Low field pediatric brain magnetic resonance Image Segmentation and quality Assurance: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Low field pediatric brain magnetic resonance Image Segmentation and quality Assurance

Challenge acronym

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

LISA

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.

In early life, accurate description of structural changes in the developing human brain is crucial for understanding healthy development and identification of neurodevelopmental disorders. Existing adult brain deep learning (DL) based segmentation tools struggle with pediatric MRI due to poor gray/white matter differentiation and rapid growth. These challenges impair the ability of existing algorithms to accurately segment pediatric brain structures even with high field (1.5T or 3T) MRI systems. In low and middle-income countries, high field MRI systems are rare due to their high cost and maintenance required. To fill this gap, 0.064T Hyperfine SWOOP scanners are being tested in these settings by the UNITY Consortium, funded by the Bill and Melinda Gates Foundation. Despite lower image quality, low-field MRI offers portability, cost-effectiveness, and eliminates the need for sedation in children. To translate recent improvements in infant brain segmentation to underprivileged communities, we introduce the Low-field pediatric brain magnetic resonance Image Segmentation and quality Assurance (LISA) challenge. We propose the first two tasks of the LISA challenge, to be presented at MICCAI 2024, with additional segmentation tasks to be added as future open challenges. Task 1 involves evaluating objective, quantifiable quality assurance (QA) methods that rate the overall quality of low-field MRI to ensure the acquired MRI meets specific accuracy and consistency standards. Task 2 concerns the automatic segmentation of the bilateral hippocampi, which are pivotal subcortical structures linked to cognitive and memory functions and are often implicated in abnormal neurodevelopment. The overarching goal of this challenge is to develop optimal and publicly available DL tools to assess and segment ultra-low field T2-weighted magnetic resonance images of the brain in early childhood.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

pediatric, MRI, low field, quality assurance, segmentation, hippocampus, MICCAI, brain, neurodevelopment, Africa

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.

N/A

Duration

How long does the challenge take?

N/A

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.

N/A

Publication and future plans

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

N/A

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.

N/A

TASK 1: Quality assessment of low field neonatal MR images

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.

During the initial stages of life the human brain undergoes rapid tissue growth and development after birth. Accurately capturing and describing structural changes from magnetic resonance images MRI during this vital period is crucial for gaining new perspectives on healthy brain development and enabling the early identification of neurodevelopmental disorders.
While validated brain automatic quality analysis tools exist for adult brains, efforts to directly translate existing adult brain algorithms to pediatric MRI have generally failed. This is partly due to the poor gray/white matter differentiation of the developing brain and its rapid growth throughout the first years of life.
In low and middle income countries, high field MRI systems are rare due to the high cost and maintenance required. More accessible MRI systems like the 0.064 T Hyperfine scanner would help to assess gross anatomical abnormalities; structural delineation challenges are further compounded in these lower resolution imaging environments. Despite the loss in image quality and challenges with image analysis, there are great advantages of using low field MRI including portability at the point of care of patients, small clinical costs with usability in low resource settings and elimination of sedation for young children.
To address the quality challenges in lowfield MRI, we introduce the first task of the Lowfield pediatric brain magnetic resonance Image Segmentation and Quality Assurance (LISA) challenge. The first task focuses on the evaluation of objective quantifiable quality assurance QA methods. These methods will rate the overall quality of lowfield 0.064 T MRI and ensure that the acquired MRI meets specific accuracy and consistency standards in the provided data.

Keywords

List the primary keywords that characterize the task.

pediatric, MRI, low field, quality assurance, MICCAI, brain, neurodevelopment, Africa

ORGANIZATION

Organizers

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

Lead Organizers: Natasha Lepore (1,4), Marius George Linguraru (2,3), Sean Deoni(6,7). Other organizers: Jeffrey Tanedo(1), Rahimeh Rouhi(1), Shreyash Zanjal(1), Austin Tapp(2), Marvin Nelson(5). Data Contributors : Samson Lecurieux Lafayette(8), Victoria Nankabirwa(9)(10), Sean Deoni (6,7)
1 CIBORG Laboratory, Department of Radiology, Children's Hospital Los Angeles; 2 Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital; 3 Departments of Radiology and Pediatrics, George Washington University School of Medicine and Health Sciences; 4 Department of Pediatrics and Biomedical Engineering, University of Southern California; 6 Advanced Baby Imaging Lab, Rhode Island Hospital; 7

Departments of Pediatrics and Diagnostic Radiology, Warren Alpert Medical School at Brown University; 8 St. Thomas Hospital, King's College London; 9 Kawempe National Referral Hospital, Makerere University, Kampala, Uganda;
10 St. Francis Hospital - Nsambya, Kampala, Uganda

b) Provide information on the primary contact person.

Natasha Lepore - nlepore@chla.usc.edu
Children's Hospital Los Angeles
4650 Sunset Blvd, Los Angeles, CA, 90027, USA
Department of Radiology and Imaging
Phone: 323-361-2411

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

This challenge will be a one-time event with a fixed submission deadline.

Challenge venue and platform

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

MICCAI 2024

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

We plan to use the Synapse platform (SAGE Bionetworks; Synapse.org) for the challenge. This platform has been used for numerous other challenges, including the CBTN-CONNECT-DIPGr-ASNR-MICCAI Pediatric Brain Tumor Segmentation Challenge 2023, organized by Dr. Linguraru.

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

synapse.org Synapse Project SynID is syn53282349.

Participation policies

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

Only fully automatic methods will be allowed.

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.

Data used to train algorithms is not restricted to only the data provided by the challenge. Publically available online and/or synthetic datasets are allowed. Our aim is to solve the segmentation and quality assurance of low-field MRI with the best possible tools. Investigator -owned data may be used only if data is made available to the community before August 8th, 2024, which is when the testing phase begins.

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.

Members of the organizers' institutes may participate in the challenge but are not eligible for awards

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

First prize in each task will be awarded $1000 each.
Second prize in each task will be awarded $500 each.
Third prize in each task will be awarded $250 each.
Top three teams in each task will receive award certificates.

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 three performing methods for each task will be announced publicly and these participants will be invited to present their method. All participant rankings (training, validation and testing) will also be made publicly available on the challenge website leaderboard. All participant methods that are ranked in the testing phase will have their accompanying short LNCS style paper (see section f) made visible to organizers and challenge participants. Making the participants' paper public will support methodological reproducibility and the open-source spirit the organizers seek to embody in this challenge.

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).

Authors participating in the evaluation of their containerized model against the unseen test data are required to submit a short 8-page LNCS style paper through the LISA CMT by Thursday, August 8th at 11:59 PM Pacific Time. Papers will be reviewed for suitability for online distribution. If challenge organizers deem that the method is not reproducible via the short paper instructions, then the method will not be evaluated and thus not considered for ranking. In some cases, organizers may provide comments to improve the short paper within the deadline. Details on training time and resources required to reproduce a participant's submission will be required in the short paper. Challenge participants must abide by the guiding principles for responsible research use and data handling within the Synapse Commons Platform as described in the Synapse Governance documents and by the Challenge's Official Rules. Publication embargo: Challenge participants are permitted to publish individual challenge results restricted to their own submitted method and ranking. Overall Challenge results or analysis thereof is restricted until Challenge organizers and participants have jointly published (or pre-published) an overview paper on the results and best performing strategies. The challenge organizers intend to aggregate the

methods and summarize the results of LISA 2024 as a single journal manuscript with the goal of publication in journals of Medical Image Analysis, IEEE TMI, or a journal of similar impact factor. Participants will be contacted through Synapse-affiliated email addresses when this condition has been met. This information will also be posted within this Synapse project. Acknowledgement: Challenge participants and other data users are permitted to use, publish and present the Challenge results, after the embargo period, provided they acknowledge the LISA challenge organizing members as follows: "Data used in this publication were obtained as part of MICCAI 2024's Low field pediatric brain magnetic resonance Image Segmentation and quality Assurance (LISA) Challenge through Synapse ID syn53282349", Publications using challenge resources will be required to cite the overview paper of LISA 2024 and acknowledge the support of the Bill and Melinda Gates Foundation, which provided funding used for supporting data collection, curation, and processing.

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.

For training/validation: In task 1, participants are required to upload their output CSV file to the evaluation platform for scoring.

For testing: Participants are required to upload a containerized method, using MLCube for submission. The MLCube submission contains information about the participant selected method, including the model that will be used during inference. A detailed description of submission guidelines will be provided to participants on our challenge website. The guidelines will ensure reproducibility of submitted methods and standardize the evaluation process, allowing for solutions to be fairly ranked and selected for solving the challenge objectives. The docker containers should run on a computer with a single GPU with 8 GB RAM, 8 core CPU and 16 GB RAM.

Throughout the challenge: During all phases of the challenge, participants will have the opportunity to test the functionality of their containerized submissions to ensure compatibility with the online evaluation platform (Synapse).

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.

To permit participant's method fine tuning, we will release the validation set in July (see timetable below). Validation ground truth data will not be released, but scores based on predictions will be provided to participants. Participants can select their final test submissions based on these validation results. After the validation phase ends, participants are allowed a maximum of 2 final testing submissions for each task, or 4 total submissions. The best scoring method of each task (2 methods per team) will be selected for final ranking. Neither test images nor their ground truths will be made public.

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

May 1
Challenge launch
Registration opens
mid-May
Release of training data with ground truth labels
Late June
Release of validation data (without labels)
8 July
Validation phase opens
8 August
Validation phase ends (submission deadline of segmentation files)
Submission deadline of short papers in CMT, reporting method and preliminary results.
9 August
Testing phase opens
(available only to participants who submitted short papers)
16 August
Testing phase ends (submission deadline of MLCubes)
16 August - 13 Sept
Evaluation of MLCubes on testing data
(performed by challenge organizers)
15 September
Top performing methods contacted for preparing oral presentation slides
6 - 10 October
Presentation of Challenge, Announcement of the final top 3 ranked teams
November 2024
Camera-ready submission of extended papers for inclusion in the associated workshop 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 low-high field magnetic resonance image data that will be used for the training, validation, and testing of algorithms was acquired through the UNITY project funded by the Bill and Melinda Gates Foundation (BMGF) at

multiple international sites. Ethics approval was obtained at each site following the local standards and all data were fully de-identified before inclusion in the project. The protocol for releasing the data was approved by the institutional review board of the data contributing institution and data inclusion in the LISA challenge was approved by the BMGF. Training and validation data will be made available to challenge participants who agree to our terms and conditions through the Synapse website. Test data will be kept hidden from participants. LISA data will be available at the time of the challenge and beyond to increase its usability and impact.

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)

Challenge participants must abide by the guiding principles for responsible research use and data handling within the Synapse Commons Platform as described in the Synapse Governance documents for Controlled Data. Greater detail can be found in Synapse Commons Data Use Procedure. LISA 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.

Organizers' evaluation software, metrics and ranking code used for the challenge will be available on GitHub at: https://github.com/LISA2024Challenge/EvaluationMetrics and linked to our website.

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

Participants must submit their containerized algorithm during the testing phase. Specific instructions for submission will be provided after challenge approval and will be detailed on the challenge website (Synapse). Participant containers will be made available for public use following the final testing/ranking phase of the challenge. Source code will be made available only at code authors' request.

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.

This challenge was funded by the Bill and Melinda Gates Foundation Grant INV-047887 as part of the UNITY Consortium.

All challenge organizers and data contributors will have access to the validation and test case labels.

We are also exploring Hyperfine as a potential sponsor for this challenge.

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, education, training, diagnosis, intervention assistance, intervention follow-up, intervention planning, screening

Task category(ies)

State the task category(ies)

Examples:

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

- Tracking

Classification, recognition

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.

The target cohort will be all pediatric low field brain MRI.

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

The challenge cohort dataset will be healthy neonatal subjects approximately equal numbers of males and females scanned from the local population at :

(1) St. Thomas Hospital at King's College London, London, United Kingdom,
(2) Kawempe National Referral Hospital at Makerere University in Kampala, Uganda,
(3) Warren Alpert Medical School at Brown University, Providence, RI, USA,
and
the Advanced Baby Imaging Lab, Rhode Island Hospital, Providence, RI, USA

Imaging modality(ies)

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

Low field (0.064T) T2 MR 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).

For Task 1, challengers will be given ratings of the usability of the data in the form of number ratings across 6 artifact categories.

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

No other information about the subject will be available.

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.

Imaging data will consist of the pediatric brain and skull shown in low resolution magnetic resonance imaging (MRI) T2 data.

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.

The algorithm target will be low field T2 brain MR images

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)).

Rate overall quality of low field MR images across 6 possible artifacts - motion, contrast, noise, zipper, positioning, and banding with an error that reflects the intra-rater variability of an expert rater.

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).

For each data site, all low field images were obtained by scanning on a Hyperfine SWOOP, a 0.064T MR Scanner designed for FDA-approved portable brain imaging. The system is small enough to be put in patients' rooms and can be powered by standard electrical outlets. An attached Apple iPad handles postprocessing, from which, clinically useful images may be obtained.

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

Low field images from Hyperfine SWOOP at all three sites
Magnetic field strength 0.064T, sequence type: spin echo, TR 1.5s, TE 5ms, TI 400ms

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.

Data was acquired at (1) St. Thomas Hospital, King's College London, London, United Kingdom, (2) Kawempe National Referral Hospital, Makerere University, Kampala, Uganda, and. (3) Warren Alpert Medical School at Brown University, Providence, RI, USA, and the Advanced Baby Imaging Lab, Rhode Island Hospital, Providence, RI, USA.

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).

All images were collected by MRI techs with experience imaging patients at the institutions listed above. All quality assurance evaluations were performed by an expert medical image evaluator.

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.

One case encompasses a trio of orthogonally acquired low field T2 MR images and ratings in each of the 6 artifact classes.

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

For Task 1 (QA), we will provide 285 subjects of which 201, 42, and 42 subjects will be considered for training, validation, and testing, respectively.

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

The total number of cases has been determined based on the available data. We collected data from different sites to introduce diversity, thereby avoiding overfitting and achieving better generalizability for the proposed models in the tasks. To prevent overlap between scans from one subject in the training cohort and those in the validation and test cohorts (due to multiple scans from a single subject), we will consider subjects rather than individual scans during the data division for training and validation. We follow the standard division of data for training (70%), validation (15%), and test (15%) and apply it to each site to get the final division for training, validation and test cohorts.

The current distribution of subjects and images across different sites is as follows:
KCL, Uganda and Brown with 75, 95, and 115 subjects respectively.
KCL, Uganda and Brown with 47, 107, and 131 low field scans respectively.
Numbers will increase with growth in participating sites.

The number of subjects in training, validation and test cohorts is as follows.
KCL, Uganda, and Brown with 53, 67, and 81 training cases respectively.
KCL, Uganda and Brown with 11, 14, and 17 validation cases respectively.
KCL, Uganda and Brown with 11, 14, and 17 test cases respectively.

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 cases will be rated across 6 domains (motion, contrast, noise, zipper, FOV, and banding) in a three point scale (0,1,2).

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.

A single expert medical image evaluator was used to assess the quality of the data through manual image examination. A neuroradiologist reviewed and approved the assessment protocol.

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 quality assessment protocol was based on the architecture found [1] in order to rate motion, contrast, and noise. For the purpose of assessing the low field images, we also added categories for zipper, positioning and banding artifacts with the definitions below.

Zipper
0: no zipper
1: Zipper is visible around the brain and interferes with brain contrast
2: Zipper is visible and causes severe distortion with brain contrast

Complete FOV
0: Skull and cortex are within FOV
1: Any portion of the skull is missing from FOV
2: Both skull and any portion of the cortex have been cut from FOV

Banding
0: No Banding
1: Banding across the MR image where tissue contrast is still visible
2: Severe banding where tissue contrast is no longer visible

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.

The rater has 10 years of experience in assessing MR image quality.

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.

n/a

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.

All images were first converted to NIFTI from DICOM using dcm2nii. These low field images are defaced for full anonymization using SynthRad2023 processing pipeline's anonymize.py. No other information will be included that could include risk of reidentification. The repository is freely available at this link : https://github.com/SynthRAD2023/preprocessing

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.

We will provide intra-rater reliability scores to estimate the magnitude of human error in image annotation.

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)

We will employ five distinct metrics including accuracy, F1 score, F2 score, precision, and recall.

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

Metric choices were based on recommendations found in [4], ensuring magnetic resonance imaging samples meet minimum quality requirements and contain no significant image artifacts.

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.

The ranking system described will be used to obtain a fair and complete comparison across all metrics between participants. We desire a ranking scheme that is empirical, objective, and provides both participants and organizers with a clear understanding of the 'best' challenge solution.

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

Submissions that have missing results or cannot provide results because inference fails on a test case will be assigned the lowest possible score for the failed test case using metrics relevant to the task. For example, if a missing result or failed inference occurs in Task 1, the values for accuracy, F1 score, F2 score, precision, and recall would all be set to 0.

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

We will take the mean across their respective metrics in order obtain rankings. Standard deviations will also be calculated and given.

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.

Permutational analyses will be employed to evaluate uncertainties in rankings. Each team's QA performance will be assessed for all 6 artifact classes, considering each metric. Rankings for the measures will be combined by averaging, and statistical significance will be assessed exclusively for QA performance measures. This assessment will involve permuting the relative ranks for each QA measure, considering each artifact class for QA per subject in 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 2: Segmentation of bilateral hippocampi from low field neonatal MR images

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.

During the initial stages of life, the human brain undergoes rapid tissue growth and development after birth. Accurately capturing and describing structural changes from magnetic resonance images (MRI) during this vital period is crucial for gaining new perspectives on healthy brain development and enabling the early identification of neurodevelopmental disorders.
While validated brain segmentation tools exist for adult brains, efforts to directly translate existing adult brain algorithms to pediatric MRI have generally failed. This is partly due to the poor gray/white matter differentiation of the developing brain and its rapid growth throughout the first years of life. These challenges impair the ability of existing algorithms to accurately segment pediatric brain structures even with high field (1.5T or 3T) MRI systems.
In low and middle-income countries, high field MRI systems are rare, due to the cost and maintenance required. More accessible MRI systems like the 0.064T Hyperfine scanner would help to assess gross anatomical abnormalities; structural delineation challenges are further compounded in these lower resolution imaging environments. Despite the loss in image quality and challenges with image analysis, there are great advantages of using low field MRI, including portability at the point of care of patients, small clinical costs with usability in low resource settings, and elimination of sedation for young children.
To drive advancements in automatic segmentation methods for low-field MRI, the LISA challenge introduces the second task. Participants are tasked with developing and evaluating deep learning methods for the automatic segmentation of the bilateral hippocampi in ultra low field (0.064T) T2-weighted magnetic resonance images of the healthy brain in early childhood.
The second task is centered around the automatic segmentation of the bilateral hippocampi which are pivotal subcortical structures linked to cognitive and memory functions, and often implicated in abnormal neurodevelopment.
The LISA challenge provides challenge participants with the first-ever publicly available low-field MRI dataset, ensuring that the training, validation, and testing phases utilize the same dataset obtained from the Synapse platform. The overarching goal of the entire challenge is obtaining optimal deep learning tools for the assessment and segmentation of low-field MRI images in early childhood.

Keywords

List the primary keywords that characterize the task.

pediatric, MRI, low field, quality assurance, segmentation, hippocampus, MICCAI, brain, neurodevelopment, Africa

ORGANIZATION

Organizers

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

Lead Organizers: Natasha Lepore (1,4), Marius George Linguraru (2,3), Sean Deoni(6,7). Other organizers: Jeffrey Tanedo(1), Rahimeh Rouhi(1), Shreyash Zanjal(1), Austin Tapp(2), Marvin Nelson(5). Data Contributors : Samson Lecurieux Lafayette(8), Victoria Nankabirwa(9)(10), Sean Deoni (6,7)
1 CIBORG Laboratory, Department of Radiology, Children's Hospital Los Angeles; 2 Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital; 3 Departments of Radiology and Pediatrics, George Washington University School of Medicine and Health Sciences; 4 Department of Pediatrics and Biomedical Engineering, University of Southern California; 6 Advanced Baby Imaging Lab, Rhode Island Hospital; 7 Departments of Pediatrics and Diagnostic Radiology, Warren Alpert Medical School at Brown University; 8 St. Thomas Hospital, King's College London; 9 Kawempe National Referral Hospital, Makerere University, Kampala, Uganda;
10 St. Francis Hospital - Nsambya, Kampala, Uganda

b) Provide information on the primary contact person.

Natasha Lepore - nlepore@chla.usc.edu
Children's Hospital Los Angeles
4650 Sunset Blvd, Los Angeles, CA, 90027, USA
Department of Radiology and Imaging
Phone: 323-361-2411

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

This challenge will be a one-time event with a fixed submission deadline.

Future tasks added to the challenge will be in open call format.

Challenge venue and platform

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

MICCAI 2024

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

We plan to use the Synapse platform (SAGE Bionetworks; Synapse.org) for the challenge. This platform has been used for numerous other challenges, including the CBTN-CONNECT-DIPGr-ASNR-MICCAI Pediatric Brain Tumor Segmentation Challenge 2023, organized by Dr. Linguraru.

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

synapse.org Synapse Project SynID is syn53282349.

Participation policies

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

Only fully automatic methods will be allowed.

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.

Data used to train algorithms is not restricted to only the data provided by the challenge. Publically available online and/or synthetic datasets are allowed. Our aim is to solve the segmentation and quality assurance of low-field MRI with the best possible tools. Investigator -owned data may be used only if data is made available to the community before August 8th, 2024, which is when the testing phase begins.

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.

Members of the organizers' institutes may participate in the challenge but are not eligible for awards.

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

First prize in each task will be awarded $1000 each.
Second prize in each task will be awarded $500 each.
Third prize in each task will be awarded $250 each.
Top three teams in each task will receive award certificates.

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 three performing methods for each task will be announced publicly and these participants will be invited to present their method. All participant rankings (training, validation and testing) will also be made publicly available on the challenge website leaderboard. All participant methods that are ranked in the testing phase will have their accompanying short LNCS style paper (see section f) made visible to organizers and challenge participants. Making the participants' paper public will support methodological reproducibility and the open-source spirit the organizers seek to embody in this challenge.

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).

Authors participating in the evaluation of their containerized model against the unseen test data are required to submit a short 8-page LNCS style paper through the LISA CMT by Thursday, August 8th at 11:59 PM Pacific Time. Papers will be reviewed for suitability for online distribution. If challenge organizers deem that the method is not

reproducible via the short paper instructions, then the method will not be evaluated and thus not considered for ranking. In some cases, organizers may provide comments to improve the short paper within the deadline. Details on training time and resources required to reproduce a participant's submission will be required in the short paper. Challenge participants must abide by the guiding principles for responsible research use and data handling within the Synapse Commons Platform as described in the Synapse Governance documents and by the Challenge's Official Rules. Publication embargo: Challenge participants are permitted to publish individual challenge results restricted to their own submitted method and ranking. Overall Challenge results or analysis thereof is restricted until Challenge organizers and participants have jointly published (or pre-published) an overview paper on the results and best performing strategies. The challenge organizers intend to aggregate the methods and summarize the results of LISA 2024 as a single journal manuscript with the goal of publication in journals of Medical Image Analysis, IEEE TMI, or a journal of similar impact factor. Participants will be contacted through Synapse-affiliated email addresses when this condition has been met. This information will also be posted within this Synapse project. Acknowledgement: Challenge participants and other data users are permitted to use, publish and present the Challenge results, after the embargo period, provided they acknowledge the LISA challenge organizing members as follows: "Data used in this publication were obtained as part of MICCAI 2024's Low field pediatric brain magnetic resonance Image Segmentation and quality Assurance (LISA) Challenge through Synapse ID syn53282349", Publications using challenge resources will be required to cite the overview paper of LISA 2024 and acknowledge the support of the Bill and Melinda Gates Foundation, which provided funding used for supporting data collection, curation, and processing.

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.

For training/validation: In task 2, participants are required to upload their output segmentation prediction files to the evaluation platform for scoring.

For testing: Participants are required to upload a containerized method, using MLCube for submission. The MLCube submission contains information about the participant selected method, including the model that will be used during inference. A detailed description of submission guidelines will be provided to participants on our challenge website. The guidelines will ensure reproducibility of submitted methods and standardize the evaluation process, allowing for solutions to be fairly ranked and selected for solving the challenge objectives. The docker containers should run on a computer with a single GPU with 8 GB RAM, 8 core CPU and 16 GB RAM.

Throughout the challenge: During all phases of the challenge, participants will have the opportunity to test the functionality of their containerized submissions to ensure compatibility with the online evaluation platform (Synapse).

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.

To permit participant's method fine tuning, we will release the validation set in July (see timetable below). Validation ground truth data will not be released, but scores based on predictions will be provided to participants. Participants can select their final test submissions based on these validation results. After the validation phase ends, participants are allowed a maximum of 2 final testing submissions for each task, or 4 total submissions. The best scoring method of each task (2 methods per team) will be selected for final ranking. Neither test images nor their ground truths will be made public.

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

May 1
Challenge launch
Registration opens
mid-May
Release of training data with ground truth labels
Late June
Release of validation data (without labels)
8 July
Validation phase opens
8 August
Validation phase ends (submission deadline of segmentation files)
Submission deadline of short papers in CMT, reporting method and preliminary results.
9 August
Testing phase opens
(available only to participants who submitted short papers)
16 August
Testing phase ends (submission deadline of MLCubes)
16 August - 13 Sept
Evaluation of MLCubes on testing data
(performed by challenge organizers)
15 September
Top performing methods contacted for preparing oral presentation slides
6 - 10 October
Presentation of Challenge, Announcement of the final top 3 ranked teams
November 2024
Camera-ready submission of extended papers for inclusion in the associated workshop 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 low-high field magnetic resonance image data that will be used for the training, validation, and testing of algorithms was acquired through the UNITY project funded by the Bill and Melinda Gates Foundation (BMGF) at multiple international sites. Ethics approval was obtained at each site following the local standards and all data were fully de-identified before inclusion in the project. The protocol for releasing the data was approved by the institutional review board of the data contributing institution and data inclusion in the LISA challenge was approved by the BMGF. Training and validation data will be made available to challenge participants who agree to our terms and conditions through the Synapse website. Test data will be kept hidden from participants. LISA data will be available at the time of the challenge and beyond to increase its usability and impact.

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)

Challenge participants must abide by the guiding principles for responsible research use and data handling within the Synapse Commons Platform as described in the Synapse Governance documents for Controlled Data. Greater detail can be found in Synapse Commons Data Use Procedure. LISA 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.

Organizers' evaluation software, metrics and ranking code used for the challenge will be available on GitHub at: https://github.com/LISA2024Challenge/EvaluationMetrics and linked to our website.

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

Participants must submit their containerized algorithm during the testing phase. Specific instructions for submission will be provided after challenge approval and will be detailed on the challenge website (Synapse). Participant containers will be made available for public use following the final testing/ranking phase of the challenge. Source code will be made available only at code authors' request.

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.

This challenge was funded by the Bill and Melinda Gates Foundation Grant INV-047887 as part of the UNITY Consortium.

All challenge organizers and data contributors will have access to the validation and test case labels.

We are also exploring Hyperfine as a potential sponsor for this challenge.

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, education, training, diagnosis, intervention assistance, intervention follow-up, intervention planning, screening

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.

The target cohort will be all pediatric low field brain MRI.

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

The challenge cohort dataset will be healthy neonatal subjects approximately equal numbers of males and females scanned from the local population at :

(1) St. Thomas Hospital at King's College London, London, United Kingdom,
(2) Kawempe National Referral Hospital at Makerere University in Kampala, Uganda.

Imaging modality(ies)

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

Low field (0.064T) T2 MR images and ONLY segmentations derived from matching high field (3T or 1.5T) T2 MR 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).

For Task 2, challengers will have access to manually segmented bilateral hippocampi data from matching high field data.

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

No other information about the subject will be available.

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.

Imaging data will consist of the pediatric brain and skull shown in low resolution magnetic resonance imaging (MRI) T2 data. For Task 2, there will also be a set of bilateral hippocampi segmentations obtained from high field pediatric T2 MR image 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.

The algorithm target will be MRI segmentations of bilateral hippocampi, annotated from high field images, and obtained from matching low field T2 brain MR images

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)).

Find high field bilateral hippocampi segmentations for pediatric low field MR T2 images. Metrics include Dice, Hausdorff 95th, ASSD, and relative volume error.

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).

For each data site, all low field images were obtained by scanning on a Hyperfine SWOOP, a 0.064T MR Scanner designed for FDA-approved portable brain imaging. The system is small enough to be put in patients' rooms and can be powered by standard electrical outlets. An attached Apple iPad handles postprocessing, from which, clinically useful images may be obtained.

For high field images which form the background of hippocampi segmentations, high field T2 scans were obtained at King's College London and Makerere University on, respectively, a Philips 3T Achieva and Siemens 1.5T Sempra scanner within days of scanning on the Hyperfine SWOOP in order to obtain low and high field matching T2 image

pairs.

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

Low field images from Hyperfine SWOOP used at both sites
Magnetic field strength 0.064T, sequence type: spin echo, TR 1.5s, TE 5ms, TI 400ms
High field images at King's College London
Magnetic field strength 3T, sequence type: spin echo, TR 12s, TE 156 ms
High field images at Makerere University
Magnetic field strength 1.5T, sequence type: spin echo, TR 5s, TE 95 ms

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.

Data was acquired at (1) St. Thomas Hospital, King's College London, London, United Kingdom and (2) Kawempe National Referral Hospital, Makerere University, Kampala, Uganda

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).

All images were collected by MRI techs with experience imaging patients at the institutions listed above. All hippocampi segmentations were approved by a board certified pediatric neuroradiologist.

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.

One case encompasses a preprocessed low field T2 MR image and a single accompanying high field MR segmentation of the bilateral hippocampi where the left hippocampus has label 2, the right hippocampus has label 1 and the background has label 0.

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

The images in validation and testing used in the QA task that also have matched high field data will be considered for validation and testing in the segmentation task. Some data may drop out due to quality for segmentation. However, additional data will be acquired through the UNITY Consortium. We are currently in discussion with St. Thomas Hospital at King's College London, Aga Khan University Hospital in Karachi, Pakistan, Kalafong Hospital in Pretoria, South Africa and the Training and Research Unit of Excellence in Zomba, Malawi to obtain permission to use additional data for the challenge.

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

The total number of cases has been determined based on the available data. We collected data from different sites to introduce diversity, thereby avoiding overfitting and achieving better generalizability for the proposed models in the tasks. To prevent overlap between scans from one subject in the training cohort and those in the validation and test cohorts (due to multiple scans from a single subject), we will consider subjects rather than individual scans during the data division for training and validation. We follow the standard division of data for training (70%), validation (15%), and test (15%) and apply it to each site to get the final division for training, validation and test cohorts.

The current distribution of subjects and images across different sites is as follows:
KCL and Uganda with 75 and 95 subjects respectively.
KCL and Uganda with 47 and 107 low field scans respectively.
KCL and Uganda with 40 and 75 matching high field scans respectively.
Numbers will increase with growth in participating sites.

The number of subjects in training, validation and test cohorts is as follows.
KCL and Uganda with 53 and 67 training cases respectively.
KCL and Uganda with 11 and 14 validation cases respectively.
KCL and Uganda with 11 and 14 test cases respectively.

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 high field segmentations will be in NIFTI format where 0 represents background, 1 represents the left hippocampus and 2 represents the right hippocampus.

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.

Initial segmentation/annotation was performed manually by research assistants, and were corrected manually by an expert rater. A neuroradiologist reviewed and approved the segmentation protocol.

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.

An outline of the segmentation protocol is detailed below.

Hippocampus Segmentation Protocol in ITK-SNAP

1. Loading the Main Image:
Launch ITK-SNAP, and load the main MRI image containing the brain structure of interest.

2. Adjusting Image Contrast:
Navigate to the "Tools" menu.
Adjust Image contrast to optimize contrast settings for differentiating white matter and gray matter.

3. Defining Labels:
Label 1 (red): Represents the left hippocampus.
Label 2 (green): Represents the right hippocampus.

4. Main Toolbar:
Use the crosshair mode to select a voxel of interest, which will be displayed in all three planes.
Adjust the zoom level as per your preference.
Utilize the paint brush mode for segmenting regions.
Using a Square Brush Shape (Size 2), maintain a consistent 2x2 voxel boundary for accurate segmentation.

6. Starting in Sagittal View:
Identify a sagittal slice that exhibits the maximum extent of the hippocampus.
Begin segmenting the gray matter while adjusting the brush opacity as needed.
Pay attention to the boundaries where the anterior and posterior sides of the hippocampus meet, especially at the ventricles.

7. Lateral Segmentation in Sagittal View:
Gradually move laterally (away from the brain center) and continue segmenting the hippocampus slice by slice.
Hippocampus will gradually become smaller and thinner.

8. Medial Segmentation in Sagittal View:
After reaching the lateral boundary, move medially (toward the brain center).
Identify the "claw-shaped" structure, which may represent the ventricles.
The "thumb" of this claw determines when to split the hippocampus in the sagittal view.
Ensure a gradual split of the hippocampus without abrupt shape shifts, maintaining the integrity of the structure.

9. Segmenting the Head and Tail:
After the split, continue segmentation until the anterior head of the hippocampus disappears.
The posterior tail will gradually vanish after the head disappears.

10. Coronal View Segmentation:
Begin segmentation from the posterior side in the coronal view.
Progress slice by slice towards the anterior side, identifying contrast differences as well as consistency in the shape shifting.
Employ crosshair mode more frequently in the coronal view to validate segmentation from the sagittal view.

11. Axial View Segmentation:
Start segmentation from the superior side in the axial view.
Move slice by slice towards the inferior side.

As you approach the endpoint in the axial view, ensure that the segmentation concludes before the corpus callosum joins. This serves as your reference for the boundary in the axial view.

12. Maintaining 2x2 Voxel Density:
Throughout the segmentation process, prevent holes or any 1x1 voxel stranding.
Maintaining a 2x2 voxel density is crucial for accurate boundary preservation.

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.

Our neuroradiologist has a number of decades of experience in assessing hippocampi segmentations.

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.

n/a

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.

All images were first converted to NIFTI from DICOM using dcm2nii. These low field images are defaced for full anonymization using SynthRad2023 processing pipeline's anonymize.py. No other information will be included that could include risk of reidentification. The repository is freely available at this link : https://github.com/SynthRAD2023/preprocessing

Each orthogonally acquired trio of anisotropic low field T2 images are reconstructed to form a single higher resolution combined isotropic T2 image as described in [2].
These images and all background high resolution images then follow a processing pipeline of bias field correction, skullstripping and linear registration as detailed in [3].

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.

We will provide intra-rater reliability scores to estimate the magnitude of human error in image annotation.

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)

We will employ four metrics, Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), Average Symmetric Surface Distance (ASSD), and Relative Volume Error (RVE) on the hippocampi segmentation predictions.

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

Metric choice was chosen from recommendations found in [5] which presents a checklist which authors of biomedical image analysis challenges are encouraged to include in their submission when giving a paper on a challenge into review. The purpose of the checklist is to standardize and facilitate the review process and raise interpretability and reproducibility of challenge results by making relevant information explicit.

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.

The ranking system described will be used to obtain a fair and complete comparison across all metrics between participants. We desire a ranking scheme that is empirical, objective, and provides both participants and organizers with a clear understanding of the 'best' challenge solution.

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

Submissions that have missing results or cannot provide results because inference fails on a test case will be assigned the lowest possible score for the failed test case using metrics relevant to the task. For example, if a missing result or failed inference occurs in Task 2, values would be set to 0% for DSC, the maximum size of the image for both HD and ASSD, and 100% for relative volume error.

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

We will take the mean across their respective metrics in order obtain rankings. Standard deviations will also be calculated and given.

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.

Permutational analyses will be employed to evaluate uncertainties in rankings. Each team's performance will be evaluated separately for the left and right hippocampus structures. Rankings for the measures will be combined

by averaging, and statistical significance will be assessed exclusively for segmentation performance measures. This assessment will involve permuting the relative ranks for each segmentation measure, considering left and right hippocampus for segmentation, per subject in 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

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.

[1] Bottani S, Burgos N, Maire A, Wild A, Ströer S, Dormont D, Colliot O; APPRIMAGE Study Group. Automatic quality control of brain T1-weighted magnetic resonance images for a clinical data warehouse. Med Image Anal. 2022 Jan;75:102219. doi: 10.1016/j.media.2021.102219. Epub 2021 Sep 3. PMID: 34773767.

[2] Deoni, S. C., O'Muircheartaigh, J., Ljungberg, E., Huentelman, M., & Williams, S. C. (2022). Simultaneous high-resolution T2-weighted imaging and quantitative T-2 mapping at low magnetic field strengths using a multiple TE and multi-orientation acquisition approach. Magnetic Resonance in Medicine, 88(3), 1273-1281

[3] Rouhi, R., Tanedo, J., Wei, I.M., Valder, M., Zanjal, S., Gajawelli, N., ... & Lepore, N. (2023, November). Automated Segmentation of the Hippocampus in Pediatric Imaging. In 2023 19th International Symposium on Medical Information Processing and Analysis (SIPAIM) (pp. 1-5), IEEE.

[4] Ravi D; Alzheimer's Disease Neuroimaging Initiative; Barkhof F, Alexander DC, Puglisi L, Parker GJM, Eshaghi A. An efficient semi-supervised quality control system trained using physics-based MRI-artefact generators and adversarial training. Med Image Anal. 2024 Jan;91:103033. doi 10.1016/j.media.2023.103033 Epub 2023 Nov 20. PMID 38000256

[5] Lena Maier-Hein, Annika Reinke, Michal Kozubek, Anne L. Martel, Tal Arbel, Matthias Eisenmann, Allan Hanbury, Pierre Jannin, Henning Müller, Sinan Onogur, Julio Saez-Rodriguez, Bram van Ginneken, Annette Kopp-Schneider, Bennett A. Landman, BIAS: Transparent reporting of biomedical image analysis challenges, Medical Image Analysis, Volume 66, 2020, 101796, ISSN 1361-8415, https://doi.org/10.1016/j.media.2020.101796.

Further comments

Further comments from the organizers.

N/A