Automatic Lung Cancer Detection and Classification in Whole-slide Histopathology: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Automatic Lung Cancer Detection and Classification in Whole-slide Histopathology

Challenge acronym

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

ACDC@LungHP

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.

Digital pathology has been gradually introduced in clinical practice. Although the digital pathology scanner could give very high resolution whole-slide images (WSI) (up to 160nm per pixel), the manual analysis of WSI is still a time-consuming task for the pathologists. Automatic analysis algorithms offer a way to reduce the burden for pathologists. Our proposed challenge will focus on automatic detection and classification of lung cancer using Whole-slide Histopathology. This subject is highly clinical relevant because lung cancer is the top cause of cancer-related death in the world. The first stage of the challenge (ACDC2019) was already successfully held in 2019 in ISBI (https://acdc-lunghp.grand-challenge.org/). ACDC2019 mainly focused on the detection of lung cancer region in WSIs. ACDC2020 will focus on classifying the main lung cancer subtypes (e.g. squamous carcinoma, adenocarcinoma) using WSI.

Challenge keywords

List the primary keywords that characterize the challenge.

Digital pathology; Lung Cancer; Whole-slide Histopathology

Year

The challenge will take place in ...

2020

FURTHER INFORMATION FOR MICCAI ORGANIZERS

Workshop

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

None

Duration

How long does the challenge take?

Half day.

Expected number of participants

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

The number of participants (on-site) from our previous challenge (ACDC Challenge at ISBI 2019) is about 20 in total. We would expect 20-30 participants for ACDC 2020.

Publication and future plans

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

The organizers of the challenge will disseminate the result to the scientific community through the publication of one overview paper after the challenge in a high-impact journal. Top 10 teams will be invited as co-authors of the paper. Each team could have a maximum of two co-authors in the 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.

We will use the grand challenge platform to compare all algorithms as we did in ACDC 2019.
The technical equipment are listed as below:

Item	Quantity
Computers	2
Monitors	2
Projectors	2
Microphones	4

TASK: Multiclass Classification of Lung Cancer

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.

Digital pathology has been gradually introduced in clinical practice. Although the digital pathology scanner could give very high resolution whole-slide images (WSI) (up to 160nm per pixel), the manual analysis of WSI is still a time-consuming task for the pathologists. Automatic analysis algorithms offer a way to reduce the burden for pathologists. Our proposed challenge will focus on automatic detection and classification of lung cancer using Whole-slide Histopathology. This subject is highly clinical relevant because lung cancer is the top cause of cancer-related death in the world. The first stage of the challenge (ACDC2019) was already successfully held in 2019 in ISBI (https://acdc-lunghp.grand-challenge.org/). ACDC2019 mainly focused on the detection of lung cancer region in WSIs. ACDC2020 will focus on classifying the main lung cancer subtypes (e.g. squamous carcinoma, adenocarcinoma) using WSI.

Keywords

List the primary keywords that characterize the task.

Digital pathology; Lung Cancer; Whole-slide Histopathology

ORGANIZATION

Organizers

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

Technical Group:
1.National University of Defense Technology, China (Zhang Li, Dr., Xichao Teng, Dr., Jiehua Zhang)
2.The Radboud University Medical Center in Nijmegen, the Netherlands (Francesco Ciompi, Dr.)
3.The Eindhoven University of Technology, the Netherlands (Tao Tan, Dr.)
4.Nanjing University of Information Science & Technology(Jun Xu, Dr.)
5.Memorial Sloan Kettering Cancer Center, USA (Peter Schüffler, Dr.)
6.Inception Institute of Artificial Intelligence, UAE (Dwarikanath Mahapatra, Dr.)
7.Hunan Lanxi Biotechnology Co. Ltd., China (Xiangjun Feng, Prof.).

Medical Group:
1.The First Hospital of Changsha, China (Yuling Tang, Prof., Hui Chen, Prof)
2.Hunan Cancer Hospital, the Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, China (PhD, Zhihong Liu, Prof., Jun Hu, Prof.)
3.The Second Xiangya Hospital, Central South University, China (Daiqiang, Li, Prof., Jiang, Yi Prof.)

b) Provide information on the primary contact person.

Zhang Li, zhangli_nudt@163.com

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.

Examples:

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

One-time event with fixed submission deadline

Challenge venue and platform

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

MICCAI.

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

grand-challenge.org

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

https://acdc-lunghp.grand-challenge.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 in the challenge but are not eligible for awards.

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

1000 euro for the first team, 600 euro for the second team, and 400 euro for the third team if source codes are released before the conference.

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 10 teams 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).

The organizers of the challenge will disseminate the result to the scientific community through the publication of one overview paper after the challenge in a high-impact journal. Top 10 teams will be invited as co-authors of the paper. Each team could have a maximum of two co-authors in the 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.

The participant will submit their result through the grand-challenge platform as we did in ACDC2019. The submission instructions will be announced on the 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.

In total ten submissions are allowed for each team. And the best submission of each team will be chosen for the ranking.

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

2020.04.30 Update the ACDC2019 Challenge website to ACDC2020 and open for registration.

2020.05.30 Upload first batch of dataset

2020.06.30 Upload training dataset

2020.09.01 Upload test and validation dataset

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 have the ethics approval for all date that used for this challenge

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.

Additional comments: ONLY for research purposes.

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.

It will be embed in grand-challenge website as we did for ACDC2019. We will also provide evaluation codes on github.

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

Participant are not required to share their codes. However, only the team that share their codes (e.g. github) will get the prize for their ranking.

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.

None conflicts of interest.

The funding will be given by NSFC 61801491.

Only the organizers have the access to all the case labels around next June.

MISSION OF THE CHALLENGE

Field(s) of application

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

Examples:

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

Diagnosis, Decision support.

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.

WSI from patients with clinically suspected lung cancers.

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

WSI from patients with histologically confirmed lung cancers.

Imaging modality(ies)

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

Digital pathology (Whole slide imaging)

Context information

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

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

Each image may contain tissues that correspond to one subtype of lung cancer. Four main subtypes will be considered in this challenge. There are squamous cell carcinoma, adenocarcinoma, small cell carcinoma and undifferentiated carcinoma.

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

The multi-center consecutive patients with diagnosed stage I -stage IV lung cancer from January 2016, through November 2018 were recruited for this challenge. Other inclusion criteria included: 1)no radiotherapy before surgery; 2) aged between 30 and 90 yr;3) detailed clinical information is available. The exclusion criteria were: 1) multiple primary cancers; 2) lung metastases from other primary malignant diseases; 3) patients with immune-deficiency or organ-transplantation history; 4) patients without detailed clinical information; 5) patients who did not provide informed consent.

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.

Lung tissue shown in digital pathology.

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.

AI algorithms for the subtype classification of lung cancers.

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)).

Sensitivity, Accuracy.

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).

The scanner of 3dhistech (https://www.3dhistech.com/) will be used as one of the main image devices as we did for ACDC 2019. Other digital scanners may be used for algorithm evaluation purposes.

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

The images used for this challenge will be H&E stained WSI. The image acquisition will follow the standard preparation and imaging procedure of pathologists.

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.

Data are mainly acquired from the challenge organizers' institutes.
The data of ACDC2020 will be uploaded to Zenodo, Google Drive and Baidu Pan for participants from different regions.

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).

Surgeon.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

Training and test cases both represent a WSI images of lung biopsy.
Training and test cases both have a image-level label (subtype of lung cancer).

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

Training: 2500 cases from different institutes (with NORMAL and CANCER cases)
Validation and test: 500 cases from different institutes (with NORMAL and CANCER cases)
Depending on the data cleaning process, more cases may be released in the end.

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

The total number of the data may enough to classify the main subtype of the lung cancer. The specific proportion

is just follow the standard rules.

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.

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.

Each image has one label related to one subtype.

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.

Following their clinical protocols.

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.

At least two experienced pathologists (over 20 years of work) will make the labels for the data. The final annotations are based on the consensus of the two pathologists.

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.

Final annotations are based on a consensus.

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.

Images from different digital pathology scanners will be converted to the multi-resolution TIFF format.

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.

We will have two experienced pathologists to make labels. The final annotations are based on the consensus of the two pathologists. The possible error may from the misdiagnose by both pathologists.

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

Not applicable.

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)

Image-level Micro F1.

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

Classification of main subtypes of lung cancer is a imbalanced classification problem. The Micro-F1 metric is normally used for this multi-class classification problems.

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 Micro F1 will be used as the ranking method. The Micro F1 score is performed on database case level rather than individual case and this one number metric can be directly used for ranking.

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

Submission that contains missing results will not be process on platform.

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

One number metric is used for ranking. A higher value of the metric means better classification performance.

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 test if the difference in the metrics are statistically different or not, using bootstrapping. The bootstrapping was introduced in PASCOL VOC and ImageNet challenge. For each bootstrap round, we sample N images with replacement from all N test images and calculate the Micro F1 of the submitted method on those sampled images. We repeat the bootstrap for several rounds and discard the lower and upper X% fraction (e.g. 2.5%). The range of remaining results represents the 1-2X% (e.g. 95%) confidence interval. We could compare the confidence interval to see the metrics are statistically different or not.

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

It is common to compute statistical significance between different methods.

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.

Ensembling would be applied.

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.

[1] Peter Bandi, et al., "From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge." IEEE Transactions on Medical Imaging, vol.38, no.2, pp. 550 - 560, 2019. DOI: 10.1109/TMI.2018.2867350
[2] Zhang Li, et al. Computer-aided diagnosis of lung carcinoma using deep learning - a pilot study. https://arxiv.org/abs/1803.05471, 2019.