Mycetoma MicroImage: Detect & Classify: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Mycetoma MicroImage: Detect & Classify

Challenge acronym

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

mAIcetoma

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.

Mycetoma is a chronic granulomatous inflammatory disease reported worldwide but endemic in tropical and subtropical areas. It was included in the World Health Organization (WHO) Neglected Tropical Disease (NTD) list in 2016. Mycetoma can be caused either by certain types of bacteria (Actinomycetoma, AcM) or fungi (Eumycetoma, EuM). Identification of mycetoma types plays a significant role in the treatment of mycetoma. An incorrect diagnosis of mycetoma can seriously affect the patient and the disease prognosis and outcome.
The proposed challenge will explore the feasibility of using AI to effectively diagnose mycetoma from histopathological images of mycetoma grains. This challenge is the first of its kind to address mycetoma diagnosis. It aims to promote the development of AI and computer vision solutions for the automatic detection of mycetoma infection in histopathological microscopic images and provides a classification of the disease as EuM or AcM. Researchers and data scientists from around the world are invited to participate in this challenge to advance the early diagnosis and management of mycetoma, ultimately improving patient outcomes and healthcare in affected regions.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

Mycetoma, Diagnosis, Hitopathology, Microscopic images, Detection, Classifcation.

Year

The challenge will take place in 2024

FURTHER INFORMATION FOR CONFERENCE ORGANIZERS

Workshop

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

Virtual event.

Duration

How long does the challenge take?

Half day.

Expected number of participants

Please explain the basis of your estimate (e.g. numbers from previous challenges) and/or provide a list of potential participants and indicate if they have already confirmed their willingness to contribute.

N/A

Publication and future plans

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

A challenge paper will be submitted for publication with all participants as co-authors. Participants can submit their papers after submitting the challenge paper.

Space and hardware requirements

Organizers of on-site challenges must provide a fair computing environment for all participants. For instance, algorithms should run on the same computing platform provided to all.

N/A

TASK 1: Mycetoma MicroImage: Detect and classify

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.

Mycetoma is a chronic granulomatous inflammatory disease reported worldwide but endemic in tropical and subtropical areas. It was included in the World Health Organization (WHO) Neglected Tropical Disease (NTD) list in 2016. Mycetoma can be caused either by certain types of bacteria (Actinomycetoma, AcM) or fungi (Eumycetoma, EuM). Identification of mycetoma types plays a significant role in the treatment of mycetoma. An incorrect diagnosis of mycetoma can seriously affect the patient and the disease prognosis and outcome.
The proposed challenge will explore the feasibility of using AI to effectively diagnose mycetoma from histopathological microscopic images of mycetoma grains. This challenge is the first of its kind to address mycetoma diagnosis. It aims to promote the development of AI and computer vision solutions for the automatic detection of mycetoma infection in histopathological microscopic images and provides a classification of the disease as EuM or AcM.
Researchers and data scientists from around the world are invited to participate in this challenge to advance the early diagnosis and management of mycetoma, ultimately improving patient outcomes and healthcare in affected regions.

Keywords

List the primary keywords that characterize the task.

Mycetoma, Diagnosis, Hitopathology, Microscopic images, Detection, Classifcation

ORGANIZATION

Organizers

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

Hyam Ali, University of Khartoum, Sudan.
Sahar Alhesseen, University of Khartoum, Sudan.
Lamis Elkhair, University of Khartoum, Sudan.
Rachid Jennane, Orleans University, France.
Ahmed Fahal, University of Khartoum, Sudan.

b) Provide information on the primary contact person.

Hyam Ali

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.

Challenge venue and platform

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

N/A

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

Synapse.

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

https://www.synapse.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.

Additional data allowed. Participants are required to submit details of this additional data to the organizers. The additional data should be clarified and explained, including details such as the type of data, the number of images, and any preprocessing steps applied.

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

Members can participate.

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

No prizes are planned.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

Top five performing methods will be announced publicly.

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).

A challenge paper will be submitted for publication with all participants as co-authors. Participants can submit their own paper after submitting the challenge paper.

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.

A link for submission instructions will be sent to the participants. Jupyter Notebook can be used to display the algorithms' output.

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.

Up to three submissions for the results are allowed. Only the last run is considered to compute challenge results.

Challenge schedule

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

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

The registration will start as soon as the challenge is accepted and the training dataset will be accessible. The challenge is a half-day virtual event. The test dataset will be released one week before the challenge. The announcement of the results is scheduled for the last day of the workshops.

Ethics approval

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

The data usage was approved by the Soba University Hospital Ethical Committee, Khartoum, Sudan (No. SUH 05/01/2019). Written informed consent was obtained from each patient to collect the data.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

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

CC BY-NC.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The code used by organisers for the evaluation will be publicly available on Git-Hub.

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

Participants are encouraged to share their codes in the public repositories.

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 organisers declare that they have no conflicts of interest. No funds were received to conduct this challenge. Only organisers will have access to the test case labels. Shortly after the workshop, the labels will be released.

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

Screening, Diagnosis, CAD.

Task category(ies)

State the task category(ies)

Examples:

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

Detection and 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.

Mycetoma suspected patients.

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

Mycetoma histopathological images.

Imaging modality(ies)

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

Microscopic Images.

Context information

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

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

Ground-truth segmentation of each image will be given.

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

The disease type for all the patients.

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.

Skin tissue which is shown in histopathological microscopic images.

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.

Mycetoma grains' in skin tissue.

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)).

With high accuracy, the algorithm should identify mycetoma grain(s) within the histopathological images. Then, classify the types of mycetoma present in the identified grain(s) with high sensitivity and specificity as well as the Mathew Correlation Coefficient.

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).

Nikon Eclipse 80i digital microscope.

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

Microscopic images were captured using a digital microscope with the conditions below:
Dimension: 800X600.
Magnification: 10X.
Colour space: RGB.

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.

The data was collected at the Mycetoma Research Center (MRC), University of Khartoum in Khartoum, Sudan.

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).

Pathologists with at least three years of experience in mycetoma diagnosis identified mycetoma infection in the histopathological images and determined its types. In addition, an expert data annotator from the MRC, with experience in labelling medical images, manually segmented grains on the histopathological images.

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.

For each patient, several histopathological microscopic slides/images were defined. There is an average of six images per patient.
A case refers to a single image.

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

A total of 863 images. The dataset is split into training, validation, and test sets with 70%, 10% and 20% proportions, respectively.

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

The total number of cases is sufficiently large to encompass the diversity, and variability present in mycetoma cases as well as the balance of the mycetoma types.
The dataset is partitioned into training, validation, and test sets with proportions of 70%, 10%, and 20%, respectively.

Given the average of six images per patient, a careful splitting strategy was employed to prevent potential statistical bias that might arise from allocating images from the same patient into different sets. Consequently, each patient's images are exclusively assigned to either the training, validation, or test sets, ensuring that they are not shared across multiple sets. This approach aims to maintain the integrity of patient-specific data within each split.
Our tests have demonstrated that these specific proportions are advantageous for robust detection and accurate classification, ensuring a well-balanced distribution of data for effective model training and evaluation.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

The samples in this study primarily come from Sudan, influencing the distribution of mycetoma types (EuM and AcM).

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.

The manual segmentation will be used as a ground-truth annotation for the detection of mycetoma grain, while the differentiation of the types defined by the expert pathologists will be used to assign mycetom class.
A single data annotator was involved.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The annotator used ImageJ software to annotate the images.

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.

A computer scientist who medically trained for two years on the histopathological microscopic images and mycetoma grains.

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.

NA.

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.

No pre-processing will be applied.

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.

Mycetoma grains sometimes have a solid or rigid texture which results in improper appearance of mycetoma grains. Such circumstance leads to the presence of small remnants in the image beside the main grain. These remnants were not considered during the annotation. Also, the image might include several grains, however, an individual grain is labelled.

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

Each of mycetoma types is divided into several sub-types. An imbalance in sub-type distribution can lead to biased predictions.

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)

The performance will be assessed using sensitivity, specificity, accuracy, and Matthew correlation coefficients (MCC).

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

Sensitivity and specificity focus on minimizing false negatives and false positives, respectively, which are critical in mycetoma diagnoses. Accuracy provides an overall assessment, and MCC is particularly useful for imbalanced type distribution.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

The use of a weighted sum allows for ranking based on the importance of each metric. Different participants may prioritize metrics differently. Weighted sum provides flexibility to align the evaluation with the specific goals of the challenge.

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

Submissions with missing results will be managed by assigning default scores to the missing results for each metric.
Assigning default scores helps maintain fairness in the evaluation process. The choice of these scores is based on the performance of a model with missing results as well as the importance of each metric. Hence:
Sensitivity, Specificity, and Accuracy: A default score of 0.5 (average performance).
MCC: A default score of 0 (indicating no correlation)

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

The ranking method for performance evaluation involves a weighted sum such that different weights will be assigned to different metrics based on their importance. The individual metric scores are then combined using a weighted sum to obtain a composite score for each algorithm.
The choice of weight aligns with the objectives of the challenge and it ranges in [0,1]. Sensitivity and specificity have a similar impact in the context of the challenge. Accuracy is deemed more critical than sensitivity and specificity so a higher weight is assigned to it. MCC has a higher weight since it indicates solid statistical accuracy taking into account the different sizes of classes. Therefore,
Sensitivity (w_sen): 0.15
Specificity (w_spec): 0.15
Accuracy (w_accu): 0.3
MCC (w_mcc): 0.4

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.

The individual metric scores are combined for each algorithm using a weighted sum as below:
weighted_sum = (w_sen*Sensitivity) + (w_spec*Specificity) + (w_accu*Accuracy) + (w_mcc*MCC)
Then ordering method will be used for ranking the results based on their weighted sum. The result with the highest overall score is ranked the top performance.

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

Using a combination of the aforementioned metrics ensures a comprehensive evaluation. However, assigning different weights to each metric allows for flexibility in reflecting the importance of each metric according to the specific goals of the challenge.
Using an ordering method based on the weighted sum ensures that algorithms are ranked in descending order of their overall scores. This straightforward approach simplifies the interpretation of results, making it clear which algorithm performed the best according to the specified evaluation criteria.

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.

NA.

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.

NA.

Further comments

Further comments from the organizers.

NA.