Thyroid Nodule Segmentation and Classification in Ultrasound Images: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Thyroid Nodule Segmentation and Classification in Ultrasound Images

Challenge acronym

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

TN-SCUI2020

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.

The thyroid gland is a butterfly-shaped endocrine gland that is normally located in the lower front of the neck. It secretes indispensable hormones that are necessary for all the cells in the body to work normally [1]. The term thyroid nodule refers to an abnormal growth of thyroid cells that forms a lump within the thyroid gland [2].

Statistical studies showed that the incidence of this disease increases with age, extending to more than 50 % of the world's population. Until recently, thyroid cancer was the most quickly increasing cancer diagnosis in the United States. It is the most common cancer in women 20 to 34 [3]. Although the vast majority of thyroid nodules are benign (noncancerous), a small proportion of thyroid nodules contains thyroid cancer. In order to diagnose and treat thyroid cancer at the earliest stage, it is desired to characterize the nodule accurately.
Thyroid ultrasound is a key tool for thyroid nodule evaluation. It is non-invasive, real-time and radiation-free. However, it is difficult to interpret ultrasound images and recognize the subtle difference between malignant and benign nodules. The diagnosis process is thus time-consuming and heavily depends on the knowledge and the experience of clinicians.

Recently, many computer-aided diagnosis (CAD) systems have been used to alleviate this problem. However, it is usually difficult to evaluate each of their efficacy as no benchmark was available so far. Our challenge, named TN-SCUI2020, aims to provide such a platform to validate all of the state-of-the-art methods and exchange for new ideas.

The main topic of this TN-SCUI2020 challenge is finding automatic algorithms to accurately segment and classify the thyroid nodules in ultrasound images. It will provide the biggest public dataset of thyroid nodule with over 4,500 patient cases from different ages, genders, and were collected using different ultrasound machines. Each ultrasound image is provided with its annotated class (benign or malignant) and a detailed delineation of the

nodule. The dataset comes from the Chinese Medical Ultrasound Artificial Intelligence Alliance (CMUAIA) which was initiated by Dr. Jiaqiao Zhou, Department of Ultrasound, Ruijin Hospital, School of Medicine, Shanghai Jiaotong University. This challenge will provide a unique opportunity for participants from different backgrounds (e.g. academia, industry, and government, etc.) to compare their algorithms in an impartial way.

Challenge keywords

List the primary keywords that characterize the challenge.

Thyroid Nodule, Ultrasound Images, Segmentation, Classification

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.

Advances in Simplifying Medical Ultrasound (ASMUS)

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.

Firstly, we are planning to gain a big number of students in our universities as participants for this challenge. For instance, these students come from the Biomedical Engineering Department, Electrical Engineering Department, Computer Science Department of Shenzhen University, Shanghai Jiaotong University, University of Oxford, Eindhoven University of Technology. The board members of this challenge are professors from the relevant departments and universities have authority and responsibility in announcing and broadcasting the challenge to the students via the university system. The students have showed their interest in our challenge because of the given goals.

Secondly, our challenge is hosted on grand-challenge.org, which is a well-known platform in Medical Image Analysis challenge, that could promisingly attract the participation from hundreds of researchers all over the world.

The expected number of participants is more than 500.

Publication and future plans

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

The award winning teams with outstanding techniques and excellent performances will be selected as speakers and for publications in the MICCAI 2020 workshop. All of the winning team members qualify as authors of their workshop paper. No publication rights will be given on the dataset images apart from this challenge. The participants will not be able to use the dataset images in any other publication or study.

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.

Our challenge will be named as TN-SCUI2020 and is hosted at following link https://tn-scui2020.grand-challenge.org/Home/. All of other information, discussion and latest announcement also can be found there.

All the participants are provided the same dataset and the online submission system of grand-challenge.org to make sure the same condition. After the participating teams get the prediction results of their algorithm, they will upload their results to the online submission system. The scores of participants will be immediately computed by the server and shown in our website.

For the server of valuation, we use the our server of Supermicro (SYS-4028GR-TR). The server includes 2 CPUs of Intel® Xeon® processor E5-2600, 288GB of RAM, 8 GPUs of Geforce GTX 1080 Ti 12GB. The server runs on Ubuntu 4.4.0 of 64 bit version.

TASK: Thyroid Nodule Segmentation and Classification in Ultrasound Images

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 thyroid gland is a butterfly-shaped endocrine gland that is normally located in the lower front of the neck. It secretes indispensable hormones that are necessary for all the cells in the body to work normally [1]. The term thyroid nodule refers to an abnormal growth of thyroid cells that forms a lump within the thyroid gland [2].

Statistical studies showed that the incidence of this disease increases with age, extending to more than 50 % of the world's population. Until recently, thyroid cancer was the most quickly increasing cancer diagnosis in the United States. It is the most common cancer in women 20 to 34 [3]. Although the vast majority of thyroid nodules are benign (noncancerous), a small proportion of thyroid nodules contains thyroid cancer. In order to diagnose and treat thyroid cancer at the earliest stage, it is desired to characterize the nodule accurately.

Thyroid ultrasound is a key tool for thyroid nodule evaluation. It is non-invasive, real-time and radiation-free. However, it is difficult to interpret ultrasound images and recognize the subtle difference between malignant and benign nodules. The diagnosis process is thus time-consuming and heavily depends on the knowledge and the experience of clinicians.

Recently, many computer-aided diagnosis (CAD) systems have been used to alleviate this problem. However, it is usually difficult to evaluate each of their efficacy as no benchmark was available so far. Our challenge, named TN-SCUI2020, aims to provide such a platform to validate all of the state-of-the-art methods and exchange for new ideas.

The main topic of this TN-SCUI2020 challenge is finding automatic algorithms to accurately segment and classify the thyroid nodules in ultrasound images. It will provide the biggest public dataset of thyroid nodule with over 4,500 patient cases from different ages, genders, and were collected using different ultrasound machines. Each ultrasound image is provided with its annotated class (benign or malignant) and a detailed delineation of the nodule. The dataset comes from the Chinese Medical Ultrasound Artificial Intelligence Alliance (CMUAIA) which was initiated by Dr. Jiaqiao Zhou, Department of Ultrasound, Ruijin Hospital, School of Medicine, Shanghai Jiaotong University. This challenge will provide a unique opportunity for participants from different backgrounds (e.g. academia, industry, and government, etc.) to compare their algorithms in an impartial way.

Keywords

List the primary keywords that characterize the task.

Thyroid Nodule, Ultrasound Images, Segmentation, Classification

ORGANIZATION

Organizers

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

The challenge is organized as a joint effort of different institutions and researchers in Shanghai Ruijin Hospital, China; Shenzhen University, China; the University of Oxford, UK and the Eindhoven University of Technology, Netherlands. We are mainly divided into two corresponding groups for our challenge.

Medical Group:
- Prof. Jianqiao Zhou, Department of Ultrasound, Ruijin Hospital, School of Medicine, Shanghai Jiaotong University, China.
- Dr. Xiaohong Jia, Department of Ultrasound, Ruijin Hospital, School of Medicine, Shanghai Jiaotong University, China.

Technical Group:
- Prof. Dong Ni, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Health Science Center, School of Biomedical Engineering, Shenzhen University, China.
- Prof. Alison Noble, Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK.
- Dr. Ruobing Huang, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Health Science Center, School of Biomedical Engineering, Shenzhen University, China.
- Dr. Tao Tan, Eindhoven University of Technology, The Netherlands.
- Ph.D Candidate Manh The Van, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Health Science Center, School of Biomedical Engineering, Shenzhen University, China.

b) Provide information on the primary contact person.

Ph.D Candidate Manh The Van, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Health Science Center, School of Biomedical Engineering, Shenzhen University, China.
Email: mtv@email.szu.edu.cn
Dr. Ruobing Huang, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Health Science Center, School of Biomedical Engineering, Shenzhen University, China.
Email: ruobing.huang@szu.edu.cn

Life cycle type

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

Examples:

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

One time event.

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://tn-scui2020.grand-challenge.org/Home/

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.

Be restricted to the data provided by the challenge

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.

Everyone is encouraged to enter the competition but the conflict of interests should be stated clearly by participants for paper reviewing.

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

We provide multiple awards according to the ranking list of all participants. The prizes for the winners will be confirmed and displayed on our website later.

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 three performing methods in segmentation and classification task will be announced respectively, in public in the MICCAI workshop.

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 award winning teams with outstanding techniques and excellent performance will be selected as speakers and for publications in the MICCAI 2020 workshop. All of the winning team members qualify as authors of their workshop paper.All the other teams can submit their methods at this workshop and selected papers will be published base on peer-reviewed results. No publication rights will be given on the dataset images apart from this challenge. The participants will not be able to use the dataset images in any other publication or study.

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.

After the participating teams get the prediction results of their algorithm, they will upload their results to the

online submission system. The score will be immediately computed by the server and shown on result leaderboard of challenge website.
The submission is processed by the following link
https://tn-scui2020.grand-challenge.org/evaluation/submissions/create/

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.

- Provide the large size of the training set.
- The test data will be released three weeks before the submission.
- Each team can only make 3 submissions and the last run is officially counted to compute challenge results.
- Each person can only belong to one team.

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

- Announcement of challenge's opening: March 15, 2020
- Open for registration: April 1, 2020
- Training data release: May 1, 2020
- Testing data release: July 7, 2020
- Deadline for submission: July 30, 2020
- Winners announcement and oral invitations: August 15, 2020
- Workshop day: October 8, 2020

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).

Data collection and usage have been approved by local hospital.

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 ND.

Additional comments: The data is to be released under a custom license similar to CC-BY-NC-ND but prohibiting any use outside of this challenge.

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 information on the accessibility of the organizers' evaluation software can be found on the following links
https://tn-scui2020.grand-challenge.org/evaluation/

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

The participants are encouraged to upload their code to Github and share the Github link in the PDF submission file.

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.

There are no conflicts of interest. The challenge is sponsored by Ruijin Hospital, Shanghai Jiaotong University and School of Medicine and Health Center, Shenzhen University, China.
The technical group of the challenge has access to the test case labels during the lifetime of the challenge.

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.

Task category(ies)

State the task category(ies).

Examples:

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

Segmentation 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.

The target cohort are the patients with thyroid nodules from any possible hospital or medical centers diagnosed by ultrasound.

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

The target cohort are the patients with thyroid nodule disease from any possible hospital or medical centers diagnosed by ultrasound.

Imaging modality(ies)

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

Ultrasound B mode 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).

The images are 2D B mode ultrasound images (.png format) that were collected in different sites around China using different ultrasound machines (e.g. Mindray DC-8, Philips-cx50, TOSHIBA Aplio300, SIEMENS ACUSONS 2000, SIEMENS ACUSON NX3 Elite, Esaote Mylab).

The segmentation masks are binary images in which pixels are either 1 for the foreground or 0 for the background. The annotation images named as “xxxx.png”, where “xxxx” presents patient ID (from 0001 to 4000). The annotation of the classification task will be provided in a CSV file with a header of PID (denotes for patient ID) and CATE (0 refer to benign, and 1 refer to malignant). The annotation of nodules is labeled and responsible by expert group of two experienced doctors.

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

The images are acquired from thyroid patients with a variety of gender (male or female) and age (from 9-year old to 82-year old).

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 origin is thyroid gland shown in the B mode ultrasound image.

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 thyroid nodule shown in the B mode ultrasound image of the thyroid gland. Each image has an either benign or malignant thyroid nodule.

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)).

Specificity, 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).

US images were acquired from systems of different vendors such as Mindray DC-8, Philips-cx50, TOSHIBA Aplio300, SIEMENS ACUSONS 2000, SIEMENS ACUSON NX3 Elite, Esaote Mylab.

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

The patient laid face-up on an exam table and the physician placed the ultrasound probe on the area of interest and moved back and forth until the desired images were captured.

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.

In the challenge, the participants will be provided the US nodular thyroid dataset supplied by Shanghai Ruijin Hospital, China.

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 data was collected from different sites by local doctors (organized by CMUAA), while the label were provided by Dr. JianQiao Zhou with over 20 years of experience at Department of Ultrasound, Ruijin Hospital, School of Medicine, Shanghai Jiaotong University, China.

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 an ultrasound image of the thyroid nodule of each patient. The training cases include the corresponding annotations of classification. The training cases are also accompanied by the mask annotation binary images in which pixels are either 1 for the foreground or 0 for the background. The mask annotation images named as “xxxx.png”, where “xxxx” presents patient ID (from 0001 to 4000). A case refers to a thyroid nodule patient.

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

The dataset to-be-released contains a training set of 4,000 cases and a test set of 500 cases

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

All the images from different centers with corresponding masks and labels were collected to the same folder and then were shuffled and divided into the training, validation and test sets. We did not seperate the dataset according to the proportion of data from each center to avoid overfitting to any particular center.

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 training, validation and test cases are chosen according to equal class distribution of benign and malignant. The split of training and test data is to make sure the test set is large enough to yield statistically meaningful results and is representative of the dataset. In other words, our goal is to make sure that the model generalizes well to new data.

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 mask annotation process was carried out manually by 2 experienced doctors. They firstly drew the boundary of thyroid nodule based on ultrasound images of patients. The classification annotation was based on the biopsy result of each patient.

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 annotators follows the Thyroid Imaging Reporting and Data Systems (TI-RADS) [4] and the American Thyroid Association (ATA) guidelines [5].

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

The dataset comes from the Chinese Medical Ultrasound Artificial Intelligence Alliance (CMUAA) initiated and responsible by Dr. JianQiao Zhou who has more than 20 years of experience in thyroid ultrasound. The annotation process was carried out by 2 senior doctors with more than 5 years of experience.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No merge method. There is one single annotation for one case.

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.

The US images were anonymised by first removing any patient related information on each image. Then all the images were renamed by assigning an ID number to each patient and converted to the same format (.png) . No information concerning the relations between the identity and the ID number of each patient will be released to the public.

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 delineation of structures in ultrasound images is a challenging task as some of the boundaries is less well-defined. In this case, there might not exist a ground truth of annotations.

b) In an analogous manner, describe and quantify other relevant 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)

Evaluation measures include: 1) the intersection of union (IoU) between the prediction and the ground truth; 2) thyroid nodule classification F1score.

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

Segmentation IoU score: The IoU score is calculated the area of the intersection of the two regions divided by the area of their union set. It is a good indicator of whether the prediction is consistent with the label [6].

Classification F1score: The F1 score is calculated based on the precision and recall of each class. It is the weighted average of the precision and the recall scores. The F1 score reaches its perfect value at one and worst at zero. It is a very good way to show that a classifier has a good recall and precision values [7].

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.

Thyroid nodule segmentation: For each participant pi and each test case cj: Compute the metric values for the IoU score.
Thyroid nodule classification: For each participant pi and each test case cj: Compute the metric values for the F1score.
For each participant pi and each test case cj, determine the rank corresponding to both metrics (R(IoU, pi, cj): descending order for IoU score, R(F1, pi, cj) ascending for F1score). Finally, compute the average over all case-specific ranks to get one final rank for each participant pi of each task.
In case of tied positions, perform the ranking according to the standard deviation of the two metrics.

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

If there are the missing results on the test cases, the submission will fail and the error notification will appear. The participants need to re-submit their results.

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

This ranking scheme aim to fairly evaluate the performances of the participating team algorithms on the individual and combination of the thyroid nodule segmentation task and thyroid nodule classification task. From the ranking scheme , we can compare the performances of them and define the top winners.

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.

No statistical analyses

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

No statistical analyses

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.

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] https://www.btf-thyroid.org/what-is-thyroid-disorder.
[2] https://www.thyroid.org/wpcontent/uploads/patients/brochures/Nodules_brochure.pdf.
[3] https://www.cancer.net/cancer-types/thyroid-cancer/statistics.
[4] Horvath, Eleonora, et al. "An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management." The Journal of Clinical Endocrinology & Metabolism 94.5 (2009): 1748-1751.
[5] Haugen, Bryan R., et al. "2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer." Thyroid 26.1 (2016): 1-133.
[6] Caicedo, Juan C., et al. "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images."

Cytometry Part A 95.9 (2019): 952-965.
[7] Singh, Vivek Kumar, et al. "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network." Expert Systems with Applications 139 (2020): 112855.