

Mitosis Domain Generalization Challenge 2025:

Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Mitosis Domain Generalization Challenge 2025

Challenge acronym

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

MIDOG 2025

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.

Generalization to previously unseen real-world data is the most important property of any machine learning algorithms. Yet, it is also one of the most challenging ones, as the data distribution used during training of, especially, deep learning algorithms, is commonly not representative of the real world use case. Domain generalization approaches try to mitigate this by applying tailored augmentation, sampling, or regularization techniques that encourage the model to learn more robust and transferable features.

One area, where domain generalization is a particular issue, is the field of digital pathology, where the digital image is influenced by a significant number of parameters influencing the tissue preprocessing, staining procedure, and digitization. Moreover, the biological variability in tissue is considerable, leading to strong data variance depending on the tissue or tumor type, but also on the question which part of tissue is actually annotated and presented to the algorithms.

The MIDOG 2025 challenge is providing yet another step towards the assessment of generalization for one of the most commonly tackled tasks in tumor pathology: Mitosis identification. The process of cell division (mitosis) is indicative of tumor growth and proliferation, and as such is part of the manual assessment of tumor malignancy by experts for many tumor types.

Following up on the successful MIDOG 2021 and 2022 MICCAI challenges, we propose a third iteration of our challenge, focusing even more strongly on data variation that algorithms working in clinical environments must cope with. Previous challenges were limited towards pre-selected hot-spot regions, leaving the question of generalization to whole slide images open. This year, aiming to generalize to whole slide image applicability, the algorithms have to additionally show their robustness on a wider range of tissues, including non-hotspot regions or even non-tumor areas as well as inflammatory or necrotic areas, which occur in tumor samples regularly.

Besides the main task of mitosis detection, another task of high scientific interest is covered in the challenge: mitotic figures with atypical morphologies have been shown to be an independent prognosticator of tumor malignancy. We therefore include the classification of normal vs. atypical mitotic figures as a second challenge track in the MIDOG 2025 challenge.

Challenge keywords

List the primary keywords that characterize the challenge.

domain generalization, mitosis detection, atypical mitosis

Year

2025

Novelty of the challenge

Briefly describe the novelty of the challenge.

The MIDOG 2025 challenge comprises two major novelties over its predecessors:

1) Generalization to whole slide images.

The previous 2022 and 2021 MIDOG challenges exclusively focused on the detection of hand-selected regions of interest (ROIs) from the whole slide image. The real-world application, however, also includes areas (e.g., necrosis, inflammation) that include cells that have similar appearance than mitotic figures but are not true mitotic figures, leading to a significant increase in false positives if models fail to distinguish between them. This introduces a critical challenge for the practical deployment of such models, as their utility depends on achieving high specificity while maintaining sensitivity across the entire slide.

Given that an applicability on the test set on 120 WSIs using docker containers is not feasible in the context of the challenge for budgetary reasons, we decided to include two groups of additional crops from our tumor cases in an effort to assess the generalization:

- a) random crops containing tissue
- b) challenging regions, such as inflammation or necrosis that contain hard negative mitotic figure candidate cells which can easily yield in false positives.

2) Identification of atypical mitotic figures

Atypical mitotic figures (AMFs) have been recently identified as being an independent prognosticator for some tumors, including breast cancer and mast cell tumor (10.1038/s41379-022-01080-0, 10.1002/jso.25152, 10.3390/vetsci11010005). Yet, the problem of automatic identification of AMF is still in its infancy, partially also due to the non-availability of datasets for this problem. We will make the first large-scale dataset for atypical mitotic figure detection and classification available in the scope of our challenge.

Besides being prognostically relevant, this problem is also interesting from a pattern recognition perspective: AMFs, reflecting defects in the cell division process, have a huge intra-class variability, and are additionally

comparatively rare. We thus expect the solutions provided by the challenge's participants to also provide important insights into similar, yet insufficiently solved, problems.

Task description and application scenarios

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

1. "WSI-ready" algorithms: By providing new areas from tissue sections outside of mitotic ROIs, we hope that the results will reflect how the submitted solutions would perform if they were to be applied on whole slide images, thus incentivizing solutions that could be applied immediately on real world data.
2. Diagnostic applications beyond mitotic index (MI): The inclusion of atypical mitotic figures (AMF) opens up new pathways for large-scale data analysis to further investigate their prognostic significance.

FURTHER INFORMATION FOR CONFERENCE ORGANIZERS

Workshop

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

N/A

Duration

How long does the challenge take?

Half day

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

Given that the identification of atypical mitoses is a relatively new and evolving topic, the workshop aims to facilitate meaningful discussions among participants, fostering an exchange of ideas and perspectives. Additionally, a half-day format allows us to include a keynote or invited talk by a leading expert in pathology, ensuring a comprehensive introduction to the subject and setting the stage for the subsequent presentations and discussions.

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.

Based on our experience with previous challenges and the steady interest in the topic, we expect 15-30 participants.

Publication and future plans

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

As with the 2021 and 2022 challenges, we plan to publish our challenge report in Elsevier's Medical Image Analysis. We also plan for the participants to offer challenge proceedings, as in previous years.

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 plan to conduct the challenge again using grand-challenge.org as platform. No special equipment or support needed.

TASK 1: Mitotic Figure Detection under Real-World Conditions

SUMMARY

Abstract

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

The participants shall detect mitotic figure objects from H&E-stained; histopathology images, including regions of interest (preselected by a pathologist), but also conditions that are suboptimal for detection, such as necrotic or inflammatory areas or non-tumor areas in the slide.

Keywords

List the primary keywords that characterize the task.

mitosis detection

ORGANIZATION

Organizers

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

Jonas Ammeling, Marc Aubreville, Sweta Banerjee, Christof A. Bertram, Katharina Breininger, Dominik Hirling, Peter Horvath, Nikolas Stathonikos, Mitko Veta

b) Provide information on the primary contact person.

Marc Aubreville

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 2025

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

grand-challenge.org

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

midog2025.deepmicroscopy.org

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.

Publicly available data is allowed

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

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

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

We are actively engaging with potential sponsors to secure monetary prizes, similar to those awarded in MIDOG 2021 and 2022. Prizes will be awarded to the top three positions for this task.

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.

Results will be announced during the challenge workshop (i.e., after the submission deadline)

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).

We aim to publish a summary of the challenge in a peer-reviewed journal (e.g. Medical Image Analysis), as in previous iterations.

Participants are requested to publish a description of their method and results in a preprint as a short paper (2-4 pages) together with their submission. At most two authors of that paper will qualify as authors in the summary paper. Participating teams are free to publish their own results in a separate publication. We aim to publish accepted short papers alongside with invited full papers in the challenge proceedings.

Participants may also publish papers including their official performance on the challenge data set, given proper reference of the challenge. There is no embargo time.

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.

Docker container is to be submitted towards grand-challenge. A template docker container with instructions is available through github: https://github.com/DeepPathology/MIDOG_reference_docker

We will provide guidance on how to use dockerized submission in our challenge to the participants.

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.

As in MIDOG2021/2022, we will provide the possibility to test on a preliminary test set, approximately 2 weeks before the submission deadline. Participants are allowed 1 docker container submission per day to be evaluated on this preliminary set.

Attempts to bypass this limitation might lead to exclusion from the challenge. After preliminary evaluation phase, participants have to submit a final container to be applied on the final test set.

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
- March 15, 2025: Go-Live of the challenge, registration open for participants
 - April 1st, 2025: Availability of training data and dataset description
 - August 14, 2025: Deadline for registration of participants
 - August 15, 2025: Availability of preliminary test set
 - August 31, 2025: Deadline for docker container submission and for two-page preprint abstract submission
 - Sept 13-Sept 27, 2025: Announcement of results at MICCAI 2025

Ethics approval

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

We got ethics approval by the UMC Utrecht (TCBio 20-776), Regional and Institutional Research Ethics Committee of the University of Szeged (SINGLE-MITO-01) and the ethics board of the medical faculty of FAU

Erlangen-Nürnberg (AZ 92_14B, AZ 193_18B). For samples taken from the diagnostic archive in veterinary pathology, no ethics approval is required.

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 (Attribution)

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.

Access to the evaluation docker container is available on github:

https://github.com/DeepPathology/MIDOG_evaluation_docker

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

We will provide links to the docker containers the participants submitted, given their consent. Participants will be encouraged to give permission for this and make their code publicly available.

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 organizers declare no conflict of interest. Access to the test case labels will be available to the organizers only, and on a need-to-know basis to perform the evaluation.

MISSION OF THE CHALLENGE

Field(s) of application

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

Examples:

- Diagnosis
- Education

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

Research, CAD, Decision support, Prognosis

Task category(ies)

State the task category(ies)

Examples:

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

Detection, Classification

Cohorts

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

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

The target cohort consists of patients with confirmed cancer and with biopsies taken for histopathologic assessment.

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

The challenge cohort consists of 520 cases of cancer, with a tissue area of 2mm² each, collected at the UMC Utrecht, the Universitätsklinikum Erlangen (University Hospital Erlangen), the HUN-REN Biological Research Centre Szeged and the University of Veterinary Medicine Vienna.

Imaging modality(ies)

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

Histopathology images, hematoxylin and eosin-stained, whole slide image acquisition.

Context information

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

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

We do not provide additional information along with the image set.

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

As this is not relevant for the task, we do not provide meta data about the patient.

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.

The data consists of image patches, selected from whole slide image scans of human and animal cancer tissue, acquired at six different institutions.

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

The algorithm target is the detection of mitotic figures, i.e. cells undergoing cell division. The algorithm shall generalize to 12 different tumor types, which are not disclosed to the participants, out of which the majority is not covered in the training set.

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)).

Robustness, Sensitivity, Specificity

DATA SETS

Data source(s)

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

At UMC Utrecht: Hamamatsu NanoZoomer XR

At VetMedUni Vienna: 3DHistech Panoramic Scan

At UK Erlangen: Hamamatsu Nanoszoomer S60

At FU Berlin: Aperio ScanScope CS2

At HS Flensburg: Aperio ScanScope CS

At HUN-REN Biological Research Centre: 3DHistech P1000

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

Hamamatsu XR: resolution: 0.23 microns / px at 40x

3dHistech Panoramic Scan: 0.25 microns / px at 40x

Nanoszoomer S60: resolution: 0.22 microns / px at 40x

Aperio ScanScope CS / CS2: 0.25 microns / px at 40x

3DHistech P1000: 0.25 microns / px at 40x

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.

- Histopathology archive of the UMC Utrecht
- Neuropathology archive of the Universitätsklinikum Erlangen
- Diagnostic archive of the Institute of Pathology at VetMedUni Vienna
- HUN-REN Biological Research Centre, Szeged

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

The regular diagnostic workflow for H&E-stained; histopathology images at the respective institutions were applied. We expect a dependency on the staining protocols of the institutions, but imply to capture real-world variability in our data set.

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 are given as a section of a whole slide microscopy images. Cases represent individual patients. We ensure no patient is part of training and test set.

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

Training set: 503 cases, previously released as MIDOG++ data set (10.1038/s41597-023-02327-4). Additionally, we provided two WSI datasets fully annotated for mitosis (10.1038/s41597-019-0290-4, 10.1038/s41597-020-00756-z). Participants may also use all publicly available additional training data.

The test cases consist of 12 different tumor types from human and veterinary pathology (dogs/cats), including lung adenocarcinoma, glioblastoma, round cell tumors, epithelial tumors, and mesenchymal tumors. Each tumor type is represented with 10 cases. In total, we thus have 120 cases for the test set, i.e. 120 WSIs from which we source the regions of interest, as specified below. Over the MIDOG 2022 test set, we include two more tumor types (each 10 cases) from a new lab. The tumor types of the MIDOG 2022 test set were disclosed in the MIDOG 2022 challenge report paper. The new tumor types are not disclosed.

We include regions of different difficulty, each sized 2mm², into the test set:

- Ideal regions: 120 ROIs, used in typical tumor grading (100 cases from the MIDOG 2022 test set (already annotated) + 20 new cases)
- Random regions: 120 ROIs randomly selected from the tissue-containing regions of the WSIs used for the ideal regions
- Difficult regions: At least 60 ROIs from difficult regions, such as inflammation and necrosis, which contain cells of high similarity to mitosis, which are candidates for false positives. We try to source a difficult region from each WSI but we assume that not every WSI has such a difficult region.

A preliminary test set (representing four tumor types not contained in the test set) will be made available per docker submission on grand-challenge to check for algorithm function. The aim of this dataset is to check the proper function of the algorithmic approach while not disclosing the characteristics of the test set. The preliminary test set also contains random selections and challenging regions to mimic the overall algorithm target.

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

The challenge dataset represents a good trade-off between capturing the naturally occurring variability of whole slide histopathology images and time invested for the annotation. The selected proportion of training, validation (preliminary test) and test cases allows for a realistic estimation of the performance. At the same time, it allows to test how robust the methods are with respect to domain shift across cancer types.

Additionally, it represents the largest cross-domain annotated mitotic figure data set.

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.

Participants can not expect statistical equality of the characteristics of the training and test sets, since the tissue characteristics of the training set and the test set are different.

e) 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.

We will partially reuse data previously used in the MIDOG 2022 challenge as part of the test set. This dataset was never made available publicly and no external party has had access to it, yet, i.e., it is still fully private.

We will also extend our previous test set with two additional tumor domains.

Additionally, we will annotate various new regions of interest from "adversarial regions" (inflammatory tissue, necrotic regions, crush artifacts) to simulate real-world conditions for algorithms.

Annotation characteristics

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

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

We will use a previously well-established approach for high quality mitotic figure annotation, involving a majority vote of three experts. We made the complete process available as part of the MIDOG++ paper [1].

For the newly annotated regions within the test set, we will additionally utilize slides stained against phospho-histone H3, a specific stain for mitosis. This annotation, while incredibly useful for identifying the biological truth of which cells depict mitosis, is introducing biases and label noise into the annotation [2]. Hence, we will utilize a process described in [2] to debias the mitotic figures in a secondary label cleaning step, making sure that they truly reflect discriminable mitotic figure objects in the H&E; stain.

[1] [10.1038/s41597-023-02327-4](https://doi.org/10.1038/s41597-023-02327-4)

[2] [10.1038/s41598-024-77244-6](https://doi.org/10.1038/s41598-024-77244-6)

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.

We will provide recent guidelines [3] towards mitotic figure identification to all annotators. Besides this, it is worth noting that all annotators have years of experience in identifying mitotic figures in a diagnostic and research setting.

[3] <https://doi.org/10.1177/0300985820980049>

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.

All subjects involved with the annotation process are professional pathologists or biologists with years of experience in identification of mitotic figures.

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.

See above. We define annotations for mitotic figures as belonging to the same cell if their centers are within 7.5 microns (approx. 30 px) of each other. This is approximately the size of a mitotic figure.

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.

We defined a region of interest from each whole slide image and then extracted TIFF images containing only the size of 2mm², as would be used in a typical diagnostic workflow. We performed no additional preprocessing.

Sources of error

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

The following errors are plausible to occur:

- Missed annotation: Especially for atypical mitotic figures, the machine-learning aided setup could fail to provide a remedy for this. In previous work, we estimated the error of potentially missed mitotic figures to be below 5 percent, due to the highly diligent and algorithmically aided annotation process.
- Classification error: If both pathologists independently and erroneously assign the wrong label (either mitotic figure or non-mitotic figure), it is taken as a consensus label. We previously used a clustering approach [4] to allow for re-assessment of potentially misassigned labels. The potential error was identified to be in the area of < 3% [3].
- Bias in the perception of mitotic figures: The experts will likely have a subjective bias as to what to account as a mitotic figure and what not. However, the three-expert consensus should mitigate this issue to a degree. Wilm et al. have shown that involving more than three experts will likely not increase the overall annotation quality significantly [5].

In the test set, we specifically use the a designated workflow to limit subjectivity/bias using PHH3-IHC stain-guided annotation (see [2])

[2] <https://doi.org/10.1038/s41598-024-77244-6>

[4] <https://www.nature.com/articles/s41597-019-0290-4>

[5] https://doi.org/10.1007/978-3-658-33198-6_56

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

Besides annotation errors we do not expect other sources of error.

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)

Matching:

A detected object is considered to be a true positive (TP) if the Euclidean distance between its center and a ground truth location is less than 7.5 microns (approx. 30 pixels, depending on the scanner resolution).

This value corresponds approximately to the average size of mitotic figures in the data set, and provides a reasonable tolerance for misalignment of the ground truth location and the detection. All detections not within 7.5m of a ground truth annotation are counted as false positives (FP). All ground truth objects without detection within a proximity of 7.5 microns are considered false negatives (FN).

The data set does not contain multiple annotations for the same mitotic figure. We expect participants to also run non-maximum suppression to ensure a mitotic figure object is only detected once. We will count detections matching an annotation only once as true positive, multiple annotations thus lead to false positives.

Metrics:

1) We calculate the number of True Positives, False Positives, and False Negatives accumulated over all images, and from that the total overall F1 score as the primary outcome metric.

2) In the same way, the F1 score for each cancer type will be calculated as a secondary outcome metric.

3) As further and threshold-independent metric we will be utilizing the free-response receiver operating characteristic score (FROC score), which is better suited than the previously used average precision (AP) score for rare event detection problems.

In particular, since we can expect images to not contain any objects, a correct prediction thereof does improve the FROC score, while it does not impact the AP score (which is only considering the number of TP/FN/FP and not the area).

All metrics are in-line with recommendations by the metrics reloaded toolkit (<https://metrics-reloaded.dkfz.de>).

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

We chose F1 because it combines precision and recall in a single metric. It is also the single most often used metric to compare mitotic figure performance, and has been used for the ICPR-MITOS challenges as well as the AMIDA13 and TUPAC16 challenges (mitotic figure detection subtask), and in MIDOG 2021/2022. For the medical application, which commonly involves estimating the density of mitotic figures, an overestimation is equally undesirable as an underestimation. This is why a symmetric measure like F1 is suitable.

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.

We calculate the F1 score ($F1 = 2 \cdot TP / (2 \cdot TP + FN + FP)$) across all images of the test set as the primary metric used for ranking. The other metrics are not used for ranking.

We opt for this assessment instead of averaging F1 scores across cases, because this would lead to a strong dependency on prevalence and detection rate on low grade tumors could strongly deteriorate robustness of the assessment.

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

Since docker containers need to be submitted, the challenge organizers will make sure the algorithms are run on every case of the test set.

Should the docker container fail on single cases of the test set repeatedly, the case will be counted as 0 detections and, consequently, all mitotic figures of the case will be counted as false negatives.

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

The goal of the challenge is domain generalization, and over-estimation is equally bad as under-estimation, which justifies the use of the F1 score. We calculate the F1 score from the totality of twelve undisclosed tumor types. This gives an emphasis on the generalization of the method. This also follows the recommendation by the metrics reloaded protocol by Maier-Hein et al.

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.

We will be using bootstrapping on the test set, and calculate the 95% CI for the evaluation metrics.

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

Bootstrapping is used to assess the variability of results given a different set of cases. This aims at estimating the error for a new test set drawn from the same distribution.

This allows us to report a 95% confidence interval for the F1 score the participants submitted.

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.

We will perform an in-depth analysis of inter-algorithm performance, depending on the actual tumor type and the case. We will report F1 score, precision and recall for each individual case and for each participant. This allows us to showcase strengths and weaknesses of the participating approaches. We will conduct statistical tests such linear mixed effects model analysis for estimation of influencing factors (tumor domains, region selection).

TASK 2: Classification of Atypical Mitotic Figures

SUMMARY

Abstract

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

The participants shall classify detected mitotic figure patches into typical and atypical mitotic figures. While not part of routine tumor grading yet, recent research provides evidence that the rate and density of atypical mitotic figures can be an independent prognosticator for malignancy (10.1038/s41379-022-01080-0, 10.1002/jso.25152, 10.3390/vetsci11010005).

This task is designed to be independent of the first, allowing participants the option to engage with either or both tasks. While combining algorithmic solutions from both tasks could enable the identification and classification of atypical mitotic figures, this task has been intentionally framed as standalone. This approach provides participants greater flexibility in developing novel algorithmic solutions, particularly for the challenges presented in this second task.

Keywords

List the primary keywords that characterize the task.

atypical mitotic figure classification

ORGANIZATION

Organizers

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

Jonas Ammeling, Marc Aubreville, Sweta Banerjee, Christof A. Bertram, Katharina Breininger, Dominik Hirling, Peter Horvath, Nikolas Stathonikos, Mitko Veta

b) Provide information on the primary contact person.

Marc Aubreville

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 2025

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

grand-challenge.org

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

midog2025.deepmicroscopy.org

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.

Publicly available data is allowed

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

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

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

We are actively engaging with potential sponsors to secure monetary prizes, similar to those awarded in MIDOG 2021 and 2022. Prizes will be awarded to the top three positions for this task.

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.

Results will be announced during the challenge workshop (i.e., after the submission deadline)

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).

We aim to publish a summary of the challenge in a peer-reviewed journal (e.g. Medical Image Analysis), as in previous iterations.

Participants are requested to publish a description of their method and results in a preprint as a short paper (2-4

pages) together with their submission. At most two authors of that paper will qualify as authors in the summary paper. Participating teams are free to publish their own results in a separate publication. We aim to publish accepted short papers alongside with invited full papers in the challenge proceedings. Participants may also publish papers including their official performance on the challenge data set, given proper reference of the challenge. There is no embargo time.

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.

Docker container is to be submitted towards grand-challenge. A template docker container with instructions will be made available through github and announced on the challenge website.

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.

As in MIDOG2021/2022, we will provide the possibility to test on a preliminary test set, approximately 2 weeks before the submission deadline. Participants are allowed 1 docker container submission per day to be evaluated on this preliminary set.

Attempts to bypass this limitation might lead to exclusion from the challenge. After preliminary evaluation phase, participants have to submit a final container to be applied on the final test set.

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
- March 15, 2025: Go-Live of the challenge, registration open for participants
 - April 1st, 2025: Availability of training data and dataset description
 - August 14, 2025: Deadline for registration of participants
 - August 15, 2025: Availability of preliminary test set
 - August 31, 2025: Deadline for docker container submission and for two-page preprint abstract submission
 - Sept 13-Sept 27, 2025: Announcement of results at MICCAI 2025

Ethics approval

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

We got ethics approval by the UMC Utrecht (TCBio 20-776), Regional and Institutional Research Ethics Committee of the University of Szeged (SINGLE-MITO-01) and the ethics board of the medical faculty of FAU Erlangen-Nürnberg (AZ 92_14B, AZ 193_18B). For samples taken from the diagnostic archive in veterinary pathology, no ethics approval is required.

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 (Attribution)

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.

Access to the evaluation docker container will be made available on github.

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

We will provide links to the docker containers the participants submitted, given their consent. Participants will be encouraged to give permission for this and make their code publicly available.

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 organizers declare no conflict of interest. Access to the test case labels will be available to the organizers only, and on a need-to-know basis to perform the evaluation.

MISSION OF THE CHALLENGE

Field(s) of application

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

Examples:

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

Research, Prognosis

Task category(ies)

State the task category(ies)

Examples:

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

Classification

Cohorts

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

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

The target cohort consists of patients with confirmed cancer and with biopsies taken for histopathologic assessment.

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

The challenge cohort consists of 520 cases of cancer, with a tissue area of 2mm² each, collected at the UMC Utrecht, the Universitätsklinikum Erlangen (University Hospital Erlangen) and the University of Veterinary Medicine Vienna. The mitotic figure annotations of these cases were previously used within the MIDOG 2022 challenge data set and all mitotic figures thereof are being subclassified into atypical and typical/normal mitotic figures for this challenge.

Imaging modality(ies)

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

Histopathology images, hematoxylin and eosin-stained, whole slide image acquisition.

Context information

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

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

We do not provide additional information along with the image set.

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

As this is not relevant for the task, we do not provide meta data about the patient.

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.

The data consists of image patches cropped around ground truth mitotic figure annotations, selected from whole slide image scans of human and animal cancer tissue, acquired at six different institutions.

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

The algorithm target is the classification of mitotic figures into atypical mitotic figures and normal/typical mitotic figures. The algorithm shall generate to different tumor types, which are not disclosed to the participants.

Assessment aim(s)

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

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

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

Accuracy,Robustness,Sensitivity,Specificity

DATA SETS

Data source(s)

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

At UMC Utrecht: Hamamatsu NanoZoomer XR

At VetMedUni Vienna: 3DHistech Panoramic Scan

At UK Erlangen: Hamamatsu Nanoszoomer S60

At FU Berlin: Aperio ScanScope CS2

At HS Flensburg: Aperio ScanScope CS

At HUN-REN Biological Research Centre: 3DHistech P1000

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

Hamamatsu XR: resolution: 0.23 microns / px at 40x

3dHistech Panoramic Scan: 0.25 microns / px at 40x

Nanoszoomer S60: resolution: 0.22 microns / px at 40x

Aperio ScanScope CS / CS2: 0.25 microns / px at 40x

3DHistech P1000: 0.25 microns / px at 40x

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.

- Histopathology archive of the UMC Utrecht
- Neuropathology archive of the Universitätsklinikum Erlangen
- Diagnostic archive of the Institute of Pathology at VetMedUni Vienna
- HUN-REN Biological Research Centre, Szeged

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

The regular diagnostic workflow for H&E-stained; histopathology images at the respective institutions were applied. We expect a dependency on the staining protocols of the institutions, but imply to capture real-world variability in our data set.

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 are given as a section of a whole slide microscopy images. Cases represent individual patients. We ensure no patient is part of training and test set.

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

Training set: We will provide a training set of 11,937 mitotic figure cropouts (128x128px), extracted from 503 cases, which is derived from the publicly available MIDOG++ dataset, which also serves as training set of the Task 1, and extends the just released AMI-Br dataset (github.com/DeepMicroscopy/AMI-Br/).

Test set: We will label the mitotic figure crop-outs from the 120 cases of the ideal regions of the Task 1 test set for the classes atypical mitotic figure and normal mitotic figure. As the test set is hidden to the public, we cannot provide the exact number of mitotic figures in this description, but expect an approximately proportional number compared to the training set.

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

The challenge dataset represents a good trade-off between capturing the naturally occurring variability of whole slide histopathology images and time invested for the annotation. The selected proportion of training, validation (preliminary test) and test cases allows for a realistic estimation of the performance. At the same time, it allows to test how robust the methods are with respect to domain shift across cancer types.

Additionally, it represents the largest cross-domain dataset of atypical mitotic figures.

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.

Participants can not expect statistical equality of the characteristics of the training and test sets, since the tissue characteristics of the training set and the test set are different.

e) 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.

We will use mitotic figure patches cropped out from private data, that spans multiple tumor domains and was part of the MIDOG 2022 test set.

Annotation characteristics

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

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

We use a three-expert consensus for annotation into atypical and typical/normal mitotic figures.

In the rare case that more than one mitotic figure is part of the image patch, the label refers to the mitotic figure in the center of the patch.

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.

We will provide recent guidelines [3] towards mitotic figure identification and subtyping to all annotators.

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.

Experts are pathologists with an extensive experience in mitotic figure classification and subtyping.

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.

We use a majority vote of the experts. We additionally provide individual expert opinions as part of our dataset.

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.

We center our crop-outs around mitotic figures that were identified as a three expert consensus (see task 1). We choose an equal pixel area (128x128 px) around each mitotic figure to facilitate processing.

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.

The task of atypical classification is known to have only a moderate inter expert agreement. This is our main source of error. We use a three expert consensus as a compromise between scale and diversity of data and our available resources.

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

Besides annotation errors we do not expect other sources of error.

ASSESSMENT METHODS

Metric(s)

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

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

We use the balanced accuracy as main metric for this task.

We additionally report sensitivity and specificity and ROC AUC for the task as secondary but not rank-defining metric.

Our metrics are in line with the recommendation by the metrics reloaded toolkit (<https://metrics-reloaded.dkfz.de>).

We will additionally split this down according to tumor domains.

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

Atypical mitotic figures occur at highly different rates across patients, but overall account for approximately only 20% of mitotic figures. Since we cannot assume a static prior during inference, it is necessary to use a prevalence-invariant metric, such as balanced accuracy.

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.

We use the balanced accuracy calculated over all patches of the test set.

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

Since docker containers need to be submitted, the challenge organizers will make sure the algorithms are run on every case of the test set. In case of a failure of the docker container on one or multiple cases, these will count as wrong classifications in the balanced accuracy.

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

As the rate of atypical mitotic figures is highly tumor- and thus patient dependent, we cannot assume a prior. Thus, we need to use a prevalence-independent metric and decided for balanced accuracy. This also follows the recommendation by the metrics reloaded protocol by Maier-Hein et al.

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.

We will be using bootstrapping on the test set, and calculate the 95% CI for the evaluation metrics.

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

Bootstrapping is used to assess the variability of results given a different set of cases. This aims at estimating the error for a new test set drawn from the same distribution.

This allows us to report a 95% confidence interval for the balanced accuracy the participants submitted.

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.

We will perform an in-depth analysis of inter-algorithm performance, depending on the actual tumor type and the case.

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.

N/A

Further comments

Further comments from the organizers.

N/A