General

Challenge name*

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

Big **C**ross-**M**odal **A**ttenuation **C**orrection Challenge

Acronym

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

BIC-MAC

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 advent of Long Axial Field-of-View (LAFOV) PET scanners has shifted the dosimetry paradigm in PET/CT imaging. The high sensitivity of these systems allows for substantial reductions in radiotracer activity, rendering the volumetric CT component the dominant source of ionizing radiation. For dose-sensitive populations such as pediatric and obstetric cohorts, eliminating the volumetric CT entirely is highly desirable. However, the CT serves a dual purpose: providing anatomical context and enabling attenuation correction (AC) for PET reconstruction, as the attenuation map is typically derived directly from the CT. Similarly, whole-body studies acquired on PET/MRI systems require estimation of the attenuation map from MR images. In both scenarios, the absence of a CT poses a reconstruction challenge.

To address this, the Big Cross-Modal Attenuation Correction (BIC-MAC) challenge tasks participants with synthesizing a 3D pseudo-CT from other available modalities. We present a novel multimodal dataset comprising whole-body PET, CT, Topogram (scout radiograph), and MRI for 100 healthy volunteers. The cohort is age- and sex-stratified, with data acquired on Siemens Biograph Vision Quadra and Siemens MAGNETOM Vida scanners. Participants will receive a training set of 80 cases containing Non-Attenuation Corrected (NAC) [18F]FDG PET images, scan-planning Topograms, and same-day DIXON MRI, alongside reference CT and CT-based attenuation-corrected PET (CTAC-PET) images. Critically, we also provide scatter maps, sinograms, and Docker containers with open-source reconstruction software, enabling closed-loop optimization on the training set—a capability previously restricted to hospital sites with access to proprietary vendor software.

The challenge comprises a single task: generate a pseudo-CT from the available input modalities. The pseudo-CT will be used to reconstruct PET images, which are then quantitatively compared against reference CTAC-PET images. Both static and dynamic PET reconstructions are evaluated to assess downstream accuracy across different clinical contexts. A defining technical characteristic of this challenge is the integration of modalities with different dimensionalities and acquisition geometries. While the 3D NAC-PET and 2D Topograms are spatially aligned with the target attenuation map, both lack anatomical detail. In contrast, whole-body MRI offers high bone and soft-tissue contrast but is acquired in a different scanner geometry with different patient positioning and body deformations. Consequently, participants must develop algorithms capable of fusing spatially unaligned information from 3D volumetric MRI with that of the 3D NAC-PET and 2D Topograms.

Keywords*

List the primary keywords that characterize the challenge. (Separate your inputs with comma like Keyword 1, Keyword 2)

Attenuation Correction, Cross-modal synthesis, registration, multimodal, PET, CT, MRI, Whole-body, 18F-FDG

Year*

Please indicate the year of the challenge. If you are applying for next year's conference, please write the year of that conference.

2026

Novelty of the challenge*

Briefly describe the novelty of the challenge.

1. CT-less attenuation correction of PET images is a well-established research field with extensive published literature. However, to the best of our knowledge, this is the first attenuation correction challenge using human imaging data.

2. This is the first reconstruction challenge that enables closed-loop algorithm optimization with PET reconstruction integrated into the development pipeline. Previously, such capabilities were restricted to hospital sites with access to proprietary vendor software. This limitation meant that non-hospital researchers could only evaluate performance using CT-based surrogate metrics, which often correlate poorly with downstream PET image accuracy. By providing open-source reconstruction tools, participants can now iterate on their algorithms guided directly by metrics computed on reconstructed PET images.

3. This is the first challenge to incorporate whole-body PET, CT, and MRI from the same subjects. The underlying dataset is novel, acquired on state-of-the-art PET and MRI scanners. No previous public dataset has included whole-body non-attenuation-corrected PET images, whole-body PET sinograms, or whole-body dynamic (4D) PET images.

4. This is the first challenge to evaluate algorithms on dynamic PET reconstructions. These 4D images are used in the final evaluation to quantify how attenuation correction errors propagate to time-activity curve measurements and, by extension, to parametric values derived from kinetic modeling.
5.

Task description and application scenarios

Briefly describe the application scenarios for the tasks in the challenge.

There are two primary clinical application scenarios for CT-less attenuation correction:

1. **Dose reduction in PET/CT studies.** CT-less attenuation correction is

important for limiting ionizing radiation exposure in dose-sensitive populations such as pediatric and obstetric cohorts. The CT effective dose varies with scanner, protocol, and patient characteristics, but is often comparable in magnitude to the effective dose from the injected radiotracer, making it a significant component of total examination radiation.

2. **Attenuation correction for PET/MRI systems.** Combined PET/MRI systems are an increasingly attractive alternative to PET/CT for whole-body investigations. However, in the absence of CT, an attenuation map must be derived from the MRI, the non-attenuation-corrected PET, or a combination of both.

Despite the clinical importance of these applications, two main barriers have limited the validation and widespread adoption of CT-less attenuation correction algorithms: (1) the lack of shared datasets for benchmarking, and (2) the lack of accessible reconstruction software. Research in this field has been predominantly restricted to hospital sites, as attenuation-corrected reconstruction requires access to both raw PET data (listmode or sinograms) and vendor-proprietary reconstruction software. This challenge addresses both barriers by providing PET sinograms alongside a Docker container with the open-source STIR reconstruction package, enabling participants to perform attenuation-corrected reconstruction on the training cases.

Conference

Associated workshops

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

None

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 participation from approximately 30 teams at MICCAI 2026. Since this is the first iteration of the challenge, we based this estimate on participation rates from other first-time MICCAI challenges with similar scope.

CT-less attenuation correction is a well-established research field, and we anticipate that many MICCAI attendees will already be familiar with the topic. Furthermore, while the evaluation scheme specifically targets PET attenuation correction, the underlying task - predicting one modality (CT) from other available modalities (NAC-PET, MRI, and Topogram) - represents a general cross-modal synthesis problem. The technical challenge of fusing information from images acquired in different geometries and coordinate systems is likely to attract researchers working on multimodal learning and image translation, even those without prior experience in attenuation correction.

Duration*

How long does the challenge take? Possible values: half day, full day, 2 hours, etc.

2 Hours

Longer duration explanation

In case you selected half or full day, please explain why you need a long slot for your challenge.

N/A

Publication and future plans*

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

Following the conclusion of the challenge, we will coordinate a summary paper presenting the challenge results and analysis of submitted methods. The top three performing teams will be invited as co-authors. We intend to submit the paper to a high-impact journal in medical imaging or nuclear medicine, such as:

- Medical Image Analysis
- IEEE Transactions on Medical Imaging
- Journal of Nuclear Medicine (JNM)
- European Journal of Nuclear Medicine and Molecular Imaging (EJNMMI)

We anticipate submission in Q4 2026.

MICCAI LNCS proceedings*

Indicate if you want to offer MICCAI Springer LNCS proceedings to the participants. Publishing a proceedings volume is optional and at the discretion of each challenge's organizers. At a minimum, organizers must ensure that a description of each participant's submission is publicly available. Organizers who wish to publish MICCAI Springer LNCS proceedings must adhere to the MICCAI Satellite events publication process

For this first iteration of the challenge, we do not plan to offer LNCS proceedings. Instead, participants are required to prepare a short methodology paper describing the technical approach underlying their submission. This paper must be uploaded to a public repository (e.g., https://arxiv.org/) and included with the final submission email alongside the Docker container.

Space/ hardware requirements*

Please describe the platform used for any online challenge. For on-site challenges, indicate how you plan to provide a fair computing environment. Please list any technical equipment or support needed (e.g., projectors, computers, monitors, loud speakers, microphones).

Participants are expected to train models using their own computational resources and submit a final Docker container to the organizers via email for evaluation on the unseen test data. Training data are available for download from Hugging Face for registered teams that have signed the Data Usage Agreement (DUA). The training data require approximately 50 GB of storage. Submitted Docker containers will be evaluated on a GPU-enabled Linux server at Rigshospitalet with the following specifications:

OS: Linux (Pop!_OS 22.04)
CPU: 2× Intel Xeon Gold 6346 @ 3.10 GHz (32 cores, 64 threads)
GPU_ NVIDIA A40 (46 GB VRAM, single GPU)
GPU Driver: Version 550
CUDA: Version 12.4

Collaboration with European Society of Radiology (ESR)*

In collaboration with European Society of Radiology (ESR), we also announce special clinical interest topics with associated clinicians who can help with the preparation of the proposals; the best 3 challenge proposals on these topics will get the opportunity to present their challenges at the European Congress of Radiology (ECR) 2027 in a special session. If you want to organize a challenge in collaboration with ESR on one of these topics, please reach out to the MICCAI Challenges Team (miccai-challenges-2026@dkfz-heidelberg.de) and we will put you in contact with the corresponding clinician.
Topics list

- Detection and quantification of colorectal liver metastasis on CT
- Quantification of Osteoporosis on CT
- Prediction of pulmonary function based on a single chest CT scan
- Automated RECIST assessment on baseline and follow-up Thorax-Abdomen CT
- AI-based Positive Assessment of Brain Imaging in Multiple Sclerosis (MS)
- AI-based assessment of PET imaging for oncology

- AI-based generation of full radiology report from imaging
- Ultrasound, Doppler, and MRI-Based Multimodal Segmentation and Characterization of Parotid Tumours
- From single to multi-sequence synthetic MRI for MSK imaging

No.

Create your task(s)

Task: A challenge may deal with multiple different tasks for which separate assessment results are provided. For example, a challenge may target the problem of segmentation of human organs in computed tomography (CT) images. It may include several tasks corresponding to the different organs of interest.

Tip: Our recommendation is to create all tasks first and then fill in the necessary information. This way you can complete everything in one step and save time.

Enter task name here

Pseudo-CT Attenuation Correction

Fill task details

Title*

(You can change the title if necessary, otherwise no need to modify)

Pseudo-CT Attenuation Correction

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.

(Same as challenge abstract since the challenge only contains one task)

The advent of Long Axial Field-of-View (LAFOV) PET scanners has shifted the dosimetry paradigm in PET/CT imaging. The high sensitivity of these systems allows for substantial reductions in radiotracer activity, rendering the volumetric CT component the dominant source of ionizing radiation. For dose-sensitive populations such as pediatric and obstetric cohorts, eliminating the volumetric CT entirely is highly desirable. However, the CT serves a dual purpose: providing anatomical context and enabling attenuation correction (AC) for PET reconstruction, as the attenuation map is typically derived directly from the CT. Similarly, whole-body studies acquired on PET/MRI systems require estimation of the attenuation map from MR images. In both scenarios, the absence of a CT poses a reconstruction challenge.

To address this, the Big Cross-Modal Attenuation Correction (BIC-MAC) challenge tasks participants with synthesizing a 3D pseudo-CT from other available modalities. We present a novel multimodal dataset comprising whole-body PET, CT, Topogram (scout radiograph), and MRI for 100 healthy volunteers. The cohort is age- and sex-stratified, with data acquired on Siemens Biograph Vision Quadra and Siemens MAGNETOM Vida scanners. Participants will receive a training set of 80 cases containing Non-Attenuation Corrected (NAC) [18F]FDG PET images, scan-planning Topograms, and same-day DIXON MRI, alongside reference CT and CT-based attenuation-corrected PET (CTAC-PET) images. Critically, we also provide scatter maps, sinograms, and Docker containers with open-source reconstruction software, enabling closed-loop optimization on the training set—a capability previously restricted to hospital sites with access to proprietary vendor software.

The challenge comprises a single task: generate a pseudo-CT from the available input modalities. The pseudo-CT will be used to reconstruct PET images, which are then quantitatively compared against reference CTAC-PET images. Both static and dynamic PET reconstructions are evaluated to assess downstream accuracy across different clinical contexts. A defining technical characteristic of this challenge is the integration of modalities with different dimensionalities and acquisition geometries. While the 3D NAC-PET and 2D Topograms are spatially aligned with the target attenuation map, both lack anatomical detail. In contrast, whole-body MRI offers high bone and soft-tissue contrast but is acquired in a different scanner geometry with different patient positioning and body deformations. Consequently, participants must develop algorithms capable of fusing spatially unaligned information from 3D volumetric MRI with that of the 3D NAC-PET and 2D Topograms.

Keywords

List the primary keywords that characterize the challenge. (Separate your inputs with comma like Keyword 1, Keyword 2)

Attenuation Correction, Cross-modal synthesis, registration, multimodal, PET, CT, MRI, Whole-body, 18F-FDG

Organizing team*

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

Christian Hinge

- Department of Clinical Physiology and Nuclear Medicine, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

Claes Nøhr Ladefoged

- Department of Clinical Physiology and Nuclear Medicine, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
- Department of Mathematical Modelling and Computation, Technical University of Denmark, Kongens Lyngby, Denmark

Flemming Littrup Andersen

- Department of Clinical Physiology and Nuclear Medicine, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
- Department of Clinical Medicine, University of Copenhagen, Copenhagen,

Denmark

Ian Law
- Department of Clinical Physiology and Nuclear Medicine, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

Kirsten Korsholm
- Department of Clinical Physiology and Nuclear Medicine, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

Contact Person*

Provide information on the primary contact person.

Christian Hinge

Clinicians part of the organizing team*

Indicate whether clinicians are part of the organizing team. If yes, describe their role.

Yes. Clinical Professor Ian Law and Kirsten Korsholm are Nuclear Medicine physicians specializing in brain and whole-body PET imaging, respectively. Both have been involved in the data collection for this challenge and have made substantial contributions to the quantitative and qualitative evaluation of CT-less PET attenuation correction algorithms in prior research. Their expertise has directly influenced the design of the evaluation scheme for the BIC-MAC challenge.

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*
- *Repeated event as open call challenge*

One-time event with fixed conference submission deadline.

The life cycle of the BIC-MAC challenge will largely depend on the participant feedback and the participation rates for the 2026 MICCAI challenge. A second iteration has not been planned, but the challenge may remain open for submissions following the conference deadline.

Event

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

MICCAI 2026

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

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

Please note: If you would like to run your challenge on grand-challenge.org, please also fill out their challenge request form as soon as possible. You can upload the PDF from your MICCAI application and then fill most fields in their form with "see PDF". You will also need to provide details regarding your compute and storage requirements. You can find more information about that here. Finally, please also note that the Grand Challenge platform strongly encourages open science and hence requires that you publish your training data with a permissive CC-BY license and that you encourage your participants to publish their source code with an appropriate license as well.

The challenge platform for BIC-MAC will be Codabench: https://www.codabench.org/competitions/12555/
(Note that the Codabench project page has not yet been completed)

Website

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

The challenge website for BIC-MAC is: https://bic-mac-challenge.github.io/
(Note that the challenge website may undergo modifications)

Allowed user interaction*

Define the allowed user interaction of the algorithms assessed. This includes the policy regarding any curation, (pre-)processing and (pre-)training steps. Examples: i) no user interaction is allowed at any step; ii) user interaction is allowed for curating training data (i.e. excluding some training samples).

(You can select multiple values and/or add a custom input)

Fully Interactive Select option(s) or start typing to add a custom input..

Please select at least one option.

Training data policy*

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 may also include publicly available data including (open) pre-trained nets. Clarify whether such additional data needs to be publicly available at the time of the challenge launch. Clarify whether adding (private) annotations of the public data is allowed.

Examples:

- *No policy defined.*
- *No additional data allowed.*
- *Private data is allowed.*
- *Publicly available data is allowed.*
- *Publicly available data is allowed but private annotations of such data is prohibited.*

No additional data allowed

Organizer policy*

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.

Examples:

- *May not participate.*
- *May participate but not eligible for awards and not listed in leaderboard.*

May participate but not eligible for awards and not listed in leaderboard

Award policy*

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

The top three teams will receive monetary awards. We are currently finalizing the agreement with the sponsor, but we expect a total prize pool of $500.

Result announcement policy*

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.*

The scores of all teams will be published on the CodaBench leaderboard. Leaderboard entries will include team name, metric scores on the test set, and overall performance rank. Teams may opt out of inclusion in the summary paper by notifying the organizing committee via email.

The top three performing teams will be announced publicly at MICCAI and on social media platforms at the discretion of the organizers. Social media announcements will include team names and performance metrics only. Announcements at MICCAI will additionally include team members' names and affiliations.

Publication policy*

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).

Up to four members from each of the top three performing teams will be invited as co-authors on the challenge summary paper. Participating teams may publish their own results independently following a three-month embargo period after the conclusion of MICCAI 2026.

Submission method*

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:*
- *Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.*
- *Algorithm container submission (type 2) on Grand Challenge.*

Submissions consist of a Docker container sent to the organizers via email. Detailed submission instructions, container validation procedures, and baseline examples will be made available on CodaBench and the GitHub project repository (https://github.com/bic-mac-challenge/challenge-codebase).

Pre-evaluation*

Provide information on the possibility for participating teams to evaluate their 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.

The challenge comprises only training and test sets; participants may split the training data for validation purposes at their own discretion.

We provide participants with the exact Docker containers used for PET image reconstruction and calculation of metrics 1–3 (see Metrics section). However, containerized solutions may exhibit unexpected behavior across different hardware environments. To address this, we offer pre-evaluation submissions: participants may send their containerized solution via email to the lead organizer, who will run the container on two predetermined training cases using the same hardware as the final evaluation. Results will be returned within 72 hours, allowing teams to verify consistent model behavior on the evaluation hardware.

Due to the computational cost of whole-body PET reconstruction (approximately

30 minutes per case), pre-evaluation submissions are limited to one successful submission per team per month. No validation leaderboard will be maintained during the challenge.

For the final evaluation on the unseen test set, each team is permitted two Docker submissions. Only the results from the most recent submission will be used for ranking.

Schedule*

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

- the release date(s) of the training cases (if any) URL
- 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

The timeline for the BIC-MAC challenge is as follows:

Date	Milestone
March 15	Training data released (n=80 cases) on Hugging Face
April 1	Registration opens via Google Forms
May 15	Pre-evaluation period opens
June 15	Final evaluation period opens
August 15	Registration closes
August 24	Pre-evaluation and final evaluation close
September 1	CodaBench leaderboard updated with test scores; top three winners announced

Note on data availability: All imaging data for both training and test cases have already been acquired. The training cases are currently available in BIDS format on PublicNeuro (https://doi.org/10.70883/GIOX3828); however, this release does not include PET sinograms and contains PET images reconstructed with clinical software. A curated subset including sinograms and STIR-reconstructed PET images will be made available on Hugging Face.

Ethics approval

Indicate whether ethics approvalis 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).

Regional ethics approval was required for original acquisition and sharing of the PET, CT and MRI scans. The project, *H-23065644 - Rask Digital Tvilling: Et*

Syntetisk PET billede til Personlig Medicin, was approved by the Danish regional Ethics Committé September 13, 2023.

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 (click here for more information).

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

The data can be acquired from huggingface upon signing a custom data user agreement (DUA), which is to be announced. The DUA will prohibit sharing the original data.

Code availability of the organizers*

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.

Evaluation code will be made publicly available on GitHub at https://github.com/bic-mac-challenge/challenge-codebase. This code will enable participants to evaluate their algorithms on the training set and understand the containerized evaluation process. The evaluation code can be run on any machine that supports Docker, Python 3.8+, and the nibabel package.

Code availability of the participating teams*

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

Submitted Docker containers and source code from participating teams will not be made public. Teams are encouraged to publish their code on open-source platforms at their own discretion.

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.

Siemens Healthineers has provided Rigshospitalet with a grant supporting the PhD salary of Christian Hinge and participant compensation for the volunteers comprising the training and test cohorts.

Only members of the organizing committee have had access to the test case data. This access was limited to data preparation tasks: converting images to NIfTI format, performing quality assurance, and organizing data according to the BIDS specification. No other individuals will have access to test case data for the duration of the challenge.

Field of application(s)*

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*

(You can select multiple values and/or add a custom input)

Research

Please select at least one option.

Task Category(ies)*

State the task category(ies).

Examples:

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

(You can select multiple values and/or add a custom input)

Classification Select option(s) or start typing to add a custom input...

Please select at least one option.

Target cohort*

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)

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

The target cohort is defined by the two clinical applications for CT-less attenuation correction: dose reduction in PET/CT studies and enabling attenuation correction in PET/MRI studies.

For the first application, the target cohort comprises primarily patients with oncological conditions, with particular emphasis on pediatric and obstetric populations for whom low-dose protocols are especially important. For the second application, the target cohort includes primarily patients with oncological disease and neurological disorders.

Both target cohorts typically include multiple studies per patient, acquired on different scanners as part of routine staging, interim treatment assessment, end-of-treatment assessment, and follow-up imaging. Depending on institutional protocols and patient characteristics, studies may employ either low-dose or standard-dose protocols. The age distribution of the target cohort is skewed toward elderly patients.

Challenge cohort*

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

The challenge cohort comprises healthy, non-pregnant participants stratified equally across sex and four age groups (18–34, 35–49, 50–69, and 70–99). All participants were asymptomatic on the day of scanning, and images were reviewed by clinicians to exclude individuals with clinically significant findings. The challenge cohort includes only one study per participant. All data were acquired on a single LAFOV PET/CT scanner (Siemens Biograph Vision Quadra) and a single MRI scanner (Siemens MAGNETOM Vida).

Imaging technique(s)*

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

The challenge image modalities are: Topogram (2D scout view XRay), lowdose Computed tomography (CT) without contrast, ultra-lowdose (0.4MBq/kg) [18F]FDG Positron Emission Tomography (PET), and VIBE DIXON T1w Magnetic Resonance Imaging (MRI).

Context information: Image data*

Provide additional information given along with the images. The information may correspond to directly to the imaging data (e.g. tumor volume).

The decay-corrected dose of injected [18F]FDG is provided to enable SUV normalization of the AC-PET and NAC-PET images.

Context information: Patient*

Provide additional information given along with the images. The information may correspond to the patient in general (e.g. gender, medical history).

The following information is provided for each participant: age (years), sex, height (cm), and weight (kg). Note that all participants are healthy controls.

Data origin*

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.

PET/CT imaging data were acquired using a standard clinical protocol covering from the vertex to 106 cm inferiorly (approximately mid-thigh). MRI data were similarly acquired from the vertex to mid-thigh. In both cases, participants were positioned head-first supine with arms at their sides.

The data origin is the same for the target and challenge cohorts, with two exceptions:

1. Patients in the target cohort are sometimes scanned with one or both arms raised.
2. Patients in the target cohort are sometimes scanned from the base of the neck, resulting in a field of view that excludes the brain.

Algorithm Target*

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 for both the challenge and target cohorts is the attenuation-corrected PET image. In practice, this is achieved by predicting a pseudo-CT, which is then used to generate an attenuation-corrected PET image (pseudo-CTAC PET). For the challenge, PET images are reconstructed using STIR to enable closed-loop optimization. For the target cohort, however, the ultimate algorithm target is the PET image reconstructed using clinical vendor software. We hypothesize that algorithms achieving high attenuation correction accuracy on STIR-reconstructed PET images will generalize to clinically reconstructed PET 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, 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:
***Accuracy**, Applicability, Complexity, Consistency, Ergonomics, Feasibility, Hardware requirements, Interaction, Integration in workflow, **Precision**, Reliability, **Robustness**, Runtime, Sensitivity, Specificity, Usability, User satisfaction*

(You can select multiple values and/or add a custom input)

Ergonomics Select option(s) or start typing to add a custom input...

Please select at least one option.

Data acquisition device(s)*

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).

PET, CT, and Topogram images were acquired on a single long-axial-field-of-view Siemens Biograph Vision Quadra scanner. MR images were acquired on a single Siemens MAGNETOM Vida scanner. Images were reviewed by clinicians using the Siemens syngo.via platform. Metadata, including injected dose and patient weight, were recorded using standard clinical equipment.

Data acquisition details*

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

Acquisition details for CT, Topogram, PET, and MRI are provided below.

CT, PET, and Topogram

The imaging protocol followed EANM guidelines for tumor PET/CT imaging with [18F]FDG, using a vertex-to-mid-femur field of view. Participants were positioned head-first supine with arms at their sides and the head immobilized using pillows and Velcro bands across the forehead. For the final 58 participants, a soft blanket was placed beneath the participant to reduce pressure-induced ischemic artifacts at the sacrum, scapulae, and occiput. The protocol commenced with a 2D Topogram for scan range planning, followed by a low-dose CT without contrast. Acquisition parameters included a pitch factor of 1.2, revolution time of 0.5 s, reference mAs of 160, and tube voltage of 120 kVp, corresponding to an effective dose of approximately 4.5 mSv. Attenuation correction CT images were reconstructed with a reconstruction diameter of 780 mm, matrix size of 512 × 512 × 631, using the ADMIRE iterative reconstruction algorithm at strength 3.

Following CT acquisition, participants received an ultra-low-dose intravenous injection of [18F]FDG (0.4 MBq/kg; range 18–51 MBq), corresponding to an effective dose of approximately 0.5 mSv. PET acquisition was performed for 70 minutes, during which participants were instructed to remain still.

MRI

MRI scanning was performed on the same day as PET/CT acquisition for 96 participants. The remaining four participants were scanned 9, 14, 58, and 89 days after PET/CT due to scheduling conflicts or failed initial acquisition. Participants were positioned head-first supine in a 3-Tesla Siemens MAGNETOM Vida scanner (software version XA20) using three body coils and one head coil. Four breath-hold T1-weighted VIBE DIXON sequences (25 s each) were acquired, spanning from the vertex to mid-femur. The dataset includes both the individual sequences and the stitched whole-body volume. Each DIXON acquisition produced two volumes: in-phase and opposed-phase. An additional 30-second DIXON sequence was acquired covering the head.

Center(s)/institute(s)*

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

All data were acquired at the Department of Clinical Physiology and Nuclear

Medicine and the Department of Radiotherapy, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.

Characteristics of the subjects*

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).

PET/CT images were acquired and assessed by the clinical personnel of the department. MR images were acquired by student workers.

Case definition*

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) and may include context information. Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.*

Cases are almost identical in nature for the training and test cohorts. Each case comprises one PET/CT study and one MRI study from the same participant. The CT, CT-derived segmentations, and CT-attenuation-corrected PET images are considered labels; all other modalities (NAC-PET, MRI, Topogram) and study/patient metadata are considered algorithm inputs. Each case corresponds to a unique participant.

Input data

Metadata:

- Decay-corrected injected [18F]FDG activity (MBq)
- Subject age (years)
- Subject sex
- Subject height (cm)

NIfTI images:

- Topogram (2D, resampled in the coronal plane to match PET geometry)
- NAC-PET static, 50–70 min post-injection (dimensions: 440 × 440 × 531)
- Head MRI (variable dimensions)
- Whole-body MRI (variable dimensions). Note: this image is acquired from four sequential breath-hold scans; both the scanner-stitched volume and the individual station volumes are provided.

Label data

NIfTI images:

- CT (resampled to PET geometry)
- Organ segmentations (resampled to PET geometry)
- CTAC-PET static, 50–70 min post-injection (dimensions: 440 × 440 × 531)
- CTAC-PET dynamic, 0–32 s post-injection (dimensions: 8 × 440 × 440 × 531). Note: not provided for training cases.

Source data:

- PET scatter maps
- PET sinograms, static 50–70 min
- PET sinograms, dynamic 0–32 s. Note: not provided for training cases.

For all MRI acquisitions, both in-phase and opposed-phase NIfTI volumes are provided.

Number of cases*

State individually total number of training, validation and test cases.

The training and test sets comprise 80 and 20 cases, respectively. No dedicated validation set is provided; teams may partition the training data for validation at their own discretion.

Quantity of data which is already annotated*

How much of the data are already annotated (stratified by train test in percentage)?

All raw PET, CT, and MRI data have been collected. CT images, MRI images, segmentations, and metadata have been processed for all 100 cases. The ground-truth STIR-reconstructed PET images are currently being generated and will be available prior to the training data release.

Explanation of data proportion*

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

The "Big" Cross-Modal Attenuation Correction challenge contains a total of 100 cases, which for many deep learning challenges would be considered a small dataset. For CT-less attenuation correction, however, this cohort size is considered large and is consistent with cohort sizes used to train state-of-the-art methods. Pseudo-CT models are typically trained using image-to-image translation architectures, such as U-Nets, with the CT image as the reference label. Such volumetric labels provide a substantially stronger and denser supervision signal compared to scalar labels used in classification and regression tasks. This has historically enabled 3D translation models to be trained on smaller datasets without compromising generalizability.

A test set of 20 cases was chosen to provide a statistically robust estimate of performance, consistent with previous publications, while limiting computational burden. Specifically, nine PET frame reconstructions are performed per test case per submission (one for metrics 1–3 and eight for metric 4). Each STIR reconstruction takes approximately 30 minutes, and with two permitted submissions per team, this amounts to 180 hours of reconstruction time per team.

Further important characteristics of the cases*

Challenge organizers are encouraged to (partly) use unseen, unpublished data for their challenges. Describe if new data will be used for the challenge and state the number of cases along with the proportion of new data.

All data were acquired between April 8, 2024, and April 1, 2025. All cases are previously unpublished and have not been used in prior challenges.

Further important characteristics of the cases*

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 100 cases are equally distributed across sex (male/female) and four age groups (18–34, 35–49, 50–69, and 70–99), yielding eight demographic strata. This balanced distribution was chosen to ensure adequate representation across diverse demographics given the limited dataset size. We note that the dataset does not include pediatric or pregnant participants. Test cases were randomly sampled from the full cohort, with the constraint that the training set contains exactly 10 cases from each of the eight demographic groups.

Method for determining the reference annotation*

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

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

The reference annotation is the STIR-reconstructed PET image with CT-based attenuation correction. No human annotators were involved; the reference is generated algorithmically.

Instructions given to the annotators*

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.

Not applicable. Reference annotations are generated algorithmically rather than through manual annotation. During image acquisition, participants were instructed to remain still throughout and between Topogram, CT, and PET acquisitions.

Details on the subject(s)/algorithm(s) that annotated the cases*

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.

Reference PET images will be reconstructed using STIR with Ordered Subset Expectation Maximization (OSEM), configured with 4 iterations and 5 subsets.

Method(s) used to merge multiple annotations

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.

Identical pre-processing methods will be applied to training and test cases.

PET source data:

1. PET listmode (raw emission) files were converted to static and dynamic sinograms using Siemens JSRecon.
2. Scatter maps were generated from the sinograms and CT images using Siemens JSRecon.
3. Dynamic and static CTAC-PET and static NAC-PET images were reconstructed using STIR.

Image data:

1. PET, CT, and MRI images were converted to NIfTI format using dcm2niix and renamed according to the BIDS naming convention.
2. TotalSegmentator was applied to CT images to generate segmentations required for metric evaluation. The TotalSegmentator tasks used were "total," "body_parts," and "brain_regions."
3. CT images and TotalSegmentator segmentations were resampled to match PET image geometry. This accommodates the STIR reconstruction software, which requires CT images at the same resolution as PET images. Topograms were similarly resampled to the PET geometry in the coronal plane. Brain and whole-body MRI images were acquired in a different geometry; preprocessing of MRI data is left to the discretion of individual teams.

Sources of error related to the image annotation*

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.

Not applicable. Reference annotations are generated algorithmically without human annotation.

Other sources of error*

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

CT-based attenuation correction is the clinical standard for virtually all PET/CT studies; however, it represents a silver standard rather than ground truth. CT images provide an accurate snapshot of tissue attenuation at a single time point, whereas PET images are reconstructed from emissions acquired over several minutes during which respiratory and cardiac motion occurs. Depending on the respiratory phase captured by the CT, characteristic "banana" artifacts may appear, particularly in the liver and lungs. Minor patient motion (e.g., hand or finger movement) may also occur between PET and CT acquisitions. In practice, the true attenuation map cannot be directly observed. We argue that the evaluation metrics—which compare CTAC-PET with pseudo-CTAC-PET—are justified by the fact that CT-based attenuation correction is the clinically accepted standard.

Metric(s)*

Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) . 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)*

Four metrics are used to assess per-case algorithm performance. Although submitted algorithms produce pseudo-CTs, all metrics quantify similarity between pseudo-CTAC-PET and CTAC-PET images.

The first three metrics are computed on static PET images reconstructed from a 50–70 minute post-injection window, corresponding to the CTAC-PET images and sinograms provided for the training cohort.

The fourth metric, TAC bias, differs in that both the reference CTAC-PET and predicted pseudo-CTAC-PET are dynamic images containing eight 4-second frames, collectively covering the first 32 seconds of PET acquisition. These images are reconstructed from the same 70-minute acquisitions used for static reconstructions but with different time framing. Neither dynamic sinograms nor dynamic reconstructed PET images are provided for training cases. This is intentional: the TAC bias metric quantifies performance on a reconstruction target that has not been subject to closed-loop optimization.

1. **Whole-body SUV MAE:** Mean voxel-wise absolute error of the SUV-normalized PET image within the body. SUV normalization, which scales PET intensity by patient weight and injected dose, standardizes the MAE magnitude across subjects. To reduce respiratory motion artifacts, axial slices within 4 cm above and below the superior point of the liver are excluded. Body and liver segmentations are defined using the TotalSegmentator "total" and "body_parts" tasks, respectively.
2. **Head and neck SUV MAE:** Identical to whole-body SUV MAE but restricted to the head and neck region. The mask is defined using the TotalSegmentator "body_parts" task.
3. **Organ bias:** Volume-weighted mean absolute relative error of SUVmean measurements across all 117 organs defined by the TotalSegmentator "total" task. Organs with volume smaller than 5 mL are excluded to reduce the influence of noise on relative error measurements.
4. **TAC bias:** Mean absolute relative error (MARE) of the time-integrated PET signal within selected regions of interest. Time-activity curves (TACs) are computed by averaging PET activity within the aorta and selected brain regions for each time frame, producing a per-region 8-element TAC. These TACs are integrated to yield a single area-under-the-curve value per region. The MARE of these values is averaged across regions to produce a single per-case TAC bias. Aorta and brain segmentations are derived using the TotalSegmentator "total" and "brain_structures" tasks, respectively.

Justification of metric(s)*

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

We justify each of the four metrics below.

1. Whole-body SUV MAE

[18F]FDG PET images are predominantly evaluated qualitatively, with displayed intensities scaled linearly by SUV. Visual interpretation typically includes assessment of whether uptake in a volume of interest (e.g., a tumor) has increased or decreased relative to a prior scan, or whether uptake is higher or lower than in a reference region such as the liver or aorta. For such intensity-based assessment, MAE is preferable to metrics emphasizing texture or perceptual similarity.

2. Head and neck SUV MAE

The head and neck contain complex bony structures that are difficult to predict from NAC-PET images alone and can induce significant attenuation correction biases in the brain and surrounding tissues if inaccurately estimated. Because the head and neck constitute a small fraction of total body volume, this region has limited influence on metric 1. We therefore include a dedicated regional metric to emphasize its importance. Additionally, the majority of published CT-less attenuation correction literature focuses on the head and neck region.

3. Organ bias

Organ bias is a well-established metric in the CT-less attenuation correction literature. In clinical practice, low organ bias supports the validity of quantitative assessments such as those defined by PERCIST 1.0, which supplement visual interpretation. Such guidelines compute relative changes in mean SUV measured within disease and reference organ volumes of interest between sequential scans—for instance, before and after treatment.

4. TAC bias

Dynamic PET images enable kinetic analyses, such as Patlak regression, to derive parametric values for regions or voxels. Relative errors in TAC integrals typically propagate to relative errors in the resulting parametric values; therefore, minimizing TAC integral bias is essential. The aortic TAC is particularly important as it commonly serves as the input function for kinetic modeling in other organs. We focus on brain regions because the brain rapidly accumulates glucose, providing a measurable signal within the 32-second time window. Typical [18F]FDG kinetic analyses employ a larger number of time frames of varying duration; however, this substantially increases reconstruction time.

Method used to compute a performance rank*

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. Ideally, provide the ranking scheme as a concrete pseudo code.

The performance rank for each submission is computed as follows. First, each metric is averaged across all test cases to yield a mean score per metric. Submissions are then assigned a rank position for each metric (1 = best, 2 = second best, ..., N = worst). The final aggregated rank is computed as the mean of the four metric ranks. The submission with the lowest final score is declared the winner.

```
FOR each metric:
    COMPUTE mean score across all test cases for each submission
    RANK submissions from 1 (best) to N (worst) based on mean score

FOR each submission:
    COMPUTE final_score = MEAN of the 4 metric ranks

WINNER = submission with lowest final_score
```

Submissions with missing results*

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

If a submission fails on some or all test cases - due to technical errors or exceeding the 6-minute per-case time limit - the team will be notified, and the failed predictions will be replaced with outputs from the baseline water model defined in the GitHub repository. The baseline model assumes uniform attenuation corresponding to water (HU = 0) within the body and air (HU = −1024) outside the body; the body mask is defined by thresholding the NAC-PET image. Each team may submit twice, with only the most recent submission used for final evaluation.

Justification of ranking*

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

The rank-based aggregation method ensures that the magnitude and variance of each metric do not influence the combined performance rank. All four metrics directly evaluate the quality of the pseudo-CTAC-PET image against the clinical silver standard (CTAC-PET) but from different clinical perspectives, as described in the "Justification of metric(s)" section. Note that the head and brain regions are evaluated in all four metrics and are thus overrepresented in the aggregated performance rank relative to other anatomical regions. This is intentional: the head, neck, and brain contain complex bony anatomy, making these regions particularly susceptible to attenuation correction errors.

Statistics - Overview*

Provide an overview of the statistical approaches used in the scope of the challenge analysis. Details can be provided in the parameters below. For each parameter, justify why the described statistical method(s) was/were used and, if necessary, add a description of any method used to assess whether the data met the assumptions required for the particular statistical approach.

The statistical analysis employs a rank-sum method to aggregate four distinct error metrics (Whole-body SUV MAE, Head and Neck SUV MAE, Organ Bias, and TAC Bias) into a single final ranking. We use non-parametric approaches throughout: bootstrapping for confidence intervals and ranking stability assessment, and the Wilcoxon signed-rank test for significance testing.

Statistics - Precision of the performance estimates

Provide a description of how the precision of the performance estimates of individual algorithms is assessed (e.g. confidence interval of the mean on the test set computed using percentile bootstrap, confidence interval of the accuracy on the test set computed using percentile bootstrap).

Each metric measurement is derived by averaging errors over large 3D volumes containing millions of voxels (approximately 15–70 L of tissue). This extensive volumetric averaging inherently yields high-precision estimates for each case, rendering voxel-wise bootstrapping unnecessary. To assess the precision of the final aggregate score (the mean across the test set), we calculate 95% confidence intervals using percentile bootstrapping, resampling the 20 subject-level scores with replacement (1,000 replicates) to define uncertainty bounds for the mean performance on each metric.

Statistics - Performance variability across cases

Provide a description of how variability of the performance of individual algorithms across tests cases is assessed (e.g. SD across test cases, IQR, graphs, reporting outliers…).

Performance variability is analyzed to ensure algorithms are robust across the biological diversity of the challenge cohort (e.g., varying BMI and anatomy). Variability is quantified using the standard deviation (SD) and interquartile range (IQR) for each metric. Boxplots are generated for each metric to visualize distributions, identify skewness, and detect outliers.

Statistics - Rankings variability

Provide a description of how variability of rankings is assessed.

To assess the robustness of the final ranking, we employ bootstrap analysis. We generate 1,000 bootstrap samples of the test set (resampling the 20 cases with replacement) and recalculate the final rank-sum winner for each sample. We report the frequency (percentage) with which each method achieves 1st, 2nd, and 3rd rank positions to demonstrate that the winning method is not determined by a single outlier case.

Statistics - Tests for significance

Provide a description of statistical tests that are used to assess whether the differences in performance between algorithms are statistically significant.

Statistical significance of performance differences between top-performing algorithms will be assessed using the Wilcoxon signed-rank test (paired samples). The test will be applied to the 20 subject-level scores for each of the four metrics individually. This paired testing strategy evaluates whether the winning algorithm achieves consistently superior performance across the test cohort relative to the runner-up. A significance level of α = 0.05 will be used, with Bonferroni–Holm correction applied for multiple comparisons across the four metrics.

Statistics - Missing data handling

Provide a description of the missing data handling

No missing data are expected, as all imaging data have been collected and verified. Should missing data occur in the training or test cohorts, the affected cases will be excluded from analysis.

Statistics - Software

Indicate any software product that is used for all data analysis methods.

Statistical analyses will be performed in Python using NumPy and SciPy. Visualization will be performed using Matplotlib and Seaborn.

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

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.

Data DOIs:

- Static images: https://doi.org/10.70883/GIOX3828
- Dynamic images: https://doi.org/10.70883/UYAG3430
- Source data: https://doi.org/10.70883/JZJH3431

Further comments

Do you have any further comments that may be important for the challenge chairs to know?

A manuscript describing the dataset is currently under review at Nature Scientific Data and can be accessed here: https://drive.google.com/file/d/1ifXPrc1Hy3-5wbfw4Qz6Uy4MPndbsxJ9/view?usp=sharing . Please also see the challenge website at https://bic-mac-challenge.github.io/