Intracranial Aneurysm and Intracranial Artery Stenosis Detection and Segmentation Challenge: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Intracranial Aneurysm and Intracranial Artery Stenosis Detection and Segmentation Challenge

Challenge acronym

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

INSTED

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.

Stroke, including ischemic stroke and hemorrhagic stroke, is a high prevalence and life-threatening cerebrovascular condition. Intracranial aneurysms stand as a common cause of hemorrhagic strokes, whereas intracranial artery stenosis caused by atherosclerosis emerges as the predominant cause of ischemic strokes. Accurate assessment of intracranial aneurysm and intracranial artery stenosis is important for the diagnosis and treatment of cerebrovascular diseases. Magnetic resonance angiography (MRA) is widely used to visualize the cerebral arterial tree for disease diagnosis, especially for the screening of intracranial atherosclerosis and aneurysm. Accurate lesion detection and segmentation are important for quantitative analysis of cerebrovascular diseases, such as estimation of degree of luminal stenosis and the size of aneurysm. However, manual detection and segmentation of MRA can be challenging even for experts given the complex network of cerebral arteries with substantial inter-individual variations, and weak signals in small vessels due to slow or in-plane blood flow. Time-of-flight (TOF) MRA is the most widely used non-invasive imaging technique to depict the anatomy of the cerebrovascular tree without use of contrast agents. Due to its non-invasive nature and absence of ionizing radiation, TOF-MRA imposes minimal harm on the human body, making it suitable for clinical screening of cerebrovascular diseases. Private datasets and annotations are commonly used in recent cerebral artery segmentation studies, whereas open-accessible large-scale TOF-MRA data with well-labeled bounding boxes and segmentation masks of intracranial aneurysm and intracranial artery stenosis are rare, hindering the development and validation of reliable automatic intracranial lesion detection and segmentation algorithms. Hence, we attempt to host the first intracranial aneurysm and intracranial artery stenosis detection and segmentation challenge in MICCAI 2024. To the best of our knowledge, this is the first challenge focusing on the detection and segmentation of both intracranial arterial stenosis and intracranial aneurysm based on TOF-MRA data.
In this challenge, the task is to detect and segment lesion from 3D TOF-MRA images acquired from a cohort

consisting of healthy volunteers and patients with intracranial artery stenosis or intracranial aneurysms. Precise detection and segmentation of lesions will be helpful for the identification and quantitative characterization of stenosis and intracranial aneurysm, which is important in clinical practice.
Notably, this challenge is also the official grand challenge of 2024 Annual meeting of Society for Magnetic Resonance Angiography (SMRA, https://society4mra.org/), which was founded in 1989 as the MR Angio Club to bring together scientists, clinicians and industry with a common interest in MR Angiography (MRA). As an MRA imaging academic society, SMRA was holding challenges since 2021, and all of them were also joint MICCAI challenge.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

Intracranial aneurysm, Intracranial artery stenosis, detection, segmentation, TOF-MRA

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.

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.

An estimation of 200 teams based on similar challenge held previously (SMRA&MICCAI; 2023 Challenge: Cerebral artery segmentation Challenge, SMRA&MICCAI; 2022 Challenge: Carotid Vessel Wall Segmentation and Atherosclerotic Lesion Diagnosis Challenge)

Publication and future plans

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

No publication plan currently. We are awaiting the committee's decision regarding the prize amount and publication plans, and it will be determined upon the release of the challenge.

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.

We intend to use https://www.codabench.org/ for the challenge.

TASK 1: Intracranial aneurysm and intracranial artery stenosis detection and segmentation

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.

Stroke, including ischemic stroke and hemorrhagic stroke, is a high prevalence and life-threatening cerebrovascular condition. Intracranial aneurysms stand as a common cause of hemorrhagic strokes, whereas intracranial artery stenosis caused by atherosclerosis emerges as the predominant cause of ischemic strokes. Accurate assessment of intracranial aneurysm and intracranial artery stenosis is important for the diagnosis and treatment of cerebrovascular diseases. Magnetic resonance angiography (MRA) is widely used to visualize the cerebral arterial tree for disease diagnosis, especially for the screening of intracranial atherosclerosis and aneurysm. Accurate lesion detection and segmentation are important for quantitative analysis of cerebrovascular diseases, such as estimation of degree of luminal stenosis and the size of aneurysm. However, manual detection and segmentation of MRA can be challenging even for experts given the complex network of cerebral arteries with substantial inter-individual variations, and weak signals in small vessels due to slow or in-plane blood flow. Time-of-flight (TOF) MRA is the most widely used non-invasive imaging technique to depict the anatomy of the cerebrovascular tree without use of contrast agents. Due to its non-invasive nature and absence of ionizing radiation, TOF-MRA imposes minimal harm on the human body, making it suitable for clinical screening of cerebrovascular diseases. Private datasets and annotations are commonly used in recent cerebral artery segmentation studies, whereas open-accessible large-scale TOF-MRA data with well-labeled bounding boxes and segmentation masks of intracranial aneurysm and intracranial artery stenosis are rare, hindering the development and validation of reliable automatic intracranial lesion detection and segmentation algorithms. Hence, we attempt to host the first intracranial aneurysm and intracranial artery stenosis detection and segmentation challenge in MICCAI 2024. To the best of our knowledge, this is the first challenge focusing on the detection and segmentation of both intracranial arterial stenosis and intracranial aneurysm based on TOF-MRA data.
In this challenge, the task is to detect and segment lesion from 3D TOF-MRA images acquired from a cohort consisting of healthy volunteers and patients with intracranial artery stenosis or intracranial aneurysms. Precise detection and segmentation of lesions will be helpful for the identification and quantitative characterization of stenosis and intracranial aneurysm, which is important in clinical practice.
Notably, this challenge is also the official grand challenge of 2024 Annual meeting of Society for Magnetic Resonance Angiography (SMRA, https://society4mra.org/), which was founded in 1989 as the MR Angio Club to bring together scientists, clinicians and industry with a common interest in MR Angiography (MRA). As an MRA imaging academic society, SMRA was holding challenges since 2021, and all of them were also joint MICCAI challenge.

Keywords

List the primary keywords that characterize the task.

Intracranial aneurysm, Intracranial artery stenosis, detection, segmentation, TOF-MRA

ORGANIZATION

Organizers

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

Huijun Chen; Xihai Zhao; Rui Li; Haokun Li; Haozhong Sun; Ziming Xu; Haining Wei; Yan Li; Jiaqi Dou; Xueyan Li from Tsinghua University, China.

b) Provide information on the primary contact person.

Haokun Li (lhk23@mails.tsinghua.edu.cn)

Life cycle type

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

Examples:

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

One-time event with fixed conference submission deadline.

Challenge venue and platform

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

MICCAI, SMRA

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

We intend to use https://www.codabench.org/

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

https://www.codabench.org/competitions/2139/

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.

Private data is NOT 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.

Members of the organizers' institutes may participate but not eligible for awards and not listed in the leaderboard.

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

There will be an award for the top three teams. As our challenge is funded by the SMRA committee, we are awaiting the committee's decision regarding the prize details, and the details will be disclosed upon the release of our challenge.

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.

All teams are encouraged to submit a report introducing their methods. Five SMRA winning teams will be invited to both SMRA 2024 (Nov 12-15, 2024) and MICCAI 2024 (Oct 6-10, 2024) to present their methods in order to foster communication and collaboration between clinicians and image processing experts. The challenge session will be conducted online as part of MICCAI and in person during SMRA. Winning teams will be given the opportunity to present their work at both conferences. Participating teams can choose whether their performance results will be made public.

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).

There are no publication requirements in this challenge. The participants may choose the author list (participating team) for their presentations at SMRA 2024 and MICCAI 2024. Data used in this challenge are limited to the purpose of developing intracranial aneurysm and intracranial artery stenosis detection and segmentation algorithms. Other requests require approval from the organizers. After the challenge concludes, new teams will still have access to the training data. Additionally, the test data will be made available following the conclusion of the challenge (No need to submit to the organizer for evaluations after the challenge). For publications, the teams can publish their methods and results from the images of the challenge. No commercial uses are allowed. The challenge organizers will not publish a challenge paper 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.

Participating teams are required to submit their Docker containers containing their algorithms through the platform provided on the challenge website. The organizing team will apply the container to the hidden test set to assess the algorithm's performance on it.

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.

A Python demo for computing the final metric will be released. Participants can utilize it to evaluate performance on the validation data split from the training set. The Docker for testing can be submitted multiple times, and participants will only receive feedback on whether the Docker runs successfully or not. Only the final submission will be officially considered for the challenge results.

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

Training data release: June 1th, 2024
Validation submission open: July 1th 2024
Validation submission closed: July 25th 2024
Test submission open: July 25th 2024
Test submission closed: Aug 15th 2024
Announce winner: Sep 1st, 2024
Presentation at MICCAI 2024: Oct 2024
Presentation at SMRA 2024: Nov 2024

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).

Yes.
The data got approved by (1) The Institution Review Board of Tsinghua University, 20170002
(2) Wuhan Union Hospital of China.
(3) The Institution Review Board of First Affiliated Hospital of PLA General Hospital, 2018KY-KS001
(4) The Institution Review Board of First Affiliated Hospital of Aerospace Center Hospital, 2021-031
(5) The Research Ethics Committee of The Second Hospital of Hebei Medical University, 2018-C021
(6) The Institution Review Board of First Affiliated Hospital of Tangshan Worker's Hospital, GRYY-LL-2018-39

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.

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 code will be put on GitHub. See the challenge website for details.

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

Participants will choose whether to release their test 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 challenge is funded by the SMRA 2024.

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
-

Diagnosis, Screening, Intervention planning.

Task category(ies)

State the task category(ies)

Examples:

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

Detection, 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 are subjects who developed intracranial artery stenosis or intracranial aneurysm.

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

The challenge cohort is from clinical data collected at (1) Center for Biomedical Imaging Research, Tsinghua University, Beijing, China. (2) Wuhan Union Hospital of China, Hubei, China
(3) The Institution Review Board of First Affiliated Hospital of PLA General Hospital
(4) The Institution Review Board of First Affiliated Hospital of Aerospace Center Hospital
(5) The Research Ethics Committee of The Second Hospital of Hebei Medical University
(6) The Institution Review Board of First Affiliated Hospital of Tangshan Worker's Hospital

Imaging modality(ies)

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

MRI (3D-TOF-MRA)

Context information

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

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

The MRI images and segmentation labels will be provided in nii.gz format, and the information of bounding boxes will be given in json format.

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

No additional information will be given other than the medical images.

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.

Head MR exams using 3D TOF-MRA sequence.

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.

Intracranial aneurysms and the cerebral arteries within stenosis regions

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 highly accurate detection and segmentation algorithm for intracranial aneurysm and intracranial artery stenosis.

User satisfaction, Robustness, Accuracy of detection and segmentation, Applicability.

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).

3.0T MR scanners with commercialized head coil.

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

3D TOF: All the intracranial TOF-MRA was acquired on 3T MR scanners by the routine imaging protocol used in each institution, which covers various MR scanners and imaging protocols to avoid overfitting. None of the data in the previous challenges we have held will be reused. In fact, discussion with the SMRA committee is ongoing to explore the possibility of collecting international data. This endeavor aims to augment the diversity of the dataset and mitigate the risk of overfitting.

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.

(1) Center for Biomedical Imaging Research, Tsinghua University, Beijing, China.
(2) Wuhan Union Hospital of China, Hubei, China
(3) The Institution Review Board of First Affiliated Hospital of PLA General Hospital
(4) The Institution Review Board of First Affiliated Hospital of Aerospace Center Hospital
(5) The Research Ethics Committee of The Second Hospital of Hebei Medical University
(6) The Institution Review Board of First Affiliated Hospital of Tangshan Worker's Hospital

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).

Trained MR operators obtained the head MR exams using a standardized protocol.

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 3D TOF-MRA scans of the brain.
The bounding box of lesion (IA or stenosis) and the segmentation in the box are annotated for each case in the data set.

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

Training cases: 160, Closed test cases: 40. There are no separate validation cases for this challenge

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

The number of training cases was decided based on the limited manual annotations. Manual judgement and annotation for lesion of a 3D MR scan are very time-consuming (2 hours per case). Similar challenges on grand-challenge.org have similar amount of data. For example, the dataset of TDSC-ABUS2023 challenge for tumor detection in 3D breast ultrasound images consisted of 200 3D volumes in total with only 100 cases for training. The dataset of CAS2023 challenge for cerebral artery segmentation included 150 3D brain TOF-MRI images. Additionally, it's worth noting that multiple lesions are often present in a single scan. Despite the relatively limited number of cases, we believe that the dataset is adequate for training a robust model.
Training cases: 160 (The relatively large number of data were used for training a robust model. Participants could split this set to validate their algorithm on their own).
Closed test cases: 40 (The relatively large number of data were used for a fair final leaderboard).

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.

The challenge cohort was derived from clinical data collected at multiple centers. Subjects in the dataset may be healthy, present with one or more IA, or exhibit one or more stenosis. The location of lesions (IA or stenosis) is specified in the following artery segments: anterior cerebral artery, middle cerebral artery, posterior cerebral artery, internal carotid artery, vertebral artery, or basilar artery. The proportion of healthy subjects, subjects with IA and subjects with stenosis is 1:2:2, and this proportion holds in both training and test datasets.

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 on coarse annotation by automatic methods. Six annotators will be involved.

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.

1) The annotation is performed using ITK-SNAP software (Paul Yushkevich, Jilei Hao, Alison Pouch, Sadhana Ravikumar and colleagues at the Penn Image Computing and Science Laboratory (PICSL) at the University of Pennsylvania, United States; http://www.itksnap.org/pmwiki/pmwiki.php).
2) The annotation of the intracranial vascular stenosis and aneurysms is semi-automatic. Pre-labeling of the vascular stenosis and aneurysms is performed by a threshold algorithm using the threshold module implemented in ITK-SNAP.
3) Finally, five junior reviewers fine-tuned the annotation results of the intracranial vascular stenosis and aneurysms structure, and one senior reviewer peer-reviewed all the annotations, a consensus was reached among all the reviewers.

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 intracranial vascular stenosis and aneurysms structure is manually annotated by five junior reviewers with more than 2 years of professional experience and one senior reviewer with more than 15 years of professional 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.

Six medically trained reviewers prepared the annotations. Five junior reviewers with more than 2 years of professional experience used the same standard to annotate the intracranial vascular stenosis and aneurysms. And then a senior reviewer with more than 15 years of professional experience peer-reviewed all the annotations and reached a consensus among all the reviewers.

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 all cases, the preprocessing methods involve:
a) Manually review each image for quality.
b) Remove private personal information and convert DICOM files into nii.gz files.
There is no preprocessing conducted beyond the methods outlined above.

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.

Possible error may originate from poor image quality or blood flow artifacts so that clear boundaries cannot be identified confidently. And some aneurysms being small, as well as the presence of intramural hematomas, could lead to inaccurate annotations in detection and segmentation. The detection annotation error will not surpass 5% of the total number of lesions. Additionally, the segmentation annotation error magnitude should be comparable to that observed in other small lesion or tumor segmentation challenges, thus not deemed severe. Errors are equally distributed in training, validation and test cases, as the cases are randomly divided into these subsets after annotation.

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

None.

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)

Detection metrics
[1] F1 score of each class (intracranial aneurysm or stenosis)
[2] Average precision score at IoU threshold of 0.5 and 0.75 (AP50 and AP75) of each class
In the case of a given fixed threshold, the F1 score of all detected lesion boxes in each lesion class is calculated. For each lesion class (intracranial aneurysm or stenosis), by progressively changing the threshold, precision and recall of the detected lesion boxes corresponding to each threshold are computed. The precision-recall curve is plotted, and the area under the curve is calculated to obtain the AP value.

Segmentation metrics
[1] Dice similarity coefficient (DSC) of each class
[2] 95% Hausdorff Distance (95HD) of each class

The dice similarity coefficient and AHD are computed separately with respect to each class, and are computed only in the regions of lesions.

Clinical metrics
[1] mean absolute error (MAE) for percentage of stenosis
[2] MAE for aneurysm long axis length
[3] MAE for aneurysm short axis length
As for the first metric, the percentage of stenosis is defined as (1-DS/DN), where DS stands for the diameter of the artery at the site of the most severe degree of stenosis, and DN means the diameter of normal vessel near the stenosis region. The metric calculates the MAE between the true percentage of stenosis determined by an experienced clinician and the percentage of stenosis calculated based on the segmentation result of a participant's algorithm at each detected intracranial artery stenosis lesion.
As for the second and the third metric, MAE is calculated between the true long or short axis length calculated based on the ground truth segmentation and the long or short axis length calculated based on he segmentation result of a participant's algorithm at each detected intracranial aneurysm lesion.

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

In clinical practice, a predetermined threshold is established for model utilization, and the F1 score serves as a typical metric for evaluating a detection model based on this fixed threshold. To further evaluate the method independently of the threshold, the average precision score is employed as it provides a common metric for assessing a detection model, irrespective of the threshold setting.

Dice is a common metric to evaluate segmentation performance. 95HD is used to assess the goodness of fit of the segmentation to the ground truth boundary. Since the purpose of segmentation in this task is to specify the areas of lesion in the detected lesion boxes, the segmentation metrics are only calculated in the regions of lesions.

Percentage of stenosis is a crucial clinical indicator for the severeness of intracranial artery stenosis, while long and short axis length of intracranial aneurysm strongly correlates with its risk of rupture. Measuring the calculation accuracy of these clinically significant indicators can directly assess the algorithm's potential for future clinical applications.

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.

This ranking scheme is developed to take four important metrics of each class of lesion into account. It provides a balanced scheme to judge whether the method can achieve accurate detection and segmentation with better performance simultaneously. Since this challenge primarily focuses on the detection of lesions, the weights assigned to detection metrics are greater than those allocated to segmentation metrics. As AP remains independent of the threshold, it can better reflect the ability of the submitted model, so it has larger weight than F1 score. Clinical metrics can assess the algorithm's potential for future clinical applications, so it is assigned weight larger than segmentation.

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

Missing results (failed segmentation or failed detection) in the test cases will be given the worst possible score.

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

(1) Calculate the F1 score, average precision score, dice similarity coefficient and AHD of each class for all cases. Calculate the clinical metrics for all cases.
(2) Use the following formula to calculate the final score, and use it to determine the final ranking.
APIA = 0.5*(AP50IA + AP75IA), APste = 0.5 * (AP50ste + AP75ste)
Final score = 0.2*(0.5*F1IA + 0.5*F1ste) + 0.4*(0.5*APIA + 0.5*APste) + 0.1*(0.5*DSCIA + 0.5*DSCste) + 0.1*(0.5*95HDIA + 0.5*95HDste) + 0.2*(MAEpct/3 + MAEla/3 + MAEsa/3)
Where IA stands for intracranial aneurysm, ste stands for stenosis, pct stands for percentage of stenosis, la stands for long axis length, sa stands for short axis length.

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.

Ranking variability will be characterized using the Bootstrap method.

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

The Bootstrap is a simple nonparametric method that relies on minimal assumptions.

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.

No further analyses are planned as of this time

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.

Annotation tool DOI: 10.1016/j.neuroimage.2006.01.015
3D TOF-MRA, Doi: 10.1038/270722a0

Further comments

Further comments from the organizers.

The challenge will be jointly held by Annual Meeting for Society for Magnetic Resonance Angiography (SMRA) 2024