the related information (i.e., transformation matrices), we discard the bias field corrected scans, and we apply this transformation matrix towards the final co-registered output images used in the challenge).

STEP 2: Rigid Registration of T1, T2, T2-FLAIR to the T1-Gd scan, and obtain the corresponding transformation matrix.

STEP 3: Rigid Registration of T1-Gd scan to the SRI-24 atlas [9], and obtain the corresponding transformation matrix.

STEP 4: Join the obtained transformation matrices and applying aggregated transformation to the LPS-oriented scans.

STEP 5: After completion of the registration process, we perform brain extraction to remove any apparent nonbrain tissue (e.g., neck fat, skull, eyeballs) based on a deep-learning approach we developed in house, focusing on scans with apparent brain tumors and exhaustively evaluated it in both private and public multi-institutional data [11]. We then manually assessed all scans for confirming the correct brain extraction (i.e., skull stripping), where the complete brain region is included, and all non-brain tissue is excluded.

This whole pipeline, and its source code are available through the CaPTk [5-6](https://github.com/CBICA/CaPTk) and FeTS [7] (https://fets-ai.github.io/Front-End/) platforms.

[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117, 2017. DOI: 10.1038/sdata.2017.117

[9] T. Rohlfing, et al. The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798-819, 2010.

[10] R.Cox, J.Ashburner, H.Breman, K.Fissell, C.Haselgrove, C.Holmes, J.Lancaster, D.Rex, S.Smith, J.Woodward, A (Sort of) new image data format standard: NIfTI-1: WE 150, Neuroimage, 22, 2004.

[11] S.Thakur, J.Doshi, S.Pati, S.Rathore, C.Sako, M.Bilello, S.M.Ha, G.Shukla, A.Flanders, A.Kotrotsou, M.Milchenko, S.Liem, G.S.Alexander, J.Lombardo, J.D.Palmer, P.LaMontagne, A.Nazeri, S.Talbar, U.Kulkarni, D.Marcus, R.Colen, C.Davatzikos, G.Erus, S.Bakas, 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. DOI: 10.1016/j.neuroimage.2020.117081

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [8] and is outside the scope of the BraTS 2022 challenge.

[8] R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise
Sensitivity, Lesionwise
Specificity, Lesionwise
Precision Lesionwise

The regions evaluated using these metrics describe the whole tumor, the tumor core, and the enhancing tumor (when present). Note that the tumor core includes the part of the tumor that is typically resected (i.e., enhancing, non-enhancing, and necrotic tumor), and the whole tumor describes all tumor sub-regions (i.e., tumor core and edema/invasion).

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of the assessed and evaluated tumor sub-regions:
i) the enhancing tumor describes the regions of active tumor and based on this, clinical practice characterizes the extent of resection. ii) the tumor core (incl. the necrotic component) describes what is typically resected during a surgical procedure. iii) the whole tumor as it defines the whole extent of the tumor, including the peritumoral edematous tissue and highly infiltrated area.

In terms of evaluation metrics, we use:
i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance,
ii) the 95% Hausdorff distance as opposed to standard HD, in order to avoid outliers havings too much weight,
iii) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or undersegment. va
iv) Precision to complement the metric of Sensitivity (also known as recall).
v) Gini index to measure case-wise distribution for DSC and HD95

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics & Health Data Science), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (0 for the DSC and the image diagonal for the HD).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the DSC, HD95, and Gini metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot [6].

[12] Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2023, uncertainties in rankings will be assessed using permutational analyses [3]. Performance for the segmentation task will be assessed based on relative performance of each team on each tumor tissue class and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

[2] S. Bakas et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, arXiv:1811.02629 [cs, stat], Apr. 2019, Accessed: Dec. 10, 2020. [Online]. Available: http://arxiv.org/abs/1811.02629.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 8: BraTS-Synthesis: MR Image Synthesis for BraTS

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Manual segmentation of brain tumors in MR images is a tedious task with high variability among raters [1]. Many recent works have developed automated segmentation methods using deep learning (DL) [2 4] to address this issue. These algorithms mostly require four input magnetic resonance imaging (MRI) modalities (typically T1 weighted [T1w] images with and without contrast enhancement, T2 weighted [T2w] images, and FLAIR images) during the inference stage. However, in clinical routine, missing MR sequences, e.g., because of time constraints and/or image artifacts (such as patient motion) are a common challenge. Some sequences, especially FLAIR and T1, are often missing from routine MRI examinations [5]. Therefore, the substitution of missing modalities is desirable and necessary for a more widespread use of such algorithms in clinical routine.

This task, following the initial setup at BraTS 2023, again calls for algorithms capable of substituting whole MRI volumes, enabling a straightforward application of BraTS segmentation networks in centers with a less extensive imaging protocol or for analyzing historical tumor study datasets. The task of generating missing MRI sequences holds promise to address this issue and has attracted growing attention in recent years [6-7]. For example, deep learning networks based on generative adversarial networks (GANs) have been explored for this task with promising results [8-10]. From a technical standpoint, these algorithms need to overcome a multitude of challenges: First, the image resolutions of the individual sequences might differ; for example, FLAIR images tend to be acquired using 2D sequences, leading to anisotropic resolution, matching the resolution of other 3D imaging sequences only poorly. Second, motion artifacts may be presented in some of the sequences. At the same time, MRI bias fields may differ in their local impact on the different image modalities, leading to spatially inconstant artifacts. Third, a general domain shift between the training and test sets due to different acquisition settings and types of scanners can be expected to be present in almost any large and multi institutional dataset. All these effects must be considered when developing methods for synthesizing volumetric MRI. Questions about how to deal with these challenges, for example, by choosing adequate metrics or invariance properties of the algorithms and network architecture, have yet to be answered.

In previous BraTS challenges, we have set up publicly available datasets and algorithms for multi modal brain glioma segmentation [11-12]. In MRI synthesis task, we will build on these efforts, and the previously generated data sets, to further the development of much needed computational tools for data integration and homogenization. It will enable a broader application of the tumor segmentation algorithms developed in previous BraTS editions (that require a fixed set of image modalities). The resulting MRI synthesis is essential to develop effective, generalizable, and reproducible methods for analyzing high resolution MRI of brain tumors. It will include data from multiple sites well established in previous BraTS challenges, adding new inference tasks beyond glioma data. Compared to BraTS 2023, we will additionally evaluate the containerized algorithms on brain metastasis datasets [8] to test their generalizability.

Keywords

List the primary keywords that characterize the task.

Synthesis, Segmentation, Brain Tumor

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Hongwei Bran Li, [Lead Organizer]
Harvard Medical School

Benedikt Wiestler,
Technical University of Munich

Juan Eugenio Iglesias,
Harvard Medical School

Syed Muhammad Anwar,
George Washington University

Marius George Linguraru,
Children's National Hospital

Bjoern Menze
University of Zurich

Koen Van Leemput
Harvard Medical School

Florian Kofler
Helmholtz Research Center

Marie Piraud
Helmholtz Research Center

Spyridon Bakas
Indiana University

Ujjwal Baid
Indiana University

Jake Albrecht
Sage Bionetworks

Keyvan Farahani
NIH

Verena Chung
Sage Bionetworks

Gian Marco Conte
Mayo Clinic

Clinical Evaluators, Annotation Approvers, & AnnotationVolunteers:
The same 65 people that facilitated the RSNA-ASNR-MICCAI BraTS 2021 challenge.

Data Contributors
The RSNA-ASNR-MICCAI BraTS 2021 challenge data contributors
(not included here as the space provided does not allow sufficient characters to be entered.)

b) Provide information on the primary contact person.

Hongwei "Bran" Li, PhD
[Lead Organizer of the "BraTS-Synthesis: MR Image Synthesis for BraTS"]
Harvard Medical School
Email: holi2@mgh.harvard.edu

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NCI (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data, but if they do so, they MUST also discuss the potential difference in their results after using only the BraTS 2024 data, since our intention is to solve the particular segmentation problem, but also to provide a fair comparison among the participating methods.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel and 2) Neosoma Inc, we are coordinating for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge.

Note that Intel has been offering monetary awards during each of BraTS 2018-2023, and Neosoma for BraTS 2021.
NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants must send their methods' output to the evaluation platform for the scoring to occur during the training and validation phases. At the end of the validation phase, the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase and will have to submit a containerized implementation that matches the design requirements of the evaluation platform. The participants will be provided guidelines on the form of the container, as we have done in previous years.
During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk) and the Federated Tumor Segmentation (FeTS) Tool [7] (https://fets-ai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).
The organizers will confirm receiving the containerized method at the start of the testing phase and evaluate it on the hidden testing data. This will enable confirmation of reproducibility, running of these algorithms to the previous BraTS instances, and using them for additional experiments to be reported in the post-conference

journal paper.

All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

[5] C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

[6] S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

[7] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

[8] Moawad, Ahmed W., et al. The Brain Tumor Segmentation (BraTS-METS) Challenge 2023: Brain Metastasis Segmentation on Pre-treatment MRI." ArXiv (2023).

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in April together with the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We have uploaded the training and validation data to The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH)following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions).
The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Googles AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Cancer Imaging Phenomics Toolkit (CaPTk [15-16], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Platform [17] (https://fets-ai.github.io/Front-End/).

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants must submit their containerized algorithm during or after the validation phase. Specific instructions for containerization will be provided after the challenge approval. These instructions will be very similar to what we requested participants to provide during the BraTS 2022 and 2023 challenges. The organizers will keep the containers and use them in follow-up research related to the BRATS challenge, for example, when applying new testing data available in the later BRATS challenge to enable a direct comparison of performances across the different annual editions of the BRATS challenge.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Spyridon Bakas, Ujjwal Baid, SAGE Bionetworks, and the organization team will have access to the validation, and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Synthesis, Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients, diagnosed with de novo diffuse gliomas of the brain.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with de novo diffuse gliomas of the brain, clinically scanned with mpMRI acquisition protocol including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region volumes)

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.

- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Participants have to generate a full image volume that corresponds to the one missing image modality (e.g., it will be one of T1w / T2w / T1c / FLAIR). Results will be evaluated regarding the accuracy of the downstream brain tumor image segmentation using Dice scores and 95th percentile Hausdorff distance, per lesion. We will implement a BraTS algorithm (the UNet pre-trained in the FETS brain tumor segmentation initiative [17]). The same algorithm will be used to evaluate the hidden test data. Segmentation rankings and image similarity rankings will be combined using statistical methods similar to the metric fusion approaches of previous BraTS

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in our related manuscripts [11,12,14]. Since then, multiple institutions have contributed data to create the current RSNA-ASNR-MICCAI BraTS dataset and these are listed in the latest BraTS arxiv paper [11].

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe mpMRI scans, acquired with different clinical protocols and various scanners from the segmentation task of BraTS 2022.

1. University of Pennsylvania (PA, USA),
2. University of Alabama at Birmingham (AL, USA),
3. Heidelberg University (Germany),
4. University of Bern (Switzerland),
5. University of Debrecen (Hungary),
6. Henry Ford Hospital (MI, USA),
7. University of California (CA, USA),
8. MD Anderson Cancer Center (TX, USA),
9. Emory University (GA, USA),
10. Mayo Clinic (MN, USA),
11. Thomas Jefferson University (PA, USA),
12. Duke University School of Medicine (NC, USA),

13. Saint Joseph Hospital and Medical Center (AZ, USA),
14. Case Western Reserve University (OH, USA),
15. University of North Carolina (NC, USA),
16. Fondazione IRCCS Instituto Neuroligico C. Besta, (Italy),
17. Ivy Glioblastoma Atlas Project,
18. MD Anderson Cancer Center (TX, USA),
19. Washington University in St. Louis (MO, USA),
20. Tata Memorial Center (India),
21. University of Pittsburg Medical Center (PA, USA),
22. University of California San Francisco (CA, USA),
23. Unity Health,
24. University Hospital of Zurich.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at a single timepoint. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI.
Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There was no inclusion/exclusion criterion applied that related to 3d acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

Here we will focus using the RSNA-ASNR-MICCAI BraTS 2021 dataset and the test set from BraTS-METS [8]:

Training data: 1,251 cases from RSNA-ASNR-MICCAI BraTS 2021
Validation data: 219 cases from RSNA-ASNR-MICCAI BraTS 2021
Testing data: 570 cases from RSNA-ASNR-MICCAI BraTS 2021 + 65 cases from BraTS-METS.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

All four MRI sequences and the segmentation map will be available in the training data. In the validation and test sets, one modality out of four sequences in each case will be randomly dropped to evaluate the performance of submitted image synthesis methods.

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

The missing modalities will be chosen randomly for each subject

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

Key annotation of all BRATS image data sets is the tumor annotation for this task.
The tumor image annotation follows the paradigm of the BraTS 2021 challenge data. The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists), which was provided to all clinical annotators, describing in detail instructions on what the segmentations of each tumor sub-region should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions:
i) the enhancing tumor (when present) delineates the hyperintense signal of the T1-Gd, after excluding the vessels.
ii) the necrotic core (when present) outlines regions appearing dark in both T1 and T1-Gd images (denoting necrosis/cysts), and darked regions in T1-Gd that appear brighter in T1.
iii) the tumor core, which is the union of the enhancing tumor and the necrotic core described in (i) and (ii) above.
iv) the farthest tumor extent including the edema (what is called the whole tumor), delineates the tissue represented by the abnormal T2-FLAIR envelope.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers were the 2 experienced board-certified neuroradiologists (with >15 years of experience), listed in the Organizers section as clinical evaluators and annotation approvers. The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators were satisfied with the produced annotations, they were passing these to the corresponding approver. The approver is then responsible for signing off these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scans, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The exact preprocessing pipeline applied to all the data considered in the BraTS 2024 challenge is identical with the one evaluated and followed by the BraTS 2017-2024 challenges. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format [10], we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 [19]) and interpolating to the same resolution as this atlas (1 mm^3). The exact registration process comprises the following steps:

STEP 1: N4 Bias field correction (notably the application of N4 bias field correction is a temporary step. Taking into consideration we have previously [14] shown that use of non-parametric, non-uniform intensity normalization (i.e., N4) to correct for intensity non-uniformities caused by the inhomogeneity of the scanners magnetic field during image acquisition obliterates the MRI signal relating to the abnormal/tumor regions, we intentionally use N4 bias field correction in the preprocessing pipeline to facilitate a more optimal rigid registration across the difference MRI sequences. However, after obtaining the related information (i.e., transformation matrices), we discard the bias field corrected scans, and we apply this transformation matrix towards the final co-registered output images used in the challenge).

STEP 2: Rigid Registration of T1, T2, T2-FLAIR to the T1-Gd scan, and obtain the corresponding transformation matrix.

STEP 3: Rigid Registration of T1-Gd scan to the SRI-24 atlas [19], and obtain the corresponding transformation matrix.

STEP 4: Join the obtained transformation matrices and applying aggregated transformation to the LPS-oriented scans.

STEP 5: After completion of the registration process, we perform brain extraction to remove any apparent non-brain tissue (e.g., neck fat, skull, eyeballs) based on a deep-learning approach we developed in house, focusing on scans with apparent brain tumors and exhaustively evaluated it in both private and public multi-institutional data [21]. We then manually assessed all scans for confirming the correct brain extraction (i.e., skull stripping), where the complete brain region is included, and all non-brain tissue is excluded.

This whole pipeline, and its source code are available through the CaPTk [15-16](https://github.com/CBICA/CaPTk) and FeTS [17] (https://fets-ai.github.io/Front-End/) platforms.

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [18] and is outside the scope of the BraTS 2024 challenge.

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Dice Similarity Coefficient (DSC), Lesionwise
95% Hausdorff distance (HD), Lesionwise
Sensitivity, Lesionwise
Specificity, Lesionwise
Precision Lesionwise

For glioma test set, the automated segmentation will be performed by the final FeTS algorithm [17]. For the test set from METS-BRATS [8], we will use its winner solution. The regions evaluated with the two segmentation metrics describe the whole tumor, the tumor core, and the enhancing tumor (when present). Note that the tumor core includes the part of the tumor that is typically resected (i.e., enhancing, non-enhancing, and necrotic tumor), and the whole tumor describes all tumor sub-regions (i.e., tumor core and edema/invasion).
The Structural similarity Index (SSIM) is used to evaluate the quality of brain structures in synthetic images, i.e. to

compare synthetic sequences with their physically acquired counterparts, as does the L2 norm distance.

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of the assessed and evaluated three tumor sub-regions:
i) the enhancing tumor describes the regions of active tumor and based on this, clinical practice characterizes the extent of resection.
ii) the tumor core (incl. the necrotic component) describes what is typically resected during a surgical procedure.
iii) the whole tumor as it defines the whole extent of the tumor, including the peritumoral edematous tissue and highly infiltrated area.

In terms of evaluation metrics, we use:
i) the Dice Similarity Coefficient, which is commonly used in the assessment of segmentation performance,
ii) the 95th percentile Hausdorff distance as a complementary metric of overlap-based metric. 95th percentile is chose as opposed to standard HD, in order to avoid outliers having too much weight,
iii) the structural similarity index, which is commonly perceptual metric to quantify image similarity between synthetic images and reference images.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics & Health Data Science), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case (e.g., inconsistent image dimensions between the generated one and reference one), this metric will be set to its worst possible value (0 for the DSC, the image diagonal for the HD and 0 for structural similarity index).

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the weighted average of the metrics described above as a univariate overall summary measure. Notably, since we focus on image segmentation, the Dice and 95p HD. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot [16].

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2023, uncertainties in rankings will be assessed using permutational analyses [13]. Performance for the segmentation task will be assessed based on relative performance of each team on each tumor tissue class and for each segmentation measure. These will be combined by averaging ranks for the measures, and statistical significance will be evaluated only for the segmentation performance measures and will be quantified by permuting the relative ranks for each segmentation measure and tissue class per subject of the testing data.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

TASK 9: BraTS-Inpainting: MR Image Inpainting for BraTS

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

The challenges task is to inpaint healthy tissue in partially broken MRI scans. Reasons for this are often technical: there may be presence of locally isolated artifacts, incompleteness of the field of view, or corrupted/missing 2D slices. For such cases, one may want to inpaint missing information locally instead of inferring the corrupted image volume completely. Therefore, our call is for algorithms capable of inpainting corrupted image intensities within a given inpainting mask. Like in the global image synthesis challenge, this will enable the application of the downstream image processing routines. For example, brain parcellation algorithms strictly require the input of normal-appearing images, which is used in neuro-imaging studies and in brain tumor treatment planning. From a technical standpoint, these algorithms need to overcome a multitude of challenges that also apply to the global synthesis challenge:

First, the image resolutions of the individual sequences might differ; for example, FLAIR images tend to be acquired using 2D sequences, leading to anisotropic resolution, matching the resolution of other 3D imaging sequences only poorly.

Second, motion artifacts may be presented in some of the sequences. At the same time, MRI bias fields may differ in their local impact on the different image modalities, leading to spatially inconstant artifacts.

Third, a general domain shift between the training and test sets due to different acquisition settings and types of scanners can be expected to be present in almost any large and multi-institutional dataset.

All these effects must be considered when developing methods for synthesizing MRI locally and globally. Questions about how to deal with these challenges, for example, by choosing adequate metrics or invariance properties of the algorithms and network architecture, have yet to be answered. In previous BraTS challenges, we have set up publicly available datasets and algorithms for multi-modal brain glioma segmentation [11-12]. In our challenge task for MRI synthesis, we will build on these efforts, and the previously generated data sets, to further the development of much-needed computational tools for data integration and homogenization. It will enable a better integration with other downstream routines used for quantitative neuro-image analysis (that only work well for brain images without perturbations from artifacts or lesion). The resulting MRI synthesis is essential to develop effective, generalizable, and reproducible methods for analyzing high-resolution MRI of brain tumors. It will include data from multiple sites well established in previous BraTS challenges, adding new inference tasks beyond image segmentation. Resulting algorithms will have the potential to benefit automated brain (tumor) image processing and improve the clinical risk stratification tools for early interventions, treatments, and care management decisions across hospitals and research institutions worldwide.

Difference to 2023 inpainting challenge

Given the positive feedback from participants and clinicians the inpainting challenge would continue with minor refinements. The overall challenge design and evaluation metrics remain the same.

Keywords

List the primary keywords that characterize the task.

Inpainting, Synthesis, Infill, Segmentation, Brain Tumor

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Florian Kofler [Lead Organizer]
Helmholtz Research Center

Hongwei Bran Li
Harverd Medical School

Benedikt Wiestler,
Technical University of Munich

Juan Eugenio Iglesias
Harvard Medical School

Syed Muhammad Anwar
Childrens National Hospital

Marius George Linguraru
Childrens National Hospital

Bjoern Menze
University of Zurich

Koen Van Leemput
Harvard Medical School

Marie Piraud
Helmholtz Research Center

Spyridon Bakas
Indiana University

Ujjwal Baid
Indiana University

Jake Albrecht
Sage Bionetworks

Keyvan Farahani
NIH

Verena Chung
Sage Bionetworks

Gian Marco Conte
Mayo Clinic

Clinical Evaluators, Annotation Approvers, & AnnotationVolunteers:
The same 65 people that facilitated the RSNA-ASNR-MICCAI BraTS 2021 challenge.

Data Contributors
The RSNA-ASNR-MICCAI BraTS 2021 challenge data contributors
(not included here as the space provided does not allow sufficient characters to be entered.)

b) Provide information on the primary contact person.

Florian Kofler, PhD
[Lead Organizer of the "BraTS-Inpainting: MR Image Inpainting for BraTS"]
Helmholtz Munich
Email: florian.kofler@helmholtz-munich.de

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NCI (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data, but if they do so, they MUST also discuss the potential difference in their results after using only the BraTS 2023 data, since our intention is to solve the particular inpainting problem, but also to provide a fair comparison among the participating methods.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Following communication with 1) Intel and 2) Neosoma Inc, we are coordinating for the sponsorship of monetary awards for the top 3 teams. Formal confirmation can only be provided after the acceptance of the challenge.

Note that Intel has been offering monetary awards during each of BraTS 2018-2023, and Neosoma for BraTS 2021-2023.
NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants must send their methods' output to the evaluation platform for the scoring to occur during the training and validation phases. At the end of the validation phase, the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase and will have to submit a dockerized implementation that matches the design requirements of the evaluation platform. The participants will be provided guidelines on the form of the container, as we have done in previous years.
During the training and validation phases, the participants will have the chance to test the functionality of their submission through both the Cancer Imaging Phenomics Toolkit (CaPTk [5-6], https://github.com/CBICA/CaPTk) and the Federated Tumor Segmentation (FeTS) Tool [7] (https://fetsai.github.io/Front-End/) that offer the implementation of the evaluation metrics, as well as via the online evaluation platform (Synapse).
The organizers will confirm receiving the containerized method at the start of the testing phase and evaluate it on the hidden testing data. This will enable confirmation of reproducibility, running of these algorithms to the

previous BraTS instances, and using them for additional experiments to be reported in the post-conference journal paper.
All participants of the challenge will be required to accept an agreement through the synapse.org website that participation in the testing phase will automatically mean that the organizers can make the submitted containerized method publicly available through our challenge webpage.

[5] C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018

[6] S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-46643-5_38

[7] S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the validation set in April together with the training set, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. Only 2 submissions will be allowed in the final testing/ranking data/phase.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training and Validation data release
Training data (with ground truth labels) and validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

We have uploaded all data in The Cancer Imaging Archive (TCIA) and the Imaging Data Commons (IDC) of the National Institutes of Health (NIH), following their standard licensing (https://wiki.cancerimagingarchive.net/display/Public/Data+Usage+Policies+and+Restrictions). The cloud-based IDC is routinely updated with new collections from TCIA. IDC public collections are now part of the Google Public Datasets Program. This will effectively make all the BraTS data available in the Google Marketplace, increasing the potential for access to the data and downstream AI developments using Googles AI resources. IDC data are also expected to be available through the AWS (Amazon Web Services) Marketplace.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)

- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY.

Additional comments: CC-BY, but if any of the non-TCIA contributors object to this license, the specific subset of the BraTS data will be released under a CC-BY-NC license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Cancer Imaging Phenomics Toolkit (CaPTk [15-16], https://github.com/CBICA/CaPTk), and the Federated Tumor Segmentation (FeTS) Platform [17] (https://fetsai.github.io/Front-End/).

Further, the inpainting metrics are available via a Python package: https://pypi.org/project/inpainting/
The ranking is computed using challengeR https://github.com/wiesenfa/challengeR

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants must submit their containerized algorithm during or after the validation phase. Specific instructions for containerization will be provided after the challenge approval. These instructions will be very similar to what we requested participants to provide during the BraTS 2021-2023 challenges. The organizers will keep the containers and use them in follow-up research related to the BRATS challenge, for example, when applying new testing data available in the later BRATS challenge to enable a direct comparison of performances across the different annual editions of the BRATS challenge.
The National Cancer Institute takes particular interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways to make the submitted algorithms available to a broader public as well. Dr. Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repositories such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Spyridon Bakas, Ujjwal Baid, SAGE Bionetworks, and the organization team will have access to the validation, and test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation

- Tracking

Synthesis, Segmentation

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients, diagnosed with de novo diffuse gliomas of the brain.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with de novo diffuse gliomas of the brain, clinically scanned with mpMRI acquisition protocol including i) pre-contrast and ii) contrast-enhanced T1-weighted, iii) T2weighted and iv) T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) MRI.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Multi-parametric MRI scans of the brain, including T1w, T2w, FLAIR T2, contrast-enhanced T1w images.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region volumes)

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain mpMRI scans.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

In the local inpainting task, participants must generate image intensities of healthy-appearing voxels that are locally voided (covering lesions or a local artifacts). Outside of these voided area(s), the full information is available. Results will be evaluated in terms of structural similarity, peak signal to noise ration and root mean square error (residual) of the image synthesized for the inpainted area and the real image. As the task is to fill in healthy appearing images, the inpainting areas of the evaluation will be localized outside of the tumor. (Unlike glioma segmentation algorithms in the global synthesis task, there is no consensus on downstream brain parcellation tasks and algorithms. To this end, we will compare brain parcellation results only in the post-challenge result analysis, and it will not contribute to the ranking.) Similarity and residual intensity-based rankings will be combined using statistical methods similar to the metric fusion approaches of previous BraTS editions.

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The exact scanners and their technical specifications used for acquiring the TCIA cohort has been listed in the data reference published in our related manuscripts [11,12,14]. Since then, multiple institutions have contributed data to create the current RSNA-ASNR-MICCAI BraTS dataset and these are listed in the latest BraTS arxiv paper [11]. We are currently in coordination with TCIA to make the complete BraTS 2021-2024 dataset permanently available through their portal. All the acquisition details will be included together with the data availability in TCIA, and subsequently in IDC, including Google and AWS Marketplaces, as part of their Public Datasets Programs.

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The acquisition protocols are different across (and within each) contributing institution, as these represent scans of real routine clinical practice. Specific details (e.g., echo time, repetition time, original acquisition plane) of each scan of each patient will be published as supplementary material together with the challenge meta-analysis manuscript.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe mpMRI scans, acquired with different clinical protocols and various scanners from the segmentation task of BraTS 2022.

1. University of Pennsylvania (PA, USA),
2. University of Alabama at Birmingham (AL, USA),
3. Heidelberg University (Germany),
4. University of Bern (Switzerland),
5. University of Debrecen (Hungary),
6. Henry Ford Hospital (MI, USA),
7. University of California (CA, USA),
8. MD Anderson Cancer Center (TX, USA),
9. Emory University (GA, USA),
10. Mayo Clinic (MN, USA),
11. Thomas Jefferson University (PA, USA),
12. Duke University School of Medicine (NC, USA),
13. Saint Joseph Hospital and Medical Center (AZ, USA),
14. Case Western Reserve University (OH, USA),
15. University of North Carolina (NC, USA),
16. Fondazione IRCCS Instituto Neuroligico C. Besta, (Italy),
17. Ivy Glioblastoma Atlas Project,
18. MD Anderson Cancer Center (TX, USA),
19. Washington University in St. Louis (MO, USA),
20. Tata Memorial Center (India),
21. University of Pittsburg Medical Center (PA, USA),
22. University of California San Francisco (CA, USA),
23. Unity Health,
24. University Hospital of Zurich.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

People involved in MRI acquisition for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case describes multi-parametric MRI scans for a single patient at a single timepoint. The exact scans included for one case are i) unenhanced and ii) contrast-enhanced T1-weighted, iii) T2-weighted and iv) T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI.
Please note that all sequences included for each case of the provided dataset, represent the sequences with the best image quality available in the acquiring institution for this particular case. There was no inclusion/exclusion criterion applied that related to 3d acquisitions, or the exact type of pulse sequence (for example MPRAGE). We, instead, accepted all types of T1 acquisitions (with the exception of T1 FLAIR, as we did not want to mix the fluid suppressed values with non-flair scans) and then we applied the harmonized preprocessing protocol we have been using in BraTS, across the complete data. This preprocessing ensures all scans have 3D representations on a specific resolution (1mm^3), and aligned to the same anatomical atlas.

b) State the total number of training, validation and test cases.

Here we will focus only on using the RSNA-ASNR-MICCAI BraTS 2021 dataset:
Training data: 1,251 cases
Validation data: 219 cases
Testing data: 570 cases

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

For the local inpainting task, only the T1 sequence will be used. Further, we provide masks to specify which regions need to be inpainted. These masks will cover all tumor areas, in addition similar sized and shaped inpainting masks will be provided for healthy brain areas to compute evaluation metrics. We provide two sets of masks `unhealthy` and `healthy` ones to enable supervised training.
During test time, inpainting areas are voided, i.e., all image intensities inside the infill areas will be set to a predefined value.

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

We designed an algorithm that samples inpainting masks in the healthy brain area. These areas are similar in size and shape to tumor areas. The code for obtaining these is publicly available:
https://github.com/BraTS-inpainting/2023_challenge/blob/main/dataset/dataset_generation.ipynb

To curate the test set these masks were curated by trained experts.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The BraTS glioma segmentation labels follow the following instructions:
Key annotation of all BRATS image data sets is the tumor annotation for this task.
The tumor image annotation follows the paradigm of the BraTS 2021 challenge data. The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuroradiologists), which was provided to all clinical annotators, describing in detail instructions on what the segmentations of each tumor subregion should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements.

Summary of specific instructions:
i) the enhancing tumor (when present) delineates the hyperintense signal of the T1-Gd, after excluding the vessels.
ii) the necrotic core (when present) outlines regions appearing dark in both T1 and T1-Gd images (denoting necrosis/cysts), and darked regions in T1-Gd that appear brighter in T1. iii) the tumor core, which is the union of the enhancing tumor and the necrotic core described in (i) and (ii) above. iv) the farthest tumor extent including the edema (what is called the whole tumor), delineates the tissue
represented by the abnormal T2-FLAIR envelope.

These glioma segmentations serve as input for our algorithm (see above) and are then curated by trained experts to make sure they do not contain other pathologies.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers were the 2 experienced board-certified neuroradiologists (with >15 years of experience), listed in the Organizers section as clinical evaluators and annotation approvers. The annotators were given the flexibility to use their tool of preference for making the annotations, and also follow either a complete manual annotation approach, or a hybrid approach where an automated approach is used to produce some initial annotations followed by their manual refinements. Once the annotators were satisfied with the produced annotations, they were passing these to the corresponding approver. The approver is then responsible for signing off these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding MRI scans, and if the annotations were not of satisfactory quality they would be sent back to the annotators for further refinements. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these scans.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The exact preprocessing pipeline applied to all the data considered in the BraTS 2023 and 2024 challenge is identical with the one evaluated and followed by the BraTS 2017-2022 challenges. Specifically, following the conversion of the raw scans from their original DICOM file format to NIfTI file format [10], we first perform a re-orientation of all input scans (T1, T1- Gd, T2, T2-FLAIR) to the LPS/RAI orientation, and then register all of them to the same anatomical atlas (i.e., SRI-24 [19]) and interpolating to the same resolution as this atlas (1 mm^3). The exact registration process comprises the following steps:

STEP 1: N4 Bias field correction (notably the application of N4 bias field correction is a temporary step. Taking into consideration we have previously [14] shown that use of non-parametric, non-uniform intensity normalization (i.e., N4) to correct for intensity non-uniformities caused by the inhomogeneity of the scanners magnetic field during image acquisition obliterates the MRI signal relating to the abnormal/tumor regions, we intentionally use N4 bias field correction in the preprocessing pipeline to facilitate a more optimal rigid registration across the difference MRI sequences. However, after obtaining the related information (i.e., transformation matrices), we discard the bias field corrected scans, and we apply this transformation matrix towards the final co-registered output images used in the challenge).

STEP 2: Rigid Registration of T1, T2, T2-FLAIR to the T1-Gd scan, and obtain the corresponding transformation matrix.

STEP 3: Rigid Registration of T1-Gd scan to the SRI-24 atlas [19], and obtain the corresponding transformation matrix.

STEP 4: Join the obtained transformation matrices and applying aggregated transformation to the LPS-oriented scans.

STEP 5: After completion of the registration process, we perform brain extraction to remove any apparent nonbrain tissue (e.g., neck fat, skull, eyeballs) based on a deep-learning approach we developed in house, focusing on scans with apparent brain tumors and exhaustively evaluated it in both private and public multi-institutional data [21]. We then manually assessed all scans for confirming the correct brain extraction (i.e., skull stripping), where the complete brain region is included, and all non-brain tissue is excluded.

This whole pipeline, and its source code are available through the CaPTk [15-16](https://github.com/CBICA/CaPTk) and FeTS [17] (https://fets-ai.github.io/Front-End/) platforms.

Step 6: Run the inpainting dataset generation tool, available here:
https://github.com/BraTS-inpainting/2023_challenge/blob/main/dataset/dataset_generation.ipynb

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Study and evaluation of the effect of this error is addressed by the uncertainty task of BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [18] and is outside the scope of the BraTS 2022 challenge.

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

The Structural similarity Index (SSIM) is used to evaluate the quality of brain structures in synthetic images, i.e. to compare synthetic sequences with their physically acquired counterparts, as does the L2 norm distance and Peak Signal to noise ratio (PSNR).

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

We choose the three most popular metrics to evaluate image synthesis methods:
i) RMSE (root mean square error)
ii) PSNR (Peak Signal to Noise Ratio)
iii) Structural similarity index (SSIM) which is commonly perceptual metric to quantify image similarity between synthetic images and reference images.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

As the metrics work on different scales we want to aggregate in scale-agnostic fashion. Further, we want to assign equal weight to all cases. The ranking scheme was developed in collaboration with Dr. Annika Reinke (DKFZ).

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, we set the scores to the worst possible rank for this case.

c) Justify why the described ranking scheme(s) was/were used.

To measure the performance of the contributions, we will evaluate the quality of the infilled regions. Since ground truth data is only available for the masked regions with healthy tissue, the evaluation will be restricted to these. We will use the following set of well-established metrics to quantify how realistic the synthesized image regions are compared to real ones: structural similarity index measure (SSIM), peak-signal-to-noise-ratio, and mean-square-error (MSE). For the final ranking of the MICCAI challenge, an equally weighted rank-sum is computed across all three metrics. To compute the rank within each metric, we rank the participants for each case and again compute a rank-sum. For these computations we use challengeR.

(image source: Reinke 2023)

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Similar to BraTS 2017-2022, uncertainties in rankings will be assessed using permutational analyses [13]. Therefore, we conduct bootstrapping and robustness analysis with challengeR.

b) Justify why the described statistical method(s) was/were used.

We want to investigate how much our rankings are driven by individual cases or in other words how robust they are.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

In addition to the challenge metrics we will also evaluate perceptual metrics such as LPIPS and provide qualitative expert evaluations

TASK 10: BraTS-Pathology: Assessing the Heterogeneous Histologic Landscape of Glioma

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Glioblastoma is the most common primary parenchymal tumor of the brain. Clinically, glioblastoma has a grim prognosis with unusually short duration antecedent symptoms and median survival of 12-18 months. This malignant tumor is widely infiltrative in the cerebral hemispheres and well-characterized by heterogenous molecular profiles as well as histopathologic features. A major obstacle in treating these tumors is this molecular and micro-environmental landscape heterogeneity. Correctly diagnosing these tumors and assessing their heterogeneity is crucial for choosing the precise treatment and potentially enhancing patient survival rates. In the gold-standard histopathology-based approach to tumor diagnosis, detecting various morpho-pathological features of distinct histology throughout digitized tissue sections is crucial. Such "features" include the presence of cellular tumor, geographic necrosis, pseudopalisading necrosis, areas abundant in microvascular proliferation, infiltration into the cortex, wide extension in subcortical white matter, leptomeningeal infiltration, regions dense with macrophages, and the presence of perivascular or scattered lymphocytes. With these features in mind and building upon the main aim of the BraTS Cluster of Challenges, the goal of the BraTS-Path challenge is to develop deep-learning models capable of identifying tumor sub-regions of distinct histologic profile. These models aim to assist in the diagnosis and grading of conditions in a consistent manner.

In the BraTS-Path challenge dataset, we focus on glioblastoma (GBM) digitized tissue sections with representative features. A team of neuropathologists annotated the slides, by identifying these distinct regions. Subsequently, these regions were segmented into patches classified based on the presence of specific histology. This approach established a classification task aimed at accurately identifying patches with specific features.

The challenge participants can obtain the labeled training data at any point from the Synapse platform. These data will be used to develop, containerize, and evaluate their algorithms in unseen validation data until July 2024, when the organizers will stop accepting new submissions and evaluate the submitted algorithms in the hidden testing data. Ground truth reference annotations for all datasets are created and approved by expert neuropathologists for every subject included in the training, validation, and testing datasets to evaluate the performance of the participating algorithms quantitatively.

Keywords

List the primary keywords that characterize the task.

Classification, Pathology, Digital Pathology, Brain Tumor, Cancer, Challenge, Glioma, Glioblastoma, health disparities, MICCAI, NCI, DREAM, diffuse glioma

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Spyridon Bakas [Lead Organizer - Contact Person]
Indiana University

Jake Albrecht
Sage Bionetworks

Verena Chung
Sage Bionetworks

Lee A D Cooper
Northwestern University

Shahriar Faghani
Mayo Clinic

Keyvan Farahani
NIH

Mana Moassefi
Mayo Clinic

Sarthak Pati
Indiana University

Siddhesh Pravin Thakur
Indiana University

Clinical Organizers:
Robert Bell,
Indiana University

Jason Huse,
MD Anderson Cancer Center

b) Provide information on the primary contact person.

Spyridon Bakas, PhD
[Lead Organizer of the "BraTS-Pathology: Assessing the Heterogeneous Histologic Landscape of Glioma"]
Indiana University
Email: spbakas@iu.edu

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed conference submission deadline, and continuous evaluation after the conference deadline

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

Synapse.org

Following our successful collaboration with the Synapse platform (SAGE Bionetworks) since the RSNA-ASNR-MICCAI BraTS 2021 challenge [1], we have coordinated with them and following the support from NCI (represented by Dr Keyvan Farahani in the organizing committee - Chair of the NCI AI Challenges Working Group) Synapse will be used as the platform to drive the evaluation of this cluster of challenges.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in a number of ways. Dr Keyvan Farahani, a long-time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC). All aforementioned NCI platforms are implemented on the Google Cloud Platform.

This collaboration with Synapse, enabled by NCI/NIH support through ITCR grant (Jamed Eddy, PI) and other NCI resources represents a major advancement in the challenge design and leveraging of public resources.

c) Provide the URL for the challenge website (if any).

https://www.synapse.org/brats2024 - (Website will be publicly visible after the challenge approval)

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

Participants are allowed to use additional data from publicly available datasets and their own institutions, for further complementing the data, but if they do so, they MUST also discuss the potential difference in their results after using only the BraTS 2024 data, since our intention is to solve the particular classification problem, but importantly to provide a fair comparison among the participating methods.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but organizers and their immediate groups cannot be 1) eligible for awards, 2) announced as the winners of the challenge, or 3) included in the announced formal rankings. They will however be evaluated and if they are within the top-ranked ones they will be honorarily mentioned to contribute back to the community. Since organizing institutions are large, other employees from other labs/departments may participate and should be eligible for the awards and to be listed in the official leaderboard.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

The lead organizer of the challenge is in communication with 1) Intel and 2) Neosoma Inc, to sponsor monetary awards for the top 3 teams.
Formal confirmation can only be provided after the acceptance of the challenge.

Note that Intel has been offering monetary awards during each of BraTS 2018-2023, and Neosoma for BraTS 2021-2023.

NIH will also provide Certificates of Merit to the top 3 performing teams.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top 3 performing methods will be announced publicly at the conference and the participants will be invited to present their method during an oral presentation.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams’ members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

The configuration of coordinating the BraTS challenge proceedings with the BrainLes workshop proceedings provides the BraTS participants with the option to publish their methods in the associated LNCS post-conference proceedings.

Furthermore, we intend to coordinate a journal manuscript focusing on publishing and summarizing the results of the challenge.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

The participants are required to send the output of their methods to the evaluation platform for the scoring to occur during the training and the validation phases. At the end of the validation phase the participants are asked to identify the method they would like to evaluate in the final testing/ranking phase.

The organizers will then confirm receiving the containerized method and will evaluate it in unseen and permanently hidden (from future challenges too) testing data. The participants will be provided detailed guidelines on the form of the container as we have done in previous years. This will enable confirmation of reproducibility.

Submission of a containerized method will take place after the participants agree to release this container publicly available with either a CC-BY, or CC-BY-NC license as part of the challenge repository and meta-analysis manuscript.

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

We intend to release the training set in April and the validation set in June, allowing participants to tune their methods in the unseen validation data. The validation data ground truth will not be provided to the participants, but multiple submissions to the online evaluation platform will be allowed for the validation phase. However, only 2 submissions will be allowed in the final testing/ranking data/phase to avoid potential tuning of the submitted approach to the testing data.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration dates: From now until submission deadline of short papers reporting method and preliminary results (see below).

1 March 2024: Registration opens
Participants will be able to register for the challenge in synapse.org, from the date of its potential acceptance (March 1, 2024) until the short paper submission deadline (July 31, 2024).

1 April 2024: Training data release
Availability of training data (with ground truth labels).

1 June 2024: Validation data release
Availability of validation data (without ground truth labels).

31 July 2024: Short paper submission deadline
Reporting method & results on training and validation data.
The only difference with the final paper submission should be the inclusion of the testing results in the camera-ready submission

15 August 2024: Containerized algorithm submission deadline.
Evaluation on testing data by the organizers, only for participants with submitted short papers. Ranking of all participating methods, following statistical significance assessment based on multiple permutation testing.

20 August 2024: Invitation to participate
Inviting all participants with valid submissions (paper + container) to present at the conference (type of presentation will be determined within the next 2 weeks)

4 September 2024: Contacting top-performing methods for preparing slides for oral presentation.

6-10 October 2024: Challenge at MICCAI
Announcement of final top 3 ranked teams

20 October 2024: Camera-ready submission deadline
Incl. results on testing data, for inclusion in the associated LNCS proceedings.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

Not applicable, as these cases are already publicly available from The Cancer Imaging Archive (TCIA), as part of the TCGA-GBM and TCGA-LGG data collections. However, please note that expert clinicians at our end worked on their reclassification according to the latest WHO classification criteria.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY, as the exact training/validation data will follow the existing TCGA-GBM and TCGA-LGG license.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The preprocessing tools, evaluation metrics, and the ranking code used during the whole challenge's lifecycle will be made available through the Generally Nuanced Deep Learning Framework (GaNDLF - https://github.com/mlcommons/GaNDLF).

2 Pati S, Thakur SP, Hamamci E, Baid U, Baheti B, Bhalerao M, et al. GaNDLF: the generally nuanced deep learning framework for scalable end-to-end clinical workflows. Communications Engineering. 2023;2(1):1-17.

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

The participants are required to submit their containerized algorithm, during or after the validation phase. Specific instructions for the containerization will be provided after the challenge approval. These instructions will be very similar to what we were requesting participants to provide during the BraTS 2021-2023 challenges.
All participants of the BraTS-Path challenge they will be required to accept an agreement through the synapse.org website that participation in the testing phase of the challenge, will automatically mean that we can make their containerized method publicly available through our challenge webpage.

The National Cancer Institute takes special interest in the BraTS 2024 challenge and is considering providing infrastructural support in several ways. Dr Keyvan Farahani, a long time co-organizer of BraTS challenges and a project scientist on a collaborative NCI Informatics Technology for Cancer Research (ITCR) grant, is the recipient of an NIH Office of Data Science and Strategy (ODSS)-STRIDES award for a sustainable medical imaging challenge cloud infrastructure, to further implement open (continuous) challenges by supporting cloud compute and other infrastructures for (a) benchmarking of tools and automated submission of containerized tools for evaluation, (b) hosting of top-ranking tools through NCI FireCloud Resource and public tool repository such as Dockstore or ModelHub, and (c) hosting resulting image annotations as derived data in the Imaging Data Commons (IDC) on the Google Cloud Platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

Monetary awards are expected by Intel and Neosoma Inc

Spyridon Bakas, Shahriar Faghani, Siddhesh Thakur, SAGE Bionetworks (synapse.org), and the clinical evaluators will be the only ones who will have access to the validation, and test case ground truth labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Research, CAD, Decision support, Treatment planning, Diagnosis, Assistance, Surgery, Intervention planning, Education, Training.

Task category(ies)

State the task category(ies)

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration

- Retrieval
- Segmentation
- Tracking

Classification

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

Patients, diagnosed with de novo diffuse gliomas of the brain.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Retrospective multi-institutional cohort of patients, diagnosed with de novo diffuse gliomas of the brain, with clinically digitized tissue sections using the paradigm of Formalin-Fixed Paraffin-Embedded (FFPE) and stained with Hematoxylin and Eosin.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Histopathology images. Specifically, H&E-stained; FFPE digitized tissue sections.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

directly to the image data (i.e., tumor sub-region class)

b) ... to the patient in general (e.g. sex, medical history).

N/A

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Brain tumor tissue showing in H&E-stained; FFPE digitized tissue sections.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

Tumor in the brain.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Accuracy, AUC, F1 Score, Matthews Correlation Coefficient (MCC), Sensitivity, Specificity.

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

The H&E-stained; FFPE digitized tissue sections used in the BraTS-Path challenge, describe histology images acquired during standard clinical practice across the 11 International sites mentioned in (c) below. The exact staining process details and the digital scanners (with their technical specifications) used for acquiring this TCIA cohort are not publicly available neither through TCIA, nor through the Genomic Data Commons (GDC) Data Portal of the NIH/NCI.

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

The exact staining process details and the digital scanners (with their technical specifications) used across these 11 sites to acquire this TCIA cohort are not publicly available neither through TCIA, nor through the Genomic Data Commons (GDC) Data Portal of the NIH/NCI.

We appreciate that the acquisition protocols and equipment are different across (and within each) contributing institution, as these represent real routine clinical practice. We are in coordination with TCIA to identify as much of these specific details for each image of each patient and then publish this as supplementary material together with the challenge meta-analysis manuscript.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

The provided data describe H&E-stained; FFPE digitized tissue sections, acquired with different clinical protocols and various scanners from:
1) Henry Ford Hospital (MI, USA),
2) University of California (CA, USA),
3) MD Anderson Cancer Center (TX, USA),
4) Emory University (GA, USA),

5) Mayo Clinic (MN, USA),
6) Thomas Jefferson University (PA, USA),
7) Duke University School of Medicine (NC, USA),
8) Saint Joseph Hospital and Medical Center (AZ, USA),
9) Case Western Reserve University (OH, USA),
10) University of North Carolina (NC, USA),
11) Fondazione IRCCS Instituto Neuroligico C. Besta, (Italy),

Note that data from these institutions are provided through The Cancer Imaging Archive (TCIA - http://www.cancerimagingarchive.net/), supported by the Cancer Imaging Program (CIP) of the National Cancer Institute (NCI) of the National Institutes of Health (NIH).

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

Clinical experts (neuropathologists and technicians) involved in tissue staining for suspected and diagnosis of brain tumors during standard clinical practice.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

Based on the given definition (in this section - (a)) that a case encompasses data processed to produce one result that is compared to the corresponding reference result, a case in this challenge represents an individual patch extracted from an H&E-stained; FFPE digitized tissue section of a single patient tumor at a specific timepoint. We ensured that the patches were of a similar size, with each representing either a specific class present in that patch or none, in which case it was classified as 'background'.

These tissue sections exhibit a variety of features indicative of the diagnosis of a glioblastoma and have been annotated by expert neuropathologists. These annotated regions are divided into same size patches, each of them corresponding to a distinct morpho-histologic feature (or class) that the participants are expected to predict. Since this task has not been conducted before, we consider individual patches as individual cases in this challenge, with the intention of conducting a deeper analysis and offer a deeper understanding of these distinct features/classes in a more fine-grained resolution. Specifically, we would like to assess the intrinsic similarity of these classes and hence inherent difficulty of detecting individual classes, as well as which are the most confused with each other features/classes.

Throughout both the training, validation, and testing phases, these patches are classified according to their respective features/classes. The inclusion criteria for each tissue section were determined by the presence of histologic features characteristic of glioblastoma. Please note that all tissue section included for each case of the provided dataset, represent the tissue sections with the best quality available for this particular case.

b) State the total number of training, validation and test cases.

Training data: 195,000 cases (from 130 digitized tissue sections)
Validation data: 25,000 cases (from 18 digitized tissue sections)
Testing data: 60,000 cases (from 40 digitized tissue sections)
Note: None of these cases have been used in previous challenges.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Based on availability. All available data were split into training, validation, and testing following a 70%-10%-20% proportion, in line with conventional proportions used in machine learning studies.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

Class distribution in classification tasks chosen according to real-world distribution. Choice was made to ensure methods that can consider the challenging real-world problem and extend to a more difficult task in the next year.

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Reference approved or edited until consensus from 2 experienced neuropathologists, following manual annotations from 10 clinical neuropathologists (volunteers from the RANO cooperative group)

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The annotation of these data followed a pre-defined clinically-approved annotation protocol (defined by expert neuropathologists), which was provided to all clinical annotators, describing in detail instructions on what the segmentations of each histologic feature should describe (see below for the summary of the specific instructions). The annotators were given the flexibility to use their tool of preference for making the annotations, or the provided infrastructure based on the Digital Slide Archive available through a web portal by Indiana University, and follow a complete manual annotation approach.

Summary of specific histologic areas of interest:
i) presence of cellular tumor
ii) pseudopalisading necrosis
iii) areas abundant in microvascular proliferation

iv) geographic necrosis
v) infiltration into the cortex
vi) penetration into white matter
vii) leptomeningeal infiltration
viii) regions dense with macrophages
ix) presence of lymphocytes

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Each case was assigned to a pair of annotator-approver. Annotators spanned across various experience levels and clinical/academic ranks, while the approvers were the 2 experienced board-certified neuropathologists (with >10 years of experience), listed in the Organizers' section as Clinical Organizers. The annotators were given the flexibility to use their tool of preference for making the annotations, or the provided infrastructure based on the Digital Slide Archive available through a web portal by Indiana University, and follow a complete manual annotation approach. Once the annotators were satisfied with the produced annotations, they were passing these to the corresponding approver. The approver is then responsible for signing off these annotations. Specifically, the approver would review the tumor annotations, in tandem with the corresponding tissue section, and the annotations of not satisfactory quality were removed from the provided annotation. If the patches from the remaining annotations were less than approximately 1,500 patches then the tissue sections would be sent back to the annotators for further annotations. This iterative approach was followed for all cases, until their respective annotations reached satisfactory quality (according to the approver) for being publicly available and noted as final ground truth segmentation labels for these cases.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

The TCGA-GBM and TCGA-LGG data sets, which are publicly accessible via the TCIA, have been chosen for this challenge. Initially, we have reclassified these collections in line with the 2021 WHO classification of CNS tumors. This reclassification was specifically done to pinpoint all cases of GBM IDH-wildtype, which are categorized under CNS WHO grade 4. The TCGA-LGG collection, initially classified as low-grade astrocytomas, is redefined under the 2021 WHO CNS criteria as GBM due to specific molecular characteristics indicative of distinct tumor evolution. Consequently, these astrocytomas, now classified as molecular GBM, are included in this challenge to develop algorithms applicable to all clinical GBM as per WHO guidelines. Conversely, certain cases in the TCGA-GBM collection have been excluded because their molecular profiles do not align with the current WHO definition of GBM.
For this study, a single H&E-stained; tissue section from each case in the reclassified TCGA-GBM and TCGA-LGG collections is used. Focusing solely on Formalin-Fixed Paraffin-Embedded (FFPE) slides, we avoid hydration artifacts common in frozen sections.

Post annotation of histologically distinct regions by clinical experts, each region is segmented into 256x256 patches. Rigorous patch-level image curation is essential to distinguish between tissue-occupied areas and artifacts. This process involves excluding patches with excessive background or artifacts like glass reflection, pen markings, or tissue tearing. The preprocessing includes three steps:
1. Removal of patches with significant white (intensity >230) or black (intensity <25) background, based on Red-Green-Blue (RGB) values, if such colors exceed 60% of the patch.
2. Conversion to Hue-Saturation-Value (HSV) space to eliminate patches with substantial saturation or value anomalies, discarding those where the percentage of such pixels exceeds 95%.
3. Stain deconvolution into Hematoxylin-Eosin-DAB (HED) space, discarding patches where over 80% of Eosin channel pixels have an intensity below 50.
These thresholds have been empirically determined to ensure that only artifacts are removed, preserving tissue-occupied areas in the selected patches

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

N/A, as only areas of high confidence from at least 2 neuro-pathologists are used for the patch creation.

Perhaps at a later version of the challenge we could propose to study and evaluate the effect of any potential annotations error as an uncertainty task, similar to the one we did in BraTS 2019-2020 (i.e., to quantify the uncertainty in the tumor segmentations) [1], but for now this is outside the scope of the BraTS-Path 2024 challenge.

1 R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning Biomedical Imaging, 1, 26, 2022

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

Accuracy,
AUC,
F1 Score,

Matthews Correlation Coefficient (MCC),
Sensitivity,
Specificity.

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

In terms of evaluation metrics, we use:
I) This fundamental metric will provide the proportion of true results (both true positives and true negatives) among the total number of cases examined. It proves the effectiveness of the classification model.
II) MCC provides a balanced measure even when the classes are of very different sizes. It is a correlation coefficient between the observed and predicted classifications, offering a more informative and nuanced assessment than simple accuracy.
III) As a measure that balances precision and recall (sensitivity), the F1 score is crucial for scenarios where the cost of false positives and false negatives is high. It is particularly useful when dealing with imbalance between classes.
IV) AUCROC curve, quantifies the overall ability of the model to discriminate between the positive and negative classes across different thresholds. A higher AUC indicates better model performance
V) Sensitivity and Specificity to determine whether an algorithm has the tendency to over- or underclassify different classes.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

Following discussions with the biostatistician involved in the design of this challenge (Dr Kun Huang, Chair of Dept of Biostatistics & Health Data Science), and also while considering transparency and fairness to the participants.

b) Describe the method(s) used to manage submissions with missing results on test cases.

If an algorithm fails to produce a result metric for a specific test case, this metric will be set to its worst possible value (e.g., 0 for accuracy and F1 score)

c) Justify why the described ranking scheme(s) was/were used.

For ranking of multidimensional outcomes (or metrics), for each team, we will compute the summation of their ranks across the average of the metrics described above as a univariate overall summary measure. This measure will decide the overall ranking for each specific team. To visualize the results in an intuitive fashion, we propose to visualize the outcome via an augmented version of radar plot [3].

3 Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,

- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

Building upon the approach in BraTS 2017-2023, the performance of the classification task will be evaluated based on the relative performance (as an aggregate metric of the ones described above) of each team's model in classifying different tumor tissue types. The assessment will involve a detailed analysis of each model's classification performance for various tumor tissues. The results will be synthesized by averaging the performance metrics across different classes, and the statistical significance of the classification performance will be determined. This evaluation will be conducted through permutation tests, where the observed classification accuracies are compared against a distribution generated by permuting the labels of the test data.

b) Justify why the described statistical method(s) was/were used.

This permutation testing would reflect differences in performance that exceeded those that might be expected by chance.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

N/A

ADDITIONAL POINTS

References

Please include any reference important for the challenge design, for example publications on the data, the annotation process or the chosen metrics as well as DOIs referring to data or code.

BraTS GLIOMA
1 C .Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of Medical Imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018
2 S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978 3 030 46643 5_38
3 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361 6560/ac9449
4 C.Davatzikos, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018, 2018. https://doi.org/10.1117/1.jmi.5.1.011018
5 S.Pati, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham, 2019. https://doi.org/10.1007/978 3 030 46643 5_38

6 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361 6560/ac9449
7 U. Baid, et al., The RSNA ASNR MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
8 S.Bakas, 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
9 Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
10 T. Rohlfing, et al. The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798 819, 2010.
11 R.Cox, J.Ashburner, H.Breman, K.Fissell, C.Haselgrove, C.Holmes, J.Lancaster, D.Rex, S.Smith, J.Woodward, A (Sort of) new image data format standard: NIfTI 1: WE 150, Neuroimage, 22, 2004.
12 S.Thakur, J.Doshi, S.Pati, S.Rathore, C.Sako, M.Bilello, S.M.Ha, G.Shukla, A.Flanders, A.Kotrotsou, M.Milchenko, S.Liem, G.S.Alexander, J.Lombardo, J.D.Palmer, P.LaMontagne, A.Nazeri, S.Talbar, U.Kulkarni, D.Marcus, R.Colen, C.Davatzikos, G.Erus, S.Bakas, 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. DOI: 10.1016/j.neuroimage.2020.117081
13 R.Mehta, et al, QU BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
14 Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large scale Multivariate Network Meta analysis. medRxiv. 2020 Jan 1

BraTS Meningioma

1 U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
2 S.Bakas, 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
3 B. H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
4 S. Bakas, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
5 C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018
6 S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978-3-030-46643-5_38
7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449
8 R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor

Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
9 Li X, Morgan PS, Ashburner J, Smith J, Rorden C (2016) The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J Neurosci Methods. 264:47-56. doi: 10.1016/j.jneumeth.2016.03.001. PMID: 26945974
10 Schwarz CG, Kremers WK, Wiste HJ, Gunter JL, Vemuri P, Spychalla AJ, Kantarci K, Schultz AP, Sperling RA, Knopman DS, Petersen RC, Jack CR, 2021. Changing the face of neuroimaging research: comparing a new MRI de-facing technique with popular alternatives. Neuroimage 231. doi: 10.1016/j.neuroimage.2021.117845.
11 Duan R, et al., PALM: Patient-centered Treatment Ranking via Large-scale Multivariate Network Metaanalysis. medRxiv. 2020 Jan 1
12 K. Clark, et al., The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, 26(6):1045-1057 (2013)
13 L. Maier-Hein, et al., BIAS: Transparent reporting of biomedical image analysis challenges, arXiv preprint arXiv:1910.04071 (2019)

BraTS Metastasis

1 U. Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
2 S.Bakas, 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
3 B. H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
4 S. Bakas, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
5 C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018
6 S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978-3-030-46643-5_38
7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449
8 R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
9 T. Rohlfing, et al., The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798-819, 2010.
10 R.Cox, et al., A (Sort of) new image data format standard: NIfTI-1: WE 150, Neuroimage, 22, 2004.
11 S.Thakur, 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. DOI: 10.1016/j.neuroimage.2020.117081
12 Duan R, et al., PALM: Patient-centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1

13 K. Clark, et al., The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, 26(6):1045-1057 (2013)
14 L. Maier-Hein, et al., BIAS: Transparent reporting of biomedical image analysis challenges, arXiv preprint arXiv:1910.04071 (2019)

BraTS Africa

1 M. Poon, et al., Longer-term (>= 2 years) survival in patients with glioblastoma in population-based studies pre- and post-2005: a systematic review and meta-analysis. Sci Rep. 2020 Jul 15;10(1):11622. https://doi.org/10.1038/s41598-020-68011-4
2 WHO GBD 2016 Brain and Other CNS Cancer Collaborators. Global, regional, and national burden of brain and other CNS cancer, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016, Lancet Neurol. 2019 Apr;18(4):376-393. https://doi.org/10.1016/S1474-4422(18)30468-X
3 M. Adewole, et al., The Brain Tumor Segmentation (BraTS) Challenge 2023: Glioma Segmentation in Sub-Saharan Africa Patient Population (BraTS-Africa) (arXiv:2305.19369), https://doi.org/10.48550/arXiv.2305.19369
4 U. Anazodo, et al., AI for Population and Global Health in Radiology. Radiology: Artificial Intelligence, 2022. https://doi.org/10.1148/ryai.220107
5 C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018
6 S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978-3-030-46643-5_38
7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open-source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361-6560/ac9449
8 R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
9 T. Rohlfing, et al., The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798-819, 2010.
10 R.Cox, et al., A (Sort of) new image data format standard: NIfTI-1: WE 150, Neuroimage, 22, 2004.
11 S.Thakur, 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. DOI: 10.1016/j.neuroimage.2020.117081
12 Duan R, et al., PALM: Patient-centered Treatment Ranking via Large-scale Multivariate Network Meta-analysis. medRxiv. 2020 Jan 1
13 K. Clark, et al., The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, 26(6):1045-1057 (2013)
14 L. Maier-Hein, et al., BIAS: Transparent reporting of biomedical image analysis challenges, arXiv preprint arXiv:1910.04071 (2019)

BraTS PEDS

1 U. Baid, et al., The RSNA ASNR MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314

2 S.Bakas, 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

3 B. H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993 2024 (2015) DOI: 10.1109/TMI.2014.2377694

4 Kazerooni AF, et al,. The brain tumor segmentation (BRATS) challenge 2023: Focus on pediatrics (CBTN CONNECT DIPGR ASNR MICCAI BraTS PEDs). ArXiv. 2023 May 26

5 Familiar AM, Kazerooni AF, et al., A multi institutional pediatric dataset of clinical radiology MRIs by the Childrens Brain Tumor Network. arXiv preprint arXiv:2310.01413. 2023 Oct 2.

6 C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018

7 S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978 3 030 46643 5_38

8 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361 6560/ac9449

9 Fathi Kazerooni A, et al. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi institutional study. Neuro Oncology Advances. 2023 Jan 1;5(1):vdad027.

10 Vossough A, et al, Training and Comparison of nnU Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors. Under Review (American Journal of Neuroradiology)

11 Liu X, et al. From adult to pediatric: deep learning based automatic segmentation of rare pediatric brain tumors. InMedical Imaging 2023: Computer Aided Diagnosis 2023 Apr 7 (Vol. 12465, pp. 15 19). SPIE.

12 R.Mehta, et al, QU BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022

13 Duan R, Tong J, Lin L, Levine LD, Sammel MD, Stoddard J, Li T, Schmid CH, Chu H, Chen Y. PALM: Patient centered Treatment Ranking via Large scale Multivariate Network Meta analysis. medRxiv. 2020 Jan 1

BraTS GoAT

1 U. Baid, et al., The RSNA ASNR MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314

2 S.Bakas, 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

3 B. H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993 2024 (2015) DOI: 10.1109/TMI.2014.2377694

4 S. Bakas, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

5 C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI:

10.1117/1.jmi.5.1.011018
6 S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978 3 030 46643 5_38
7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361 6560/ac9449
8 R.Mehta, et al, QU BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
9 T. Rohlfing, et al., The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798 819, 2010.
10 R.Cox, et al., A (Sort of) new image data format standard: NIfTI 1: WE 150, Neuroimage, 22, 2004.
11 S.Thakur, 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. DOI: 10.1016/j.neuroimage.2020.117081

BraTS Augment

1 U. Baid, et al., The RSNA ASNR MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
2 S.Bakas, 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
3 B. H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993 2024 (2015) DOI: 10.1109/TMI.2014.2377694
4 S. Bakas, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
5 C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018
6 S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978 3 030 46643 5_38
7 S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361 6560/ac9449
8 R.Mehta, et al, QU BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
9 T. Rohlfing, et al., The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798 819, 2010.
10 R.Cox, et al., A (Sort of) new image data format standard: NIfTI 1: WE 150, Neuroimage, 22, 2004.
11 S.Thakur, 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. DOI: 10.1016/j.neuroimage.2020.117081
12 Duan R, et al., PALM: Patient centered Treatment Ranking via Large scale Multivariate Network Metaanalysis. medRxiv. 2020 Jan 1

13 K. Clark, et al., The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, 26(6):1045 1057 (2013)
14 L. Maier Hein, et al., BIAS: Transparent reporting of biomedical image analysis challenges, arXiv preprint arXiv:1910.04071 (2019)

BraTS Synthesis

1. WHO Causes of child mortality, WHO, 2020. http://www.who.int/gho/child_health/mortality/causes/en/ (accessed Jun. 07, 2020).
2. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61-78
3. Wang G, Li W, Ourselin S, Vercauteren T. Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks With Uncertainty Estimation. Front Comput Neurosci 2019;13:56
4. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans Med Imaging 2016;35(5):1240-1251
5. Conte GM, Weston AD, Vogelsang DC, Philbrick KA, Cai JC, Barbera M, Sanvito F, Lachance DH, Jenkins RB, Tobin WO, Eckel Passow JE. Generative adversarial networks to synthesize missing T1 and FLAIR MRI sequences for use in a multisequence brain tumor segmentation model. Radiology. 2021 May;299(2):313 23.
6. Iglesias JE, Konukoglu E, Zikic D, Glocker B, Leemput KV, Fischl B. Is synthesizing MRI contrast useful for inter modality analysis?. In International Conference on Medical Image Computing and Computer Assisted Intervention 2013 Sep 22 (pp. 631 638). Springer, Berlin, Heidelberg.
7. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. Journal of medical systems. 2018 Nov;42(11):1 3.
8. Li H, Paetzold JC, Sekuboyina A, Kofler F, Zhang J, Kirschke JS, Wiestler B, Menze B. DiamondGAN: unified multi modal generative adversarial networks for MRI sequences synthesis. In International Conference on Medical Image Computing and Computer Assisted Intervention 2019 Oct 13 (pp. 795 803). Springer, Cham.
9. Thomas MF, Kofler F, Grundl L, Finck T, Li H, Zimmer C, Menze B, Wiestler B. Improving Automated Glioma Segmentation in Routine Clinical Use Through Artificial Intelligence Based Replacement of Missing Sequences With Synthetic Magnetic Resonance Imaging Scans. Investigative Radiology. 2022 Mar 1;57(3):187 93.
10. Iglesias JE, Billot B, Balbastre Y, Tabari A, Conklin J, González RG, Alexander DC, Golland P, Edlow BL, Fischl B, Alzheimers Disease Neuroimaging Initiative. Joint super resolution and synthesis of 1 mm isotropic MP RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. Neuroimage. 2021 Aug 15;237:118206.
11. U. Baid, et al., The RSNA ASNR MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification, arXiv preprint arXiv:2107.02314
12. S.Bakas, 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
13. B. H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993 2024 (2015) DOI: 10.1109/TMI.2014.2377694
14. S. Bakas, et al., Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
15. C.Davatzikos, et al., Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics

and predictive modeling of clinical outcome. Journal of medical imaging, 5.1:011018 (2018) DOI: 10.1117/1.jmi.5.1.011018
16. S.Pati, et al., The cancer imaging phenomics toolkit (CaPTk): technical overview. International MICCAI Brainlesion Workshop. Springer, Cham (2019) DOI: 10.1007/978 3 030 46643 5_38
17. S.Pati, et al, The federated tumor segmentation (FeTS) tool: an open source solution to further solid tumor research, Phys. Med. Biol. 67(20), 204002, 2022. DOI: 10.1088/1361 6560/ac9449
18. R.Mehta, et al, QU BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
19. T. Rohlfing, et al., The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp. 31(5):798 819, 2010.
20. R.Cox, et al., A (Sort of) new image data format standard: NIfTI 1: WE 150, Neuroimage, 22, 2004.
21. S.Thakur, 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. DOI: 10.1016/j.neuroimage.2020.117081
22. Duan R, et al., PALM: Patient centered Treatment Ranking via Large scale Multivariate Network Meta analysis. medRxiv. 2020 Jan 1
23. K. Clark, et al., The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, 26(6):1045 1057 (2013)
24. L. Maier Hein, et al., BIAS: Transparent reporting of biomedical image analysis challenges, arXiv preprint arXiv:1910.04071 (2019)
25. Kofler F, Berger C, Waldmannstetter D, Lipkova J, Ezhov I, Tetteh G, Kirschke J, Zimmer C, Wiestler B, Menze BH. BraTS toolkit: translating BraTS brain tumor segmentation algorithms into clinical and scientific practice. Frontiers in neuroscience. 2020:125.

BraTS-Inpainting

Inpainting challenge paper preprint: https://arxiv.org/pdf/2305.08992.pdf
1 WHO - Causes of child mortality, WHO, 2020.
http://www.who.int/gho/child_health/mortality/causes/en/ (accessed Jun. 07, 2020).
2. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61-78

BraTS Pathology

1 R.Mehta, et al, QU-BraTS: MICCAI BraTS 2020 Challenge on Quantifying Uncertainty in Brain Tumor Segmentation-Analysis of Ranking Scores and Benchmarking Results, Journal of Machine Learning for Biomedical Imaging, 1, 26, 2022
2 Pati S, Thakur SP, Hamamci E, Baid U, Baheti B, Bhalerao M, et al. GaNDLF: the generally nuanced deep learning framework for scalable end-to-end clinical workflows. Communications Engineering. 2023;2(1):1-17.
3 Duan R, et al., PALM: Patient-centered Treatment Ranking via Large-scale Multivariate Network Metaanalysis. medRxiv. 2020 Jan 1

4 K. Clark, et al., The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, 26(6):1045-1057 (2013)
5 L. Maier-Hein, et al., BIAS: Transparent reporting of biomedical image analysis challenges, arXiv preprint arXiv:1910.04071 (2019)
6 Reinke A, Tizabi MD, Baumgartner M, Eisenmann M, Heckmann-Nötzel D, Kavur AE, et al. Understanding metric-related pitfalls in image analysis validation. ArXiv. Published online September 25, 2023. doi:10.3115/1072064.1072067
7 Maier-Hein L, Reinke A, Godau P, Tizabi MD, Buettner F, Christodoulou E, et al. Metrics reloaded: Recommendations for image analysis validation. arXiv [csCV]. Published online June 3, 2022. http://arxiv.org/abs/2206.01653

Further comments

Further comments from the organizers.

N/A