Anatomical Brain Barriers to Cancer Spread: Segmentation from CT and MR Images : Structured description of the challenge design

Remark: This challenge have been slightly modified. All changes are highlighted in red.

CHALLENGE ORGANIZATION

Title

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

Anatomical Brain Barriers to Cancer Spread: Segmentation from CT and MR Images

Challenge acronym

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

ABCs

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.

The goal of this challenge is to identify best methods to segment brain structures that serve as barriers to cancer spread, for use in computer assisted target definition for radiotherapy plan optimization. Target delineation is a major task in the radiotherapy workflow since it influences the overall outcome of the treatment. Accurate delineation of structures for treatment planning is associated with better local tumor control and reduced radiation dose to non-target tissues leading to improved therapeutic index.

Accuracy can be improved through automation of target definition by identifying natural anatomical barriers to tumor spread. The barriers are brain structures that are not routinely outlined during the treatment planning for radiotherapy plan optimization. Advanced expertise and time is required to identify anatomy on computed tomography images used for treatment planning. Automation of the barrier delineation will make the treatment planning workflow more efficient leading to improved scheduling and potentially increasing the size of patient panels. With improving precision of radiation dose delivery and increasing number of patients treated in the cancer centers worldwide, algorithmic assistance for the target definition is becoming a necessity. Creating standardized definition of the target and normal anatomy will also create an excellent educational resource for clinicians working at centers with low patient volume, and for medical residents.

Segmentation of the brain structures is a challenging task as different structures can be appreciated more or less favorably on different imaging modalities. For example, the skull is typically segmented using a CT image, whereas falx cerebri is better seen on an MR T1-weighted image. Furthermore, as multi-modality images are usually acquired at different time points they could present with subtle differences even for brain imaging. This presents a

unique technological challenge as information from multi-modality imaging is used by the radiation oncologist to define the clinical target volume for each individual patient's disease.

The proposed challenge will provide participants with 60 cases, each consisting of a planning CT scan and two MR scans (post-contrast T1-weighted and T2-weighted FLAIR volumes) with 12 normal structures previously contoured using co-registered CT and MR at a single institution using an imaging protocol defined for glioma radiotherapy treatment planning. This dataset is valuable to the medical imaging and radiation oncology communities as it could advance algorithm development for automated segmentation of these structures, which in turn could be used for automated radiotherapy target definition.

Challenge keywords

List the primary keywords that characterize the challenge.

Segmentation, artificial intelligence, brain imaging, cross-modality, computed tomography (CT) imaging, magnetic resonance imaging (MRI), cancer.

Year

The challenge will take place in ...

2020

FURTHER INFORMATION FOR MICCAI ORGANIZERS

Workshop

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

none

Duration

How long does the challenge take?

Half 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 expect to have 25 teams to participate in the challenge.
This number was estimated based on our past experience hosting grand challenges at the annual American Association of Physicist in Medicine meeting. Between the 2017 and 2019 auto-segmentation challenges we saw on average 15 participants. BraTS challenge hosted 12, 16, 54, and 64 participants in four consecutive year starting from 2015. Since the MICCAI meeting hosts a larger number of attendees and auto-segmentation challenges are more common in this meeting, we thought a number of 25 participants was a conservative estimate number of participants for the first-time challenge. Once the challenge is formally accepted, we will advertise the challenge by contacting individual teams working on segmentation across the world via email and through Social Media posts.

Publication and future plans

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

We plan to coordinate a publication on the challenge results.

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.

Room with internet connection.

TASK: Segmentation of brain structures that serve as anatomical barriers to cancer spread

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 goal of this task is to segment brain structures, falx cerebri, tentorium cerebelli, transverse and sagittal brain sinuses, ventricles, and cerebellum to use for automated definition of the target for radiotherapy treatment.

Keywords

List the primary keywords that characterize the task.

Segmentation, brain, computed tomography (CT), magnetic resonance imaging (MRI)

ORGANIZATION

Organizers

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

Nadya Shusharina (Massachusetts General Hospital, Harvard Medical School)
Thomas Bortfeld (Massachusetts General Hospital, Harvard Medical School)
Carlos Cardenas (The University of Texas MD Anderson Cancer Center)
Jinzhong Yang (The University of Texas MD Anderson Cancer Center)

b) Provide information on the primary contact person.

Nadya Shusharina

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.

Examples:

- One-time event with fixed submission deadline
- Open call
- Repeated event with annual fixed submission deadline

One time event.

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.

grand-challenge.org

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

https://abcs.mgh.harvard.edu/

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.

No additional data allowed.

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 not eligible for awards and not listed in leaderboard.

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

1000 EUR for 1st prize
500 EUR for 2nd prize
300 EUR for 3rd prize

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 will be announced publicly. The participants can choose to be listed on the Leaderboard.

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).

Three winning team's members will co-author a challenge paper that will be published first.

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.

Before the conference the participants will be provided with the test data and required to submit their results within 2-day period. The data will be released upon request during evaluation period of two weeks. The predicted segmentations will be uploaded to the challenge server. The results of evaluation will be announced during the conference.

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.

15 test cases will be available for off-line testing. The submissions will be processed and results will be posted on the web-based dashboard. The participants can also evaluate their algorithms with the metrics that will be used for ranking calculations. (Volumeric Dice Similarity Coefficient and Surface Dice Similarity Coefficient: https://github.com/deepmind/surface-distance.git)

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

Release of training cases: June 1, 2020
Release of off-line test cases: July 1, 2020
Registration: March 1, 2020 - August 31, 2020
Submission date: during two weeks before the conference.

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).

Ethics approval is not necessary for the data. The IRB protocol was approved for the data use.

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 NC SA.

Additional comments: The data can be used for developing segmentation algorithms as well as other educational purposes. Any commercial use of data is forbidden.
The data will be available for download from the challenge website. The participants will be required to register and will receive an approval through their institutional e-mail. The detailed instructions for data access will be

provided.

Additional comments: The data can be used for developing segmentation algorithms as well as other educational purposes. Any commercial use of data is forbidden.
The data will be available for download from the challenge website. The participants will be required to register and will receive an approval through their institutional e-mail. The detailed instructions for data access will be provided.

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 participants will be provided a link to the evaluation source code.

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

The winning teams are encouraged to publish provide as many algorithm details as possible including brief description of the network architecture, network hyperparameters, description of image pre- and post-processing to be included in the challenge publication. We will make access to monetary prizes conditional on the publication of code.

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.

The prizes are courtesy of our sponsor RaySearch Laboratories.

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

Decision support, Training, Treatment planning, Research, Education.

Additional points: The participating algorithms will provide the basis for computer-assisted delineation of the target for radiotherapy cancer treatment. Implementation of the algorithms will provide a framework for more consistent among clinicians target definition, will serve as training platform for medical residents, and will be used as a research tool for further developments.

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 glioblastoma and low-grade glioma who are undergoing radiotherapy treatment.

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

Glioblastoma and low-grade glioma patients treated with radiotherapy with curative intent at MGH.

Imaging modality(ies)

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

Post-operative CT and multimodal MRI scans (post-contrast T1-weighted and T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) volumes) of glioblastoma and low-grade glioma.

Context information

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

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

Manually delineated brain structures: falx cerebri, tentorium cerebelli, transverse and sagittal brain sinuses, ventricles, cerebellum.

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

Patients diagnosed with glioblastoma and low-grade glioma.

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.

Post-tumor-resection brain shown in computed tomography (CT) and magnetic resonance imaging (MRI) 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.

Brain structures that serve as natural anatomical barriers to tumor cells infiltration.

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.

Additional points: Find accurate brain structures segmentation using CT and MR images.

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).

CT Scanners (GE);
MR Scanners (Siemens, GE)

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

CT: standard protocol for radiotherapy treatment planning;
MR: standard protocols for diagnostic imaging.
The imaging parameters will be released as a part of the dataset.

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.

MGH, Radiation Oncology.

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).

The images were acquired by certified experienced personnel.

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.

Training and test cases both represent CT, MR (T1-weighted, and T2-weighted FLAIR) images of human brain after surgical resection of tumor mass. The MR and CT images were co-registered using rigid registration by a trained medical professional as part of the clinical workflow at the time of patient's treatment. All cases are annotated with the brain structures.

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

75 cases to be split into training (45 image sets and the ground truth), off-line test (15 image sets), and test (15 image sets).

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

45 cases will be available for training at the earliest time (image sets + ground truth segmentations), 15 will be available to offline-test submissions after 4 weeks (image sets only) and the final test (15 cases) will be released on the first day of the conference (image sets only). The first 15 offline-test cases will be used to calculate the scores that will be posted on a web-based dashboard. We will release the ground truth segmentations for the offline-test cases two weeks before the conference to assist with final model training. The other 15 cases will be used to assess the final algorithm's performance by scoring the final results and posting those on the dashboard during the MICCAI meeting.

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.

Cases are annotated with the same set of non-overlapping structures. All images show the brain after tumor resection.

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.

Manual image annotation, one annotator. We will randomly pick five cases and ask 3 raters draw the manual contours for each structure. An inter-rater accuracy will be analyzed based on these manual contours (see Yang, et al. Med Phys, 2018). Inter-rater accuracy will be provided as a reference point to assess the relative performance of the algorithms.

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.

Annotation of brain structures was performed by certified medical dosimetrist under supervision of neuro-anatomist who is performing delineations of the brain structures for conformal radiotherapy treatments (see ref. Ratiu & Talos).

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.

Professionally trained medical dosimetrist with 4 years of experience and medically trained neuro-anatomist with over 20 years of experience.

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.

Annotation was done on image fusion of CT scan used for radiotherapy treatment planning and diagnostic T1-weighted MRI.

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.

For each case, MR images were rigidly aligned with the CT image.

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.

Intra-annotator variability.

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

Errors from image registration.

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),
Surface Dice Similarity Coefficient

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

Similarity of surface of the structures is relevant for the radiotherapy target definition.

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.

Performance rank of submissions is the rank in an ordered list of performance scores. The performance score is the sum of average surface Dice similarity coefficient for all structures and all cases, and the average Dice similarity coefficient for all structures and all cases.

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

The performance score of submissions with missing results on some of cases will be reduced proportionally to the number of cases missed. Missing segmentations will be given DSC of zero and surface DSC of zero.

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

The proposed ranking scheme uniformly assesses algorithm performance across all structures and all cases.

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.

Missing data will be treated as zeros to penalize incomplete segmentations. Scipy and in-house DSC and SDSC code will be used to assess the submissions, calculate the average DSC, and establish the rank.

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

The goal of the challenge is to assess algorithms toward producing clinically acceptable segmentations. As such, we chose to rank on the sum of all DSC and SCDC to uniformly weigh all the different anatomical structures, as well as to penalize submissions with incomplete data. In the future, ranking algorithms will be reconsidered on per-structure basis given the knowledge gained from the present challenge.

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.

TASK: Segmentation of brain structures used for radiotherapy treatment plan optimization

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 goal of the task is to segment the brainstem, structures of visual pathway, optic chiasm, optic nerves, eyes, and lens, structures of auditory pathway, cochlea, and lacrimal glands to use in radiotherapy treatment plan optimization.

Keywords

List the primary keywords that characterize the task.

Segmentation, brain, computed tomography (CT), magnetic resonance imaging (MRI)

ORGANIZATION

Organizers

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

Carlos Cardenas (The University of Texas MD Anderson Cancer Center)
Jinzhong Yang (The University of Texas MD Anderson Cancer Center)
Nadya Shusharina (Massachusetts General Hospital, Harvard Medical School)
Thomas Bortfeld (Massachusetts General Hospital, Harvard Medical School)

b) Provide information on the primary contact person.

Carlos Cardenas

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.

Examples:

- One-time event with fixed submission deadline
- Open call
- Repeated event with annual fixed submission deadline

One time event.

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.

grand-challenge.org

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

https://abcs.mgh.harvard.edu/

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.

No additional data allowed.

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 not eligible for awards and not listed in leaderboard.

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

The winner will be determined by averaging results of the two tasks.

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.

Same as in Task 1.

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).

Same as in Task 1.

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.

Same as in Task 1.

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.

Same as in Task 1.

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

Same as in Task 1.

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).

Same as in Task 1.

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 NC SA.

Additional comments: Same as in Task 1.

Additional comments: Same as in Task 1.

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.

Same as in Task 1.

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

Same as in Task 1.

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.

The prizes are courtesy of our sponsor RaySearch Laboratories.

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

Decision support, Treatment planning.

Additional points: Implementation of the algorithms in radiotherapy treatment planning will reduce variability of normal anatomy delineation improving quality of the treatment plans which justifies an effective treatment.

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 with brain tumors who are undergoing radiotherapy treatment.

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

Same as in Task 1.

Imaging modality(ies)

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

Same as Task 1.

Context information

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

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

Manually delineated brain structures, structures of visual and auditory pathways: brainstem, optic chiasm, optic nerves, cochleas, eyes, lens, lacrimal glands.

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

Same as in Task 1.

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.

Same as in Task 1.

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.

Anatomical structures used for radiotherapy treatment plan optimization.

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.

Additional points: Find accurate brain structures segmentation using CT and MR images.

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).

CT Scanners (GE);
MR Scanners (Siemens, GE).

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

Same as in Task 1.

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.

MGH, Radiation Oncology.

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).

The images were acquired by certified experienced personnel.

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.

Same as in Task 1.

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

Same as in Task 1.

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

Same as in Task 1.

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.

Same as in Task 1.

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.

Manual image annotation, multiple annotators.

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.

Annotation of brain structures was performed by the residents of Harvard Medical School according to the clinical guidelines established at MGH for radiotherapy treatment plans. The plans were approved by the treating radiation oncologists to treat the patients (see refs Brouwer et al, and Basu & Bhaskar)

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.

Medically trained residents.

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.

Annotation was done on CT scan used for radiotherapy treatment planning.

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.

Same as in Task 1.

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.

Intra- and inter-annotator variability.

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

Errors from image registration.

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),
Surface Dice Similarity Coefficient.

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

Similarity of surface of the structures is required to avoid extensive manual post-processing.

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.

Same as in Task 1.

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

Same as in Task 1.

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

Same as in Task 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.

Same as in Task 1.

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

Same as in Task 1.

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.

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.

Surface Dice Similarity Coefficient:
Nikolov S, Blackwell S, Mendes R, De Fauw J, Meyer C, Hughes C, et al.
Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy 2018;arXiv:1809.04430.
Brouwer C.L., Steenbakkers R.J.H.M., Bourhisa J., et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol 2015; 117: 83-90.
Trinanjan Basu and Nithin Bhaskar (November 5th 2018). Overview of Important "Organs at Risk" (OAR) in Modern Radiotherapy for Head and Neck Cancer (HNC), Cancer Survivorship, Dil Afroze, IntechOpen, DOI: 10.5772/intechopen.80606. Available from: https://www.intechopen.com/books/cancer-survivorship/overview-of-important-organs-at-risk-oar-in-modern-radiotherapy-for-head-and-neck-cancer-hnc-.
P.Ratiu, I.-F.Talos. Cross-sectional atlas of the brain and DVD. Harvard University Press. 2006.
N. Shusharina, J. Soderberg, D. Edmunds, F. Lofman, H. Shih, T. Bortfeld. Automated delineation of the clinical target volume using anatomically constrained 3D expansion of the gross tumor volume. Radiother Oncol 2020; 146: 37-43.
J. Yang, H. Veeraraghavan, S. G. Armato III, et al. Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. Med Phys 2018; 45:4568-4581.

Further comments

Further comments from the organizers.

We will make the data openly available after the challenge. Based on the consensus from all organizers, we will make the ground truth of the test data available 6 months after the challenge.

Further comments

Further comments from the organizers.

We will make the data openly available after the challenge. Based on the consensus from all organizers, we will make the ground truth of the test data available 6 months after the challenge.